CBI Performance Metric for FY20: Report on genomic science-based advances and testing of new plant feedstocks for bioenergy purposes.

Q1 Metric: Report on the latest genomic insights into plant cell wall formation that informs the development of dedicated bioenergy crops.

January 2020

Introduction

Both quantitative trait loci (QTL) approaches, using structured pedigrees, and genome-wide association (GWAS) approaches, using arbitrary collections of unrelated genotypes, were used to gain insight in plant cell wall formation in *Populus*. This quarterly report focuses on *Populus* species as poplar was the first sequenced tree and has the most fully developed dedicated bioenergy feedstock genomic and other 'omics resources. This has also been a focus of research in both the Center for Bioenergy Innovation (CBI) as well as the prior BioEnergy Science Center (BESC). As QTLs and quantitative trait nucleotides (QTNs) from the GWAS studies were identified multiple 'omics approaches were then used to validate the statistical association found in the either approach. These 'omics approaches included RNAseq characterization, metabolic profiling, transient expression using isolated protoplasts, as well as stable transgenic and cisgenic evaluations of the targeted phenotype. Cell wall phenotypes (e.g., composition and recalcitrance) were identified and characterized using pyrolysis molecular beam mass spectrometry (pyMBMS), gas chromatography mass spectrometry (GC-MS), and nuclear magnetic resonance (NMR). These approaches were used to validate candidate cell wall formation genes which were then tested via cis- or transgenics in poplar or in model plants (e.g., *Arabidopsis thaliana*), respectively. We also used *in vitro* enzyme expression and characterization platforms based on heterologous expression in human embryonic kidney (HEK293) cells as an expression validation assay. The creation of the single nucleotide polymorphism (SNP) library, establishment of the common gardens, development of the analytical platforms and optimizing the transformation systems occurred over a 10-year period. Outcomes and conclusions from this investment are presented in chronological order starting with the QTL results followed by the GWAS results.

Highlighted Results

Quantitative trait loci study. We initially developed 148,428 simple sequence repeat (SSR) marker set from the unambiguously mapped sequence scaffolds of the *Populus trichocarpa* Nisqually-1 genome and completed the genetic map for Family 822. Here, 20 QTLs were detected for cell wall chemistry and cell wall precursor metabolites. Our findings revealed that lignin content and the S/G ratio in *Populus* stems is

Figure 1. QTLs detected for above- (red bars) and below- (green bars) ground lignin content linked to chromosome 10 (vertical bar on the left designate genetic marker position).

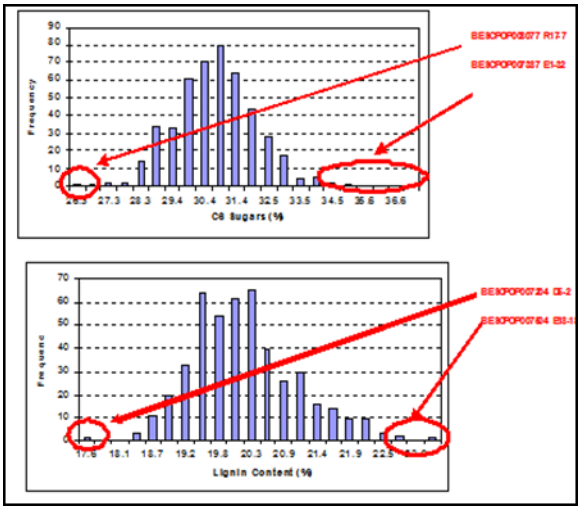

Figure 2. pyMBMS wood chemistry data, where lines in red represent those with extreme wood chemistry phenotypes.

significantly higher than lignin content and S/G ratios in roots. Large-effect QTLs were detected on chromosomes 2, 3, 6, 10 and 13 (e.g., Fig. 1). It is noteworthy that below-ground and above-ground lignin content is controlled by different QTL. We then designed an Illumina array targeting 6,000 SNP that was used to genotype 700 progeny in Family 52–124, 300 progeny in Family 331 and 100 progeny in Family 545 to verify the initial QTL results. Cell wall phenotypes, i.e., lignin content, syringyl to guaiacyl (S/G) ratios, C6 sugar content and C5 sugar content varied across the tested QTL pedigrees (Fig. 2). This variation permitted the mapping of individual QTLs for each measured trait.

Major QTLs affecting lignin content and S/G ratio were reproducibly identified from the larger mapping pedigree, where 712 individuals were subjected to molecular beam mass spectrometry (MBMS) compositional and saccharification assays. A high-density single nucleotide polymorphism (SNP) genetic map with 3,559 markers was used to delimit genomic regions harboring QTLs for several recalcitrance phenotypes. Based on this analysis, at least four distinguishable major QTLs, spanning the length of chromosome 14 of the *Populus* genome assembly, were identified (Fig. 3A). Logarithm of odds (LOD) profiles across the chromosome exhibited significant overlap between lignin and S/G ratio QTLs suggesting possible pleiotropic interactions. Subsequently, an unselected natural or wild-type population of *P. trichocarpa* was subjected to saccharification and MBMS assays and linked to 30,229 SNPs loci across the genome. This was our first attempt at association mapping and was conducted using the combined SNP and phenotypic data which showed significant marker-trait SNPs in four genes that corresponded with QTL peaks previously identified on scaffold 14 (Fig. 3B).

