COVER PAGE

DOE award #: DE-SC0008462

Name of recipient: Michigan Technological University

Title: "Functional Gene Discovery and Characterization of Genes and Alleles Affecting Wood Biomass Yield and Quality in *Populus*"

Project director: Victor Busov

Members: Robert Sykes, Wellington Muchero, Gerald Tuskan

EXECUTIVE SUMMARY

Adoption of biofuels as economically and environmentally viable alternative to fossil fuels would require development of specialized bioenergy varieties. A major goal in the breeding of such varieties is the improvement of lignocellulosic biomass yield and quality. These are complex traits and understanding the underpinning molecular mechanism can assist and accelerate their improvement. This is particularly important for tree bioenergy crops like poplars (species and hybrids from the genus *Populus*), for which breeding progress is extremely slow due to long generation cycles. A variety of approaches have been already undertaken to better understand the molecular bases of biomass yield and quality in poplar. An obvious void in these undertakings has been the application of mutagenesis. Mutagenesis has been instrumental in the discovery and characterization of many plant traits including such that affect biomass yield and quality. In this proposal we use activation tagging to discover genes that can significantly affect biomass associated traits directly in poplar, a premier bioenergy crop. We screened a population of 5,000 independent poplar activation tagging lines under greenhouse conditions for a battery of biomass yield traits. These same plants were then analyzed for changes in wood chemistry using pyMBMS. As a result of these screens we have identified nearly 800 mutants, which are significantly ($P<0.05$) different when compared to wild type. Of these majority (~700) are affected in one of ten different biomass yield traits and 100 in biomass quality traits (e.g., lignin, S/G ration and C6/C5 sugars). We successfully recovered the position of the tag in approximately 130 lines, showed activation in nearly half of them and performed recapitulation experiments with 20 genes prioritized by the significance of the phenotype. Recapitulation experiments are still ongoing for many of the genes but the results are encouraging. For example, we have shown successful recapitulation for a fascilin-like gene that when overexpressed increase many biomass-yield associated traits. Genes discovered through activation tagging showed polymorphisms in *P. trichocarpa* association mapping population linked to the traits modified by the activation tagging. This suggests that putative alleles that are associated with improvement of the trait o interest can be discovered and used in marker associated selection. This will significantly simplify and accelerate the breeding efforts.

ACCOMPLISHMENTS BY OBJECTIVES

1. **Screen a population of 5,000 activation tagged lines for changes in woody biomass quality and yield.**

To identify genes that affect traits linked to biomass yield in *Populus*, we screened a population of 5,000 poplar activation tagged lines under controlled greenhouse conditions. A total of 762 lines with at least one significantly affected trait were discovered at P<0.05 when compared to WT controls under the same conditions. We will refer to these phenotypic lines as mutants. More than half of the mutants (402) were affected in only one trait (Figure 1a). Interestingly, some of the mutations impacted simultaneously as many as 5-8 traits (Figure 1b). The various mutations had an approximately evenly distributed impact over the 10 studied traits. Number of internodes was most impacted among all traits (Figure 1a). Diameter at leaf plastochron index (LPI) 5, green density, dry weight at the base and dry leaf weight were impacted the least (Figure 1a).