Alternate alleles of two transcription factors (TF) and a protein kinase carrying SNPs with significant

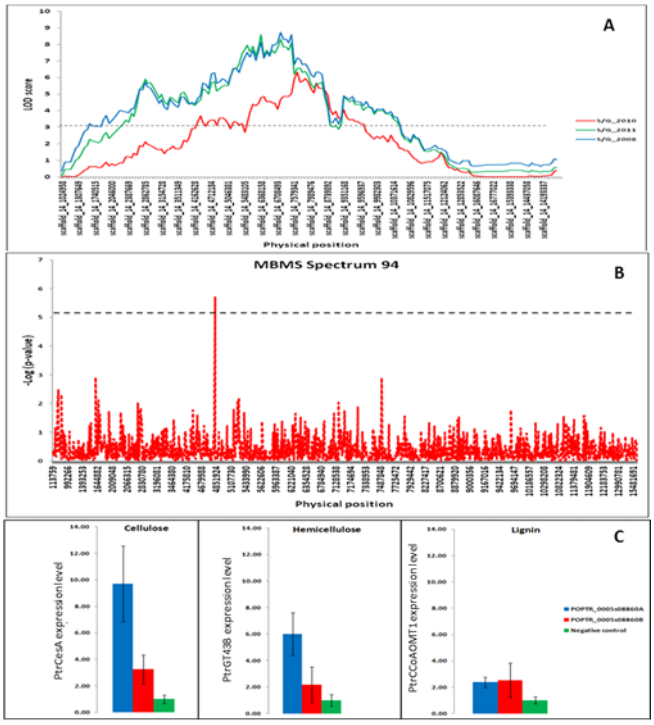

Figure 3. A) Major QTL controlling S/G ration on chromosome 14 segregating in Family 52-124, **B**) SNP locations that were significantly associated with S/G ratio. **C**) Allelic variants of the genes under lying the QTL and their impact on cell wall chemistry.

trait associations were identified and successfully cloned from a set of natural variants maintained in the greenhouse. These cDNAs were used to transform *Populus* protoplasts to assay comparative effects of alternate alleles on expression of markers genes routinely used to infer activation of the cellulose (CesA), hemicellulose (GT43B) and lignin (CoAOMT1) biosynthetic pathways (Fig. 3C).

> *These results were first indication that biomass related traits such as cell wall chemistries were genetically controlled and could be manipulated via genetic transformation or genomic selection approaches.*

Establishing the primary genome-wide association study. We initially obtained stem cuttings from 496 genotypes included in various existing *P. trichocarpa* clone trials, as well as an additional 603 genotypes growing in five natural populations, representing most of the geographic range of *P. trichocarpa*. This association mapping population was planted in four replicated field trials in Agassiz, BC, Clatskanie, OR, Corvallis, OR and Placerville, CA. At the end of the 2009 growing season, survival was high in the Clatskanie (99%) and Corvallis (98%). Survival was lower at the other sites and selected replanting was attempted. We used 10 SSR markers to characterize population substructure in two ecologically and demographically contrasting populations of *P. trichocarpa* because undetected population substructure can cause spurious phenotype-genotype correlations and is a major concern in association genetic studies. At this time, the DOE Joint Genome Institute (JGI) had completed the whole-genome Illumina resequencing of 16 *P. trichocarpa* genotypes chosen to span the geographic range represented in the association mapping population. These data were used to 1) discover single nucleotide polymorphism (SNP) markers that were used to test for phenotype-genotype associations and 2) characterize linkage disequilibrium across the genome of *P. trichocarpa*. In order to assess the reliability of polymorphisms discovered based on whole-genome, short-read Illumina re-sequencing, we also re-sequenced a set of ten candidate genes in a population of 48 trees (i.e., which includes 15 of the trees used for whole-genome re-sequencing) using traditional Sanger sequencing technology. We discovered 16,818,364 SNPs and generated a library of 169,626 potential target SNPs from the 3704 candidate genes that were categorized as high- or medium-quality. The resulting library represented a median of 43 SNPs per candidate gene with only 131 genes (3.5%) with fewer than 10 target SNPs.

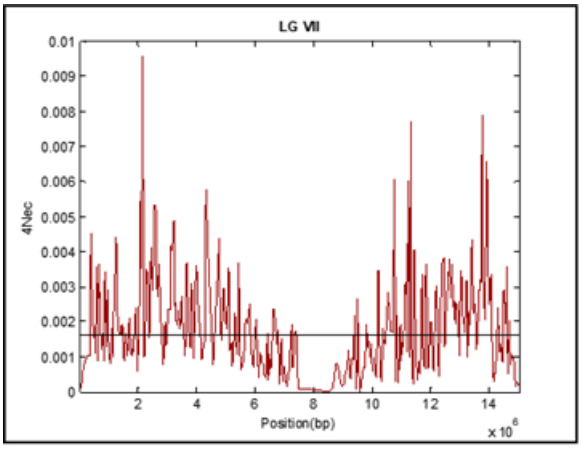

Figure 4. Variation of the population recombination rate (4Nec) for one of the 19 *Populus* chromosomes, chromosome 7.

Finally, we selected 38,000 target SNPs based on 1) information about linkage disequilibrium (LD), 2) design scores and assay types for Infinium iSelect HD Custom Genotyping, 3) annotation information, 4) degree of subpopulation differentiation, and 5) genome distribution considerations. LD (as measured by the r^2 statistic) decays steeply across the genome. However, LD appeared to decay slower in the genome-wide data than in data from resequencing of candidate genes. We concluded that LD varied across the

genome of *P. trichocarpa* and regions of high LD exist. Finally, rates of recombination (measured by 4Nec) varied dramatically among candidate genes and across the genome (e.g., Fig. 4). At this point, the potential existence of haplotype blocks and recombination hotspots in the genome of *P. trichocarpa* had to be confirmed, with major implications for our basic understanding of the biology of the species.

The closed *Populus trichocarpa* association study population was completed and consisted of 1,100 trees collected from a range of species throughout California to British Columbia. During this time, the JGI completed the draft resequencing of all ~1,100 genotypes in the population. This data is available on the JGI website. We used a population genomics approach to determine the distribution of neutral and adaptive genetic variation in this population. For example, the detection of reliable phenotype-genotype associations and molecular signatures of selection requires detailed knowledge about genome-wide patterns of allele frequency variation, linkage disequilibrium (LD) and recombination. Analyses of the population structure at multiple spatial scales revealed significant geographic differentiation that was consistent with models of Isolation by Distance (Fig. 5). Furthermore, latitudinal allele frequency gradients were strikingly common across the genome, affecting approximately 25% of SNPs. The decay of genome-wide LD relative to physical distance was much slower than expected from smaller-scale studies in *Populus*, with a r^2 value dropping below 0.2 within 6 kb. Consistent with this observation, estimates of effective population size from LD (Ne≈4000-6000) was remarkably low relative to the large census size of *P. trichocarpa* stands. Fine-scale rates of recombination varied widely across the genome but were largely predictable based on DNA sequence and methylation. Taken together, these results have implications for our understanding of the evolutionary history of *P. trichocarpa* and the design of robust selection scans. Importantly, the detected LD suggests that genomic selection from association studies in undomesticated populations may be more feasible in *Populus* than previously assumed.

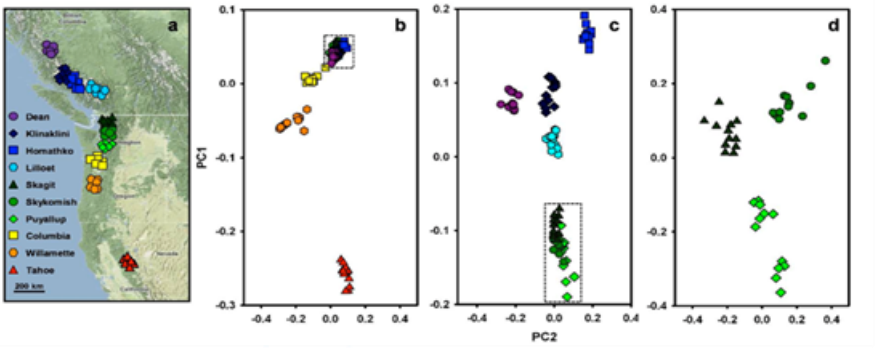

Figure 5. Population structure in *P. trichocarpa* based on Illumina SNP array. **a**) Sampling locations of the 120 *P. trichocarpa* trees from 10 subpopulations chosen to span the geographic distribution of the 16 re-sequenced trees. **b-d**) Major axis of variation based on PCA of 22,280 SNPs on a range-wide scale (**b**), after excluding tree from the clearly differentiated Tahoe, Willamette and Columbia subpopulations (**c**) and within the most homogeneous set of subpopulations in western Washington.

Field trials for the association population were maintained and measured at four sites spanning most of the latitudinal range of the collection area: Placerville, CA; Corvallis, OR; Clatskanie, OR; and Agassiz, BC. Each trial was established with three clonal replicates of all 1,100 genotypes and survival ranges between

~70% in Placerville to ~90% in Clatskanie. Genetically based phenotypic differentiation among groups of individuals was estimated as Q_{ST}, which in turn is analogous to F_{ST}, or genetic differentiation among the groups at molecular markers. Here we assumed that a neutral trait will diverge due to demographic factors and thus the neutral expectation is that Q_{ST} will be equivalent to F_{ST}. However, we found that Q_{ST} was significantly greater than F_{ST}, suggesting that divergent selection has occurred for these traits.

This analysis demonstrated that our network of geographically distributed plantations was able to accurately estimate quantitative components of phenotypic variation and ultimately identify the underlying loci controlling these traits. Such results can be utilized by breeding programs to enhance the productivity of bioenergy plantations across a wide range of environments, thus ensuring a stable supply of feedstock for lignocellulosic biofuel refineries.

Cell wall chemistry phenotyping. Using cisgenic materials, GCMS-based analysis of down-regulated Korrigan (KOR) and up-regulated UDP-glucose pyrophophorylase genes was evaluated. Metabolic profiles for both up- and down-regulated materials were similar, likely reflecting a dysfunctional wall assembly phenotype. Surprisingly, metabolic responses in all cisgenics were limited in xylem, however major increases in most caffeoyl-, caffeoyl-glycoside- and caffeoyl-shikimate-conjugates were detected in leaves and phloem. Declines in a limited number of metabolites were indicative of disrupted walls. Two unique caffeoyl-glycoside-conjugate metabolites accumulated in leaves of cisgenics. Disrupted metabolism was evident well beyond the targeted cellulose synthesis genes (e.g., changes in lignin and aromatic acid pathways) and lead to the discovery of an uncharacterized higher-order conjugates of caffeoyl-shikimate. Limited responses in the xylem coupled with pronounced responses in the phloem suggests that production of lignin precursors and their modification occurs in the phloem with products subsequently transported to the xylem.

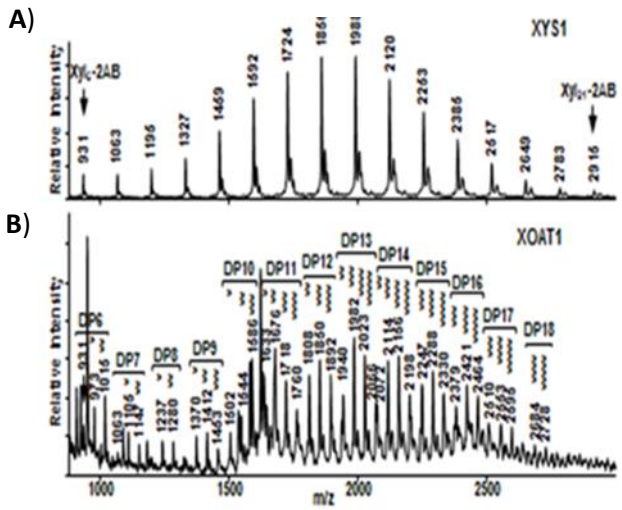

Figure 6. Xylan synthase (XYS1) and xylan *O*-acetyl transferase (XOAT1) heterologously expressed in human embryonic kidney (HEK293) cells synthesized and *O*-acetylated, respectively, xylan *in vitro*. **A**) MALDI-TOF-MS analysis of the products formed after incubation XYS1 for 96 h with a labeled acceptor (Xyl_6-2AB) and UDP-Xyl. Annotated $[M+H]^+$ ions correspond to the xylooligosaccharides with 6-21 residues. **B**) Products of the reaction in 7A were incubated with XOAT1 and acetyl-CoA for 16 h. The number of acetates added to each oligosaccharide is indicated by asterisks above it $[M+H]^+$ ion.