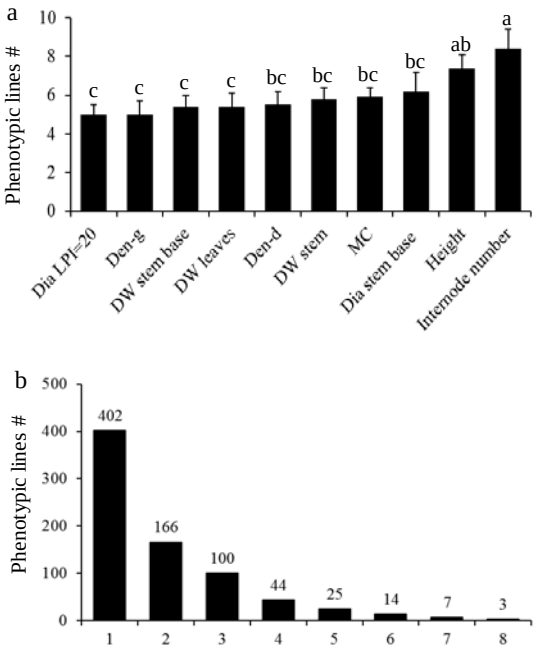


Figure 1 Mutant identification and distribution across different biomass yield traits. Phenotypic traits presented with mutant discovery rate (mean ± SE, n=25). Different letters indicate significance as determined by a one-way ANOVA followed by Fisher's test ($p < 0.05$). (b) Distribution of biomass mutants isolated after analysis of biometric data of 25 subsets each presented of approximately 120 independent lines grown in three replicates. A single line with significant increase determined by Student's *t*-test ($p<0.05$, $p<0.01$ or $p<0.001$) in at least one parameter was classified as a mutant after comparison with the entire subset.

Pair-wise phenotypic correlations were estimated for all traits across the entire experiment (Table 1). Not surprisingly, high correlations were observed between tightly dependent or derivative traits like green and dry density; dry biomass of the whole stem and dry biomass of the stem base. However, some other correlations were also found. For

	Height	Int #	Dia stb	Dia LPI20	DW stb	MC	Den-g	Den-d	DW st	DW leaves
Height	1									
Int #	0.40	1								
Dia stb	0.50*	0.60**	1							
Dia LPI20	-0.17	0.09	0.15	1						
DW stb	0.19	0.35	0.56**	0.28	1					
MC	0.12	0.14	0.23	0.13	0.33	1				
Den-g	0.01	0.34	0.20	0.28	0.34	0.55**	1			
Den-d	0.08	0.25	0.24	0.12	0.32	0.55**	0.86***	1		
DW st	0.38	0.54**	0.61***	0.19	0.70***	0.32	0.41	0.37	1	
DW leaves	0.14	0.41	0.59**	0.16	0.58**	0.31	0.17	0.26	0.51**	1

Table 1 Correlation between different phenotypic traits. Regression analysis was used to test the significance in the correlation between pair-wise traits. "*", "**" and "***" denote significant correlation at $p \leq 0.05$, $p \leq 0.01$ and $p \leq 0.001$, respectively. *Abbreviations*: int # - internode number, dia – diameter, stb - stem base, st – stem, LPI20 – leaf plastochron index 20, dw – dry weight, mc – moisture content, den-g – density green, den-d – density dry.

example, dry leaf biomass was significantly and positively correlated with diameter of the stem base and dry stem biomass. Dry stem biomass was positively correlated with internode number and diameter of the stem base.

All lines that were screened for biomass yield traits were also studied with respect to changes in their cell wall chemistry with main focus being the lignin S/G ratio and the C5 and C6 sugars. The woody stem at the base of the stem was milled into fine powder and analyzed via pyrolysis mass beam mass spectroscopy (pyMBMS) in collaboration with Robert Sykes at NREL, Boulder, Colorado. A total of 86 lines were found to be significantly different form WT at P<0.001 with highest number of lines significantly affected in lignin concentration.