Based GWAS results and transgenic validation experiments, we definitively identified the primary xylan synthase and characterized the archetypal plant polysaccharide *O*-acetyltransferase, both accomplishments represent breakthroughs in understanding hemicellulose synthesis. Specifically, we demonstrated that *A. thaliana* IRREGULAR XYLEM 10-L is a β-(1→4) xylosyl transferase, that elongates the xylan backbone and

ESKIMO1/TRICOME BIREFRINGENCE 29, is a xylan *O*-acetyltransferase. We showed that the xylan backbone can be synthesized and *O*-acetylated *in vitro* using only these two enzymes in combination (Fig. 6). This work, sheds light on integral steps in the molecular pathways utilized by plants to synthesize a major component of plant biomass and was accomplished by combining purified, recombinant CAZymes developed with high-throughput glycosyl-, methyl- and acetyltransferase assays.

Immunolabeling of stem sections of the *Arabidopsis* gaut12 (irx8) mutant, which has reduced xylan and pectin, with monoclonal antibody CCRC-M138 reactive against an unsubstituted xylopentaose epitope, revealed a bi-lamellate pattern in wild-type fiber cells and a collapsed bi-layer in irx8 cells. These results suggest that in *Arabidopsis* fiber cells GAUT12 participates in the synthesis of a specific layer of xylan or provides an architectural framework required for the native xylan deposition pattern. We propose the hypothesis that GAUT12 functions in the synthesis of a structure required for xylan and lignin deposition during secondary cell wall formation.

RNAi down-regulation of one of the two *Populus deltoides* GAUT12 homologs (PtGAUT12.1) resulted in transgenic lines with 50–67% reduced PtGAUT12 transcript expression compared to controls. These lines had 4–8% greater glucose release and 12–52% and 12–44% increased plant height and radial stem diameter, respectively, compared to the controls, which is attributed to increased phloem, vessel, fiber and palisade cell sizes. Cell wall analyses revealed the PtGAUT12.1 KD lines to have 25–47% reduced GalA and 17–30% reduced xylan without affecting total lignin content. Sugar analyses of sequential cell wall extracts proved a reduction of both glucuronoxylan and the pectins. The results show that GAUT12 function is required for xylan and pectin biosynthesis during *Populus* wood formation and that controlled down regulation of GAUT12 leads to reduced recalcitrance, more easily extracted cell walls, and increased growth in *Populus*.

We have identified covalent linkages between cell wall glycoproteins and wall polysaccharides and analyzed the function of proteoglycans isolated from *Populus* wood tissues and *Arabidopsis* cultures. In addition to the characterized proteoglycan arabinoxylan-pectin-arabinogalactan-protein 1 (APAP1), we isolated a xyloglucan-rich glycoprotein fraction from *Arabidopsis*, a glucuronoxylan-rich glycoprotein fraction from *Populus*, and an apiose-rich arabinogalactan-protein fraction from duckweed. Our current data indicates that there are additional covalent linkages between wall polysaccharides and wall glycoproteins or proteoglycans, and thus, that models of the wall that depict only polysaccharides and lignin are oversimplified. The glycan-protein covalent linkages may contribute to cell wall recalcitrance. To test this, transgenic and gene family knock down (KD) transgenics of two *Populus* AtAGP57C/APAP1 orthologs were generated. A subset of the overexpression lines exhibits a dwarf phenotype after 5 months of growth, which is the same phenotype exhibited by *Arabidopsis* AtAGP57C overexpression plants. We also analyzed the expression pattern of AtAGP57C/APAP1 in *Arabidopsis* using GUS staining and qPCR. Overexpression of AtAGP57C caused a short stem phenotype compared to wild-type, the inverse phenotype of the apap1 mutants. Both glycome-profiling and sugar composition analyses showed that overexpression plants had more tightly bound arabinogalactan, xylan, and pectin epitopes than wild-type, an inverse chemotype of the APAP1 mutants. We concluded that AtAGP57C is involved in stem growth and affects cell wall structural integrity.

Based on cell wall chemistry characterization and its superior performance across common garden test sites, a new *Populus* reference biomass clone (GW-9947), representative of the most comprehensive phenotypic and molecular genetics data, was selected as our internal standard for all future cell wall chemistry, sugar release assays and metabolic profiling experiments. In the most comprehensive test of Populus cell wall properties, 1,000 increment cores from Corvallis, OR, were processed and characterized using NMR, py-MBMS, and C5/C6 compositional analyses. We also subjected these samples to the newly established X-ray tomography imaging platform at NREL. Finally, we summarized field performance data for *Populus deltoides* transgenics engineered for enhanced biofuels yields and found no significant differences for productivity and sustainability phenotypes, suggesting that altered cell wall composition in the tested transgenic lines does not negatively impact plant performance (Fig. 7).

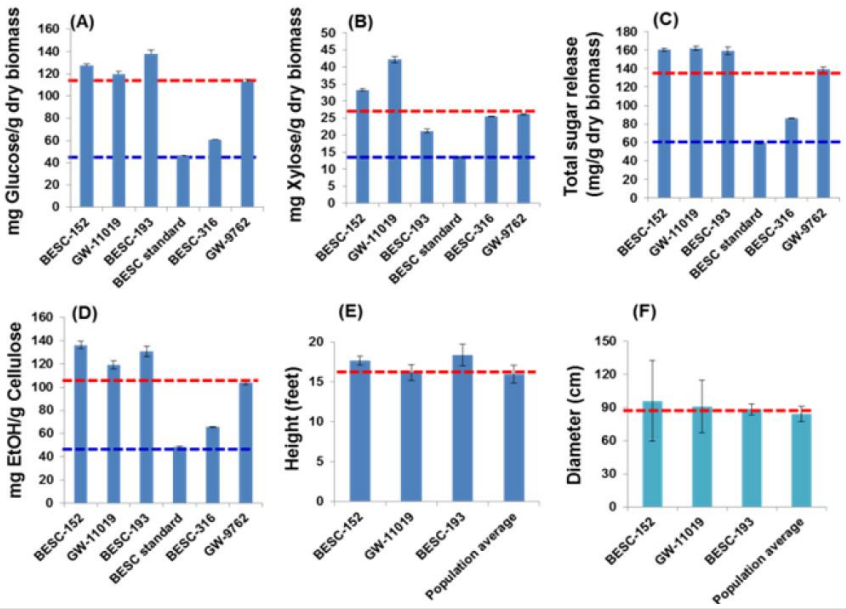

Figure 7. *Populus* TOP Line natural variants. Sugar release data (**A-C**) and ethanol yield (**D-F**) data were collected by using separate hydrolysis and fermentation of un-pretreated biomass collected from 4-year-old trees growing in Clatskanie, Oregon.