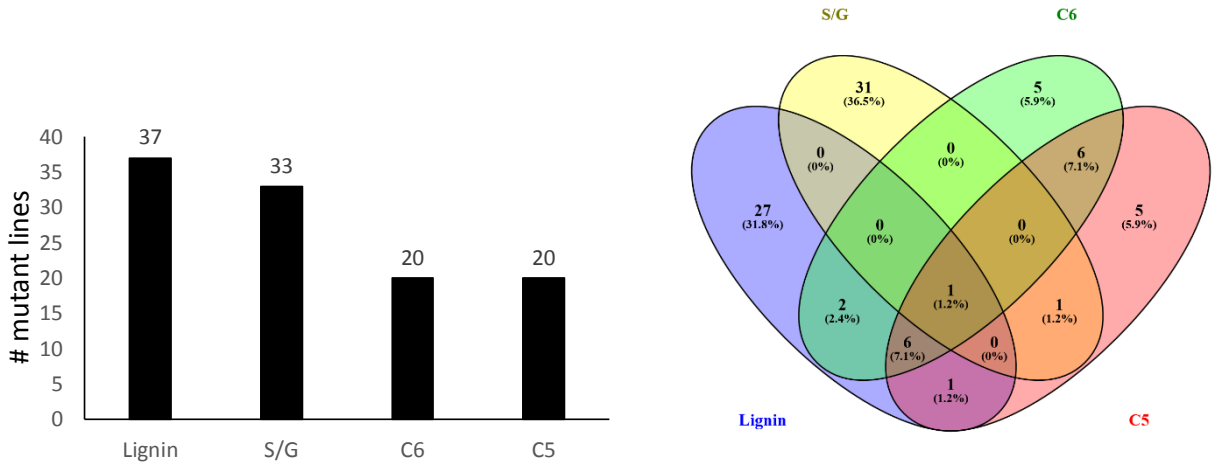


Figure 2. Distribution of the cell wall mutants among the different characteristics. C6 and C5 indicate C6 and C5 sugars in the cell wall like cellulose and hemicellulose respectively.

2. Identify estimated +160 tagged genes in all mutants through positioning in the genome and validation of activation.

We have recovered the tag in approximately 130 mutant lines. These all have been sequenced and positioned in the poplar reference genome and for a subset the activation of the putative tagged gene validated through real time RT-PCR comparison with WT control plants. Majority (71.3%) of the insertions were found to be in the proximal 10 Kbp of the 5′/3′ intergenic regions of the genes, 12% were in introns and 16.7% in exons. Insertions were approximately evenly distributed among the chromosomes (Figure 2a) with number of insertions highly significantly correlating with the size of the chromosome (Figure 2b). Activation tag in

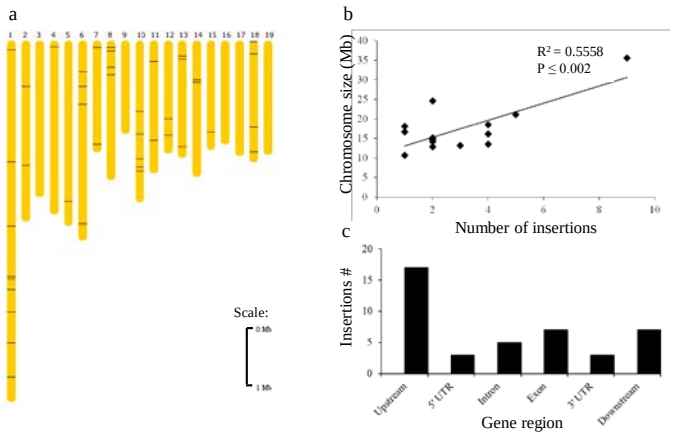


Figure 2 Characterization of the T-DNA integration in the *Populus* genome. (a) Distribution of the insertions within the chromosomes. (b) High correlation between chromosome length and number of insertions per chromosome. (c) T-DNA insertions are predominantly identified upstream the gene coding region.

mutant lines was predominantly found in the upstream putative promoter region flanking the coding sequence (Figure 2c). These are most 'productive' insertions with respect to activation of the gene and thus consistent with their predominance in mutant lines.

3. Recapitulate through retransformation the effect of 20 tagged genes prioritized by novelty and severity of the phenotype.

We have recapitulated the effects of 10 biomass yield and 10 biomass quality activation tagged genes. These studies are in various stages of characterizations and will be published and full comprehensive analyses become available. Biomass quality mutations characterization is proceeding slower as in addition to the biometric analyses, cell wall composition is being undertaken. We have selected genes form mutants that either have strong phenotypic effect or have multi-trait effects. For example, one of the lines selected for recapitulation experiments was A630-7. This line was selected because affected simultaneously many (5) traits. However, most importantly, the mutation positively impacted the most important aspect of biomass production – stem dry weight. In addition, the line also showed increased leaf dry biomass and as mentioned earlier, the leaf dry biomass was positively correlated with stem dry weight (Table 1). We therefore were interested in the gene that confers these properties and potentially drives these correlations. The activation tagged gene in the line showed highest homology to *FLA10* from Arabidopsis and thus named *PtaFLA10*. We thus produced overexpression constructs for *PtaFLA10* and transformed them in transgenic plants. Numerous independent events were regenerated with the *PtaFLA10* overexpression construct (oe-*PtaFLA10*). *PtaFLA10* overexpression positively affected several phenotypic traits linked to biomass growth compared to WT-717 plants (Figure 4a-h). For

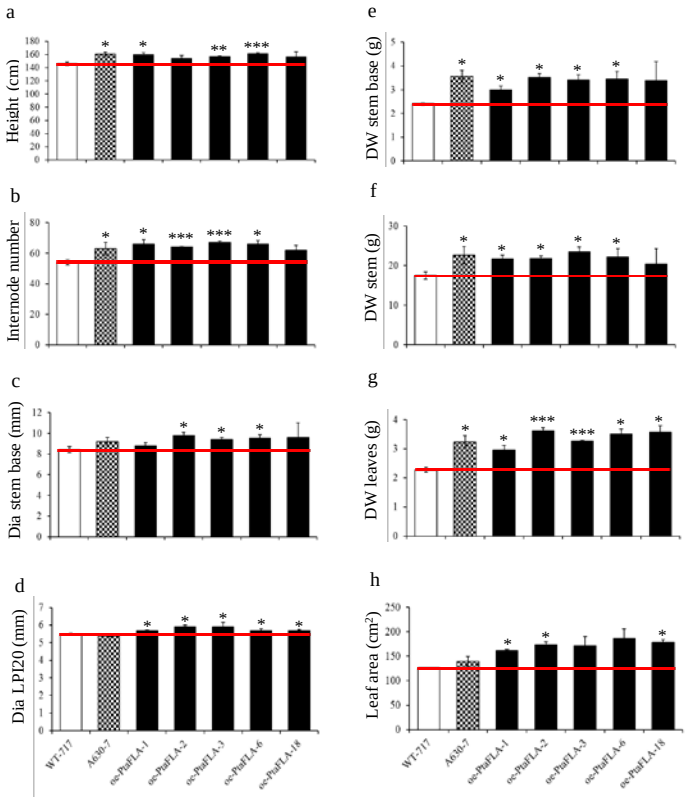
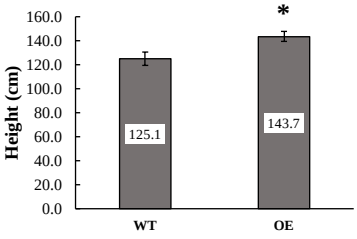


Figure 4 Recapitulation of A630-7 phenotype. (a) – (h) Changes of traits in mutant and *oe-PtaFLA* lines. Values are reported as mean ± SE (n=4). White bars represent wild type, checker board bars indicate mutant line, black bars – PtaFLA OE=overexpression lines. Asterisks indicate significant differences between transgenics and wild type plants as determined by Student's *t*-test (*, ** and *** denoting $p<0.05$, $p<0.01$ and $p< 0.001$, respectively). Red line corresponds to the wild types' threshold.

instance, both the original mutant A630-7 line and oe-*PtaFLA* lines were about 9% taller than the wild type. Similarly, the number of the internodes was increased with 15% in the A630-7, while the oe-*PtaFLA* lines on average had 20 % more internodes. The *oe-PtaFLA10* lines showed changes that were not observed in the original mutant and likely result from the strong overexpression. For example, significant increase was measured in the diameter at the 20th internode and at the stem base and. Most importantly, 40% increase in the dry stem biomass was observed in both the original A630-7 mutant line and oe-*PtaFLA10* transgenics. Consistent with the strong correlation between leaf and stem dry weight (Table 1), leaf dry weight was also impacted in the mutant and recapitulation transgenics (Figure 4g).

Figure 5. Recapitulation of line 273p-11. The plants to the left are typical wild type (WT) plants, while the plants to the right are three ramets of a representative line overexpressing (OE) the activated gene. Graph on the bottom show significant height difference of the OE lines compared to WT using Student T-test. *-P<0.05.



We are also recapitulating a number of the tagged genes which affect cell wall properties. As mentioned earlier these studies are still ongoing with respect to estimation of cell wall content. However preliminary data from the greenhouse growth of the recapitulation transgenics shows that the lines are indeed phenotypic (Figure 5). For example, the gene isolated from line 273p-11 when overexpressed causes increased height growth (Figure 5). This line was originally isolated because of increased cellulose and decreased lignin content.