Finally, several *Populus* natural variants that displayed increased rates of sugar release and uncompromised plant growth were selected as TOP Lines (Fig. 8). TOP Line #15 (BESC-152) has normal lignin content and syringyl/guaiacyl ratio (S/G), TOP Line #16 (BESC-193) has reduced lignin content and TOP Line #17 (GW-11019) has increased S/G ratio. These results suggest that these three TOP Lines positively reduce recalcitrance through alternate, distinct mechanisms. SnpEff revealed that in each of these natural variants, a number of genes relevant to cell wall biosynthesis possess homozygous high-impact SNPs. The causal SNPs for the observed sugar release and ethanol yield phenotypes remain unknown. It is likely that the observed phenotypes are caused by a combination of high-impact SNPs. Thus, these TOP Lines appear to represent a "natural stacking" of a number of cell wall biosynthesis genes.

Based on the above, we were able to develop high-throughput phenotyping assays that accurately predict cell wall chemistry traits and thus allowed that assessment of hundreds of biomass samples. In combination with the genome-wide association population and DNA marker libraries we were able to link variation at the phenotypic level to specific changes in DNA structure, such as single nucleotide

polymorphisms and single amino acid substitutions in functioning proteins. Many such genes were identified and were subsequently tested and verified using cis- and trans-genic validation assays and genomic selection [see the following two sections].

Genome-wide cis- and trans-regulatory elements. We observed that allelic variants of EPSP synthase-like gene produce biomass with reduced lignin and increased sugar release and ethanol yield from non-pretreated biomass (Fig. 8). GWAS analysis of the *Populus* natural variation population using composition pipeline data from the *Populus* common gardens identified SNPs in two EPSPs paralogs. The SNPs of both paralogs were independently associated with lignin content. EPSP synthases are functionally conserved from prokaryote to eukaryotes and have, until now, been known only to catalyze the sixth step in the shikimate pathway. We have now shown that a novel *Populus* isoform encodes a protein that, unlike ancestral EPSPs, localizes in the nucleus and carries an N-terminus region showing structural homology to the DNA binding helix-turn-helix motif. We have recognized this EPSP gene as a putative transcriptional regulator of phenylpropanoid and flavonoid pathways. We have also shown that over expression of this gene in *Populus* and *Medicago* results in upregulation of genes involved in secondary cell wall and flavonoid biosynthesis, including the canonical secondary cell wall transcriptional master switches—MYB46 and NST1. Rare allelic variants of this locus result in reduced lignin content (as low as 15.7%) and up to 280% increased sugar release and ethanol yield from non-pretreated biomass. These results suggest that an EPSP synthase transcriptional regulator can be used to metabolic engineering plants to lower lignin.

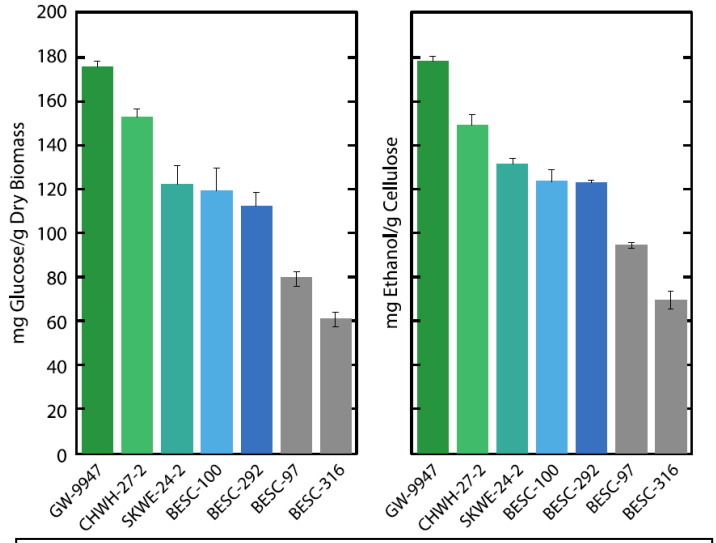

Figure 8. Glucose release and ethanol yield of rare EPSP variants (green and blue) compared to medium and high lignin genotypes (gray).

To test this hypothesis transgenic lines of *Populus deltoides* and natural variants of *P. trichocarpa* were characterized by metabolite analysis using gas chromatography mass spectrometry. The results established metabolomics changes associated with the specific EPSP synthase perturbations in controlling carbon flow through the lignin pathway. The gene deletion on EPSP synthase paralog 14 primarily resulted in leaf profiles elevated in tryptophan and reduced in phenolic metabolites, including coniferin, catechin, quercetin, and salicylic acid and phenolic acids upstream of monolignol production, such as p-coumaric acid, caffeic acid, and ferulic acid. The homologous gene deleted on chromosome 2 resulted in severe depletions of higher-order salicylates in leaves down to 6–23% that of wildtype plants. Metabolomics confirmed the broader impacts beyond the targeted gene of the genetic perturbations. This gene has been licensed to a forage breeding company as a means of reducing lignin content in alfalfa.

We discovered an alternate pathway to higher cellulose content based on GWAS results. This gene PdIQD1 in PdIQD1 RNAi transgenics, enhanced cellulose content and crystallinity while lignin and hemicellulose (xylose and arabinose) content were unaffected. An *in vivo* protein interaction assay approach revealed that PDIQD protein interacts with two *Populus* proteins with homology to *Arabidopsis* kinesin light chain reaction protein. These results combined with earlier co-expression analyses suggest that PdIQD1 participates in stem development and/or cell wall biosynthesis either as a signaling protein to nuclear TFs or as a part of a directional cargo (small protein/metabolite) provision protein complex participating directly in cellulose biosynthesis.