4. **Use the resequencing and phenotypic data for 1,100 *P. trichocarpa* individuals to identify rare genotype-phenotype associations in the population of the +160 tagged genes.**

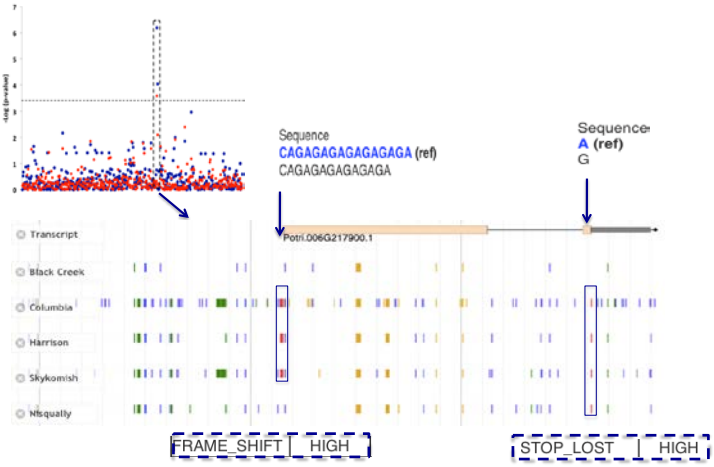


Figure 6. High level of association between cell wall characteristics and polymorphisms in one of the activation tagged genes.

All of the tagged genes were sent to our collaborators Drs. Muchero and Tuskan from ORNL, Oakridge, TN for analyses of potential associations of the trait characteristics found to be affected in the mutants with the variation of similar traits in the natural populations with SNP polymorphisms in 1,100 re-sequenced individuals. A number of genes did show strong associations. For example, one of the

tagged genes identified in a mutant line with changes in lignin characteristics was found to harbor polymorphisms among the 1,100 re-sequenced individuals that show significant association with lignin content (Figure 6).

PRODUCTS

Publications

1. Dash M., Yordanov Y.S., Georgieva T., Tschaplinski TJ., Yordanova E., Busov V. (2016) Poplar PtabZIP1-like enhances lateral root formation and biomass growth under drought stress. doi: 10.1111/tpj.13413
2. Dash, M., Yordanov, Y.S., Georgieva, T., Kumari, S., Wei, H., and Busov, V. (2016). A network of genes associated with poplar root development in response to low nitrogen. Plant Signal Behavior. 10.1080/15592324.2016.1214792
3. Busov, V., Carneros, E., and Yakovlev, I. (2016). EARLY BUD-BREAK1 (EBB1) defines a conserved mechanism for control of bud-break in woody perennials. Plant Signal Behav. 10.1080/15592324.2015.1073873
4. Haiwei Lu,Venkatesh Viswanath; Cathleen Ma;Elizabeth Etherington; Palitha Dharmawardhana; Olga Shevchenko; David W Pearce; Stewart B Rood; Victor B Busov; Steven H Strauss (2015) Recombinant DNA modification of gibberellin metabolism alters growth rate and biomass allocation in Populus 11: 127 DOI 10.1007/s11295-015-0952-0
5. Dash, M. , Yordanov, Y.S., Georgieva, T., Kumari, S., Wei, H., and Busov, V. (2015). A systems biology approach identifies new regulators of poplar root development under low nitrogen. The Plant Journal 84, 335-346.
6. Yordanov YS, Ma C, Strauss SH, Busov VB (2014) EARLY BUD-BREAK 1 (EBB1) is a regulator of release from seasonal dormancy in poplar trees. Proc Natl Acad Sci U S A 111: 10001-10006
7. Xiang Li, Yufeng Guo, Chathura Gunasekara, Hang Zhang, Liang Lei, Sermsawat Tunlaya-Anukit, Vincent Chiang, Victor Busov, Hairong Wei (2014) Pop's Pipes: Poplar Gene Expression Data Analysis Pipelines. Tree Genetics and Genomes 10:1093–1101.
8. Zawaski, Christine and Victor B. Busov. (2014) Roles of gibberellin catabolism and signaling in growth and physiological response to drought and short-day photoperiods in *Populus* trees. PLOS ONE, 9 (1), e86217
9. Wei, Hairong, Yordan Yordanov, Tatyana Georgieva, Sapna Kumari, Victor Busov Genetic networks involved in poplar root response to low nitrogen. (2013) Plant Signaling and Behavior. 8(11) e27211