Similarly, through GWAS a HDZipIII TF family member was discovered whose expression leads to higher cellulose content. *Populus* PdHB3-KD transgenic lines had a significant increase in enzymatic glucose release (~50% increase) and fermentative ethanol conversion (<60% increase). An increase in plant height (up to 24%), stem diameter (up to 14%) and cellulose (by ~7%) was observed in transgenic lines. PdHB3 was shown to localize in the nucleus and RT-PCR assays using a panel of ~40 proposed secondary cell wall TFs confirmed its position in transcriptional hierarchy of known cell wall related transcription factors. Our results are the first to link PdHB3, as a TF, with increased cellulose, total biomass, and decreased biomass recalcitrance.

Again, using GWAS analyses, Angustifolia, a TF in *Populus*, contains variation in polyglutamine (polyQ) motifs that affects its function. These allelic variants affect lignin content. Using protoplast assays we demonstrated that longer polyQ motifs prevent nuclear localization of the Angustifolia whereas variants with the exact same sequence but shorter polyQ motifs successfully localize in the nucleus. This is the first report of a gene containing a polyQ variant affecting cell wall chemistry. *Arabidopsis* knock-out lines of the *Populus* gene had significantly reduced lignin content.

We functionally characterized a serine hydroxymethyltransferase (SHMT) from *Populus* and evaluated its potential for developing lignocellulosic feedstocks. SHMT plays an important role in cellular one-carbon pathways. However, little is known about its function in plant cell wall–related processes. Among nine *SHMT* genes in poplar, *SHMT2* was highly expressed in the developing xylem and co-expressed with secondary cell wall biosynthetic genes. In *P. deltoides* transgenic plants overexpressing *PdSHMT2*, the biomass yield and sugar (glucose and xylose) release were increased and lignin content was decreased. Transcriptomics and metabolomics analyses revealed that genes and metabolites related to secondary cell wall biosynthesis were affected by *PdSHMT2* overexpression. Based on the transcription factor binding sites of differentially expressed genes in *PdSHMT2* overexpression lines, a total of 27 hub transcription factors were identified. We proposed a regulatory model of action of *PdSHMT2* with transcriptional master switches of secondary cell wall biosynthesis. On a phenotypic level, *PdSHMT2* lines showed decreased lignin content (7.6%), increased glucose release (12%) and increased xylose release (31%).

Five transgenic lines, belonging to *Populus* IQD, HDZipIII, Extensin, Aux/IAA and ARF gene families, have been identified through studies of genes that positively affect cellulose synthesis. Specifically, 1) a putative signaling protein in the IQD family with a likely functional role in cellulose deposition and secondary wall formation, 2) an HDZip III gene family transcription factor that affects cellulose content and radial and

vertical stem growth, 3) EXT, extension, genes that impact biomass formation and cell wall composition and 4) Aux/IAA and ARF, auxin transcription factor genes that impact biomass formation. We also determined that manipulation of expression of *Populus* PdUGPase results in reduced growth and altered cell wall composition.

For each of the high-quality predictions based on GWAS results, we were able to validate that the identified candidate genet did in fact control the associated phenotype. We were also able in many cases to identify the causal substitution at the amino acid level and thus understand the mechanisms leading to the change in phenotype.

Targeted quantitative trait nucleotide (QTN) stacking and genomic selection. Based on RNA-seq analysis of leaf and xylan tissues in the GWAS population, we established an eQTN mapping pipeline to conduct global prediction of regulatory elements for >20,000 genes from both leaf and xylem transcriptomes. Of these, ~10,000 were shared between the two tissues and ~5,000 exhibited tissue-specific expression. We demonstrated numerous instances where regulatory elements mapped to the same genomic position across tissues, supporting our predictions of master regulatory loci (Fig. 9). Additionally, three-layer transcriptional regulatory hierarchies were constructed to identify transcriptional master regulators that target more than 100 cell wall-relevant genes. From this analysis, two candidate genes (*XBAT* and *GOLGIN*) have been selected for cis-genic CRISPR-Cas 9 validation.

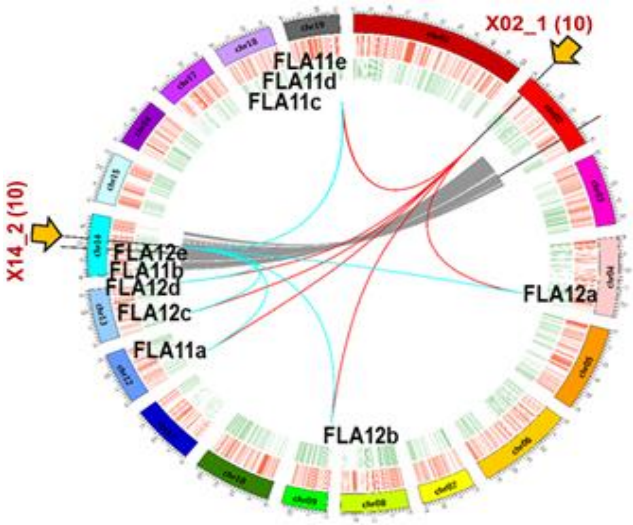

Figure 9. Mapping novel transcriptional regulators on chromosome 2 and 14 of ten members of the Fasciclin gene family in *Populus*. This gene family has been widely implicated in cell wall biosynthesis in numerous plant genera, but this is the first report of these specific transcriptional regulators targeting multiple members of the family.

Our previous studies have demonstrated that *P. trichocarpa* PtrEPSP-TF, an isoform of 5-enolpyruvylshikimate 3-phosphate (EPSP) synthase, has transcriptional activity and regulates phenylpropanoid biosynthesis in *Populus*. In this study, we report the identification of a SNP of *PtrEPSP-TF* that defines its functionality. *Populus* natural variants carrying this SNP have been shown to have reduced lignin content. We have also demonstrated that the SNP-induced substitution of 142nd amino acid (PtrEPSP-TF^{D142E}) dramatically impairs the DNA binding and transcriptional activity of PtrEPSP-TF. When introduced to a monocot species rice (*Oryza sativa*), in which an EPSP synthase isoform with the DNA-binding helix-turn-helix motif is absent, PtrEPSP-TF, but not PtrEPSP-TF^{D142E} variant, activated genes in the phenylpropanoid pathway. More importantly, heterologous expression of *PtrEPSP-TF* uncovered five new transcriptional regulators of phenylpropanoid biosynthesis in rice (Fig. 10). Collectively, this study

identifies the key amino acid required for PtrEPSP-TF functionality and provides a strategy to uncover new transcriptional regulators in phenylpropanoid biosynthesis.

Finally, based on GWAs results and validation cis- and trans-genic assays we have developed a series of genomic selection algorithms that predict progeny performance of selected cross in P. trichocarpa. These algorithms are based on and target increased biomass production, improved cell wall chemistry traits (i.e., lignin) and enhanced disease resistance. Crosses have been made and progeny are now currently growing in field trials and have completed their first year of growth. We will not know that final outcome and success of this prediction until after the end of the fourth growing season (approximately fall 2022).

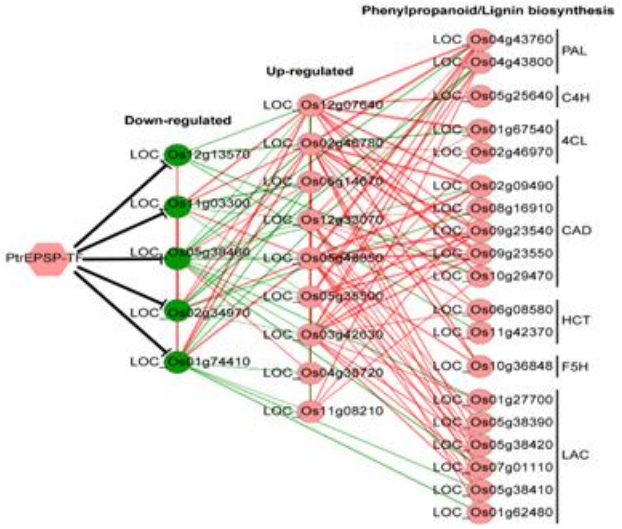

Figure 10. Model of PtrEPSP-TF associated transcription factor (TF) network regulating phenylpropanoid biosynthesis in rice.

Summary

The development of genomic resources in poplar in conjunction with sustained common gardens of sequenced and phenotyped genotypic individuals has led to many genomic-based insights into plant cell wall formation. This has included a number of TFs for regulation of plant cell wall composition and structure. We have also been able to identify potential functions of a number of biosynthetic genes. Not surprisingly, these plant-cell-wall genes have been shown to have functional homology beyond poplar into other plant species. Finally, we have shown and validated the possibility of hybrid functions – such as the EPSP genes which has both biosynthetic and regulatory functions. This work has led to the development of improved genetic selection algorithms for improved breeding strategies.

References

Quantitative trait loci study

1. Zhang, J., Yang, Y., Zheng, K., Xie, M., Feng, K., Jawdy, S.S., Gunter, L.E., Ranjan, P., Singan, V.R., Engle, N., Lindquist, E., Barry, K., Schmutz, J., Zhao, N., Tschaplinski, T.J., LeBoldus, J., Tuskan, G.A., Chen, J-G., and Muchero, W., "Genome-wide association studies and expression-based quantitative trait loci analyses reveal roles of HCT2 in caffeoylquinic acid biosynthesis and its regulation by defense-responsive transcription factors in *Populus*", 220(2):502-516, *New Phytologist*, 220(2):502-516, 2018. doi: 10.1111/nph.15297
2. Bdeir, R., Muchero, W., Yordanov, Y., Tuskan, G.A., Busov, V., and Gailing, O., "Genome-wide association studies of bark texture in *Populus trichocarpa*", *Tree Genetics and Genomes*, 15(14), 2019. doi: 10.1007/s11295-019-1320-2

3. Yin, T.M., Zhang, X.Y., Gunter, L.E., Li, S.X., Wullschleger, S.D., Huang, M.R., Tuskan, G.A., "Microsatellite primer resource for *Populus* developed from the mapped sequence scaffolds of the Nisqually-1 genome", *New Phytologist*, 181:498-503, 2018. doi: 10.1111/j.1469-8137.2008.02663.x
4. Kang, B.G., Osburn, Lori, Kopsell, Dean, Tuskan, G.A., Cheng, Z.M., Micropropagation of *Populus trichocarpa* 'Nisqually-1': The genotype deriving the *Populus* reference genome, *Journal of Plant Biotechnology*, 99(3):251-257, 2009. doi: 10.1007/s11240-009-9596-9
5. Muchero, W., Sondreli, K.L., Chen, J-G., Urbanowicz, B.R., Zhang, J., Singal, V., Yang, Y., Brueggeman, R.S., Franco-Coronado, J., Abraham, N., Yang, J-Y., Moremen, K.W., Weisberg, A.J., Chang, J.H., Lindquist, E., Barry, K., Ranjan, P., Jawdy, S., Schmutz, J., Tuskan, G.A., LeBoldus, J.M., "Association mapping, transcriptomics, and transient expression identify candidate genes mediating plant-pathogen interactions in a species of forest tree", *PNAS*, 115:45(11573-11578), 2018. doi: 10.1073/pnas.1804428115

Cell wall chemistry phenotyping

1. Payyavula, R.S., Tschaplinski, T.J., Jawdy, S.S., Sykes, R.W., Tuskan, G.A., Kalluri, U.C., "Metabolic profiling reveals altered sugar and secondary metabolism in response to UGPase overexpression in *Populus*", *BMC Plant Biology*, 14(265), 2014. doi: 10.1186/s12870-014-0265-8
2. R.A., Yang, X., Winkeler, K., Collins, C., Mohanty, S.S., Ryno, D., Tan, L., Gelineo-Albersheim, I., Hunt, K., Sykes, R.W., Turner, G.B., Ziebell, A., Davis, M.F., Decker, S.R., Hahn, M.G., Mohnen, D., "Working towards recalcitrance mechanisms: increased xylan and homogalacturonan production by overexpression of GAlactUronosylTransferase12 (GAUT12) causes increased recalcitrance and decreased growth in *Populus*", *Biotechnology for Biofuels*, 11(9), 2018. doi: 10.1186/s13068-017-1002-y
3. Hao, Z., Avci, U., Tan, L., Zhu, X., Glushka, J., Pattathil, S., Eberhard, S., Scholes, T., Rothstein, Lukowitz, W., Orlando, R., Hahn, M.G., Mohnen, D., "Loss of Arabidopsis GAUT12/IRX8 causes anther indehiscence and leads to reduced G lignin associated with altered matrix polysaccharide deposition", *Frontiers in Plant Science*, 5:357, 2014.
4. Hao, Z., Mohnen, D., "A review of xylan and lignin biosynthesis: foundation for studying Arabidopsis irregular xylem mutants with pleiotropic phenotypes", *Critical Reviews in Biochemistry and Molecular Biology*, 49(3):212-214, 2014. doi: 10.3109/10409238.2014.889651
5. Urbanowicz, B.R., Pena, M.J., Moniz, H.A., Moremen, K.W., York, W.S., "Two Arabidopsis Proteins Synthesize Acetylated Xylan *in vitro*", *The Plant Journal*, 80(2):197-206, 2014. doi: 10.1111/tpj.12643
6. Biswal, A.K., Hao, Z., Pattathil, S., Yang, X., Winkeler, K., Collins, C., Mohanty, S.S., Richardson, E.A., Gelineo-Albersheim, I., Hunt, K., Ryno, D., Sykes, R.W., Turner, G.B., Ziebell, A., Gjersing, E., Lukowitz, W., Davis, M.F., Decker, S.R., Hahn, M.G., Mohnen, D., "Downregulation of GAUT12 in *Populus* deltoides by RNA silencing results in reduced recalcitrance, increased growth and reduced xylan and pectin in a woody biofuel feedstock", *Biotechnology for Biofuels*, 8(41), 2015. doi: 10.1186/s13068-015-0204-4

7. Biswal, A.K., Atmodjo, M.A., Pattathil, S., Amos, R.A., Yang, X., Winkeler, K., Collins, C., Mohanty, S.S., Ryno, D., Tan, L., Gelineo-Albersheim, I., Hunt, K., Sykes, R.W., Turner, G.B., Ziebell, A., Davis, M.F., Decker, S.R., Hahn, M.G., Mohnen, D., "Working towards recalcitrance mechanisms: increased xylan and homogalacturonan production by overexpression of GAlactUronosylTransferase12 (GAUT12) causes increased recalcitrance and decreased growth in *Populus*", *Biotechnology for Biofuels*, 11(9), 2018. doi: 10.1186/s13068-017-1002-y
8. Tan, L., Eberhard, S., Pattathil, S., Warder, C., Glushka, J., Yuan, C., Hao, Z., Zhu, X., Avci, U., Miller, J.S., Baldwin, D., Pham, C., Orlando, R., Darvill, A., Hahn, M.G., Kieliszewski, M.J., Mohnen, D., An Arabidopsis cell wall proteoglycan consists of pectin and arabinoxylan covalently linked to an arabinogalactan protein, *The Plant Cell*, 25(1):270-287, 2012. doi: 10.1105/tpc.112.107334
9. Amos, R.A., and Mohnen, D., "Critical review of plant cell wall matrix polysaccharide glycosyltransferase activities verified by heterologous protein expression" *Frontiers in Plant Science*, 10:915, 2019. doi: 10.3389/fpls.2019.00915
10. Amos, R.A., Pattathil, S., Yang, J-Y., Atmodjo, M.A., Urbanowicz, B.R., Moremen, K.W., Mohnen, D., "A two-phase model for the non-processive biosynthesis of homogalacturonan polysaccharides by the GAUT1:GAUT7 complex", *Journal of Biological Chemistry*, 2018. doi: 10.1074/jbc.RA118.004463
11. Atmodjo, M.A., Hao, Z., Mohnen, D., "Evolving views of pectin biosynthesis", *Annual Review of Plant Biology*, 64:747-779, 2013. doi: 10.1146/annurev-arplant-042811-105534

Genome-wide cis- and trans-regulatory elements

1. Xie, M., Zhang, J., Singan, V.R., McGranahan, M.J., LaFayette, P.R., Jawdy, S.S., Engle, N., Doeppke, C., Tschaplinski, T.J., Davis, M.F., Lindquist, E., Barry, K., Schmutz, J., Parrott, W.A., Chen, F., Tuskan, G.A., Chen, Jin-Gui, and Muchero, W., "Identification of functional single nucleotide polymorphism of *Populus trichocarpa* PtrEPSP-TF and determination of its transcriptional effect", *Plant Direct*, 2019. doi: 10.1002/pld3.178
2. Xie, M., Muchero, W., Bryan, A.C., Yee, K., Guo, H-B., Zhang, J., Tschaplinski, T.J., Singan, V.R., Lindquist, E., Payyavula, R.S., Barros-Rios, J., Dixon, R., Engle, N., Sykes, R.W., Davis, M., Jawdy, S.S., Gunter, L.E., Thompson, O., DiFazio, S.P., Evans, M., Winkeler, K., Collins, C., Schmutz, J., Guo, H., Kalluri, U., Rodriguez, M., Feng, K., Chen, J-G., Tuskan, G.A., "A 5-enolpyruvylshikimate 3-phosphate synthase functions as a transcriptional repressor in *Populus*", *Plant Cell*, 30:1645-1660, 2018. doi: 10.1105/tpc.18.00168
3. Zhao, Q., Zeng, Y., Yin, Y., Pu, Y., Jackson, L.A., Engle, N.L., Martin, M.Z., Tschaplinski, T.J., Ding, S-Y., Ragauskas, A.J. and Dixon, R.A., "Pinoresinol reductase 1 impacts lignin distribution during secondary cell wall biosynthesis in Arabidopsis", *Phytochemistry*, 112:170-178, 2014. doi: 10.1016/j.phytochem.2014.07.008
4. Muchero, W., Guo, J., DiFazio, S.P., Chen, J-G., Ranjan, P., Slavov, G.T., Gunter, L.E., Jawdy, S., Bryan, A.C., Sykes, R., Ziebell, A., Klapste, J., Porth, I., Skyba, O., Unda, F., El-Kassaby, Y.A., Douglas, C.J., Mansfield, S.D., Martin, J., Schackwitz, W., Evans, L.M., Czarnecki, O., Tuskan, G.A.,

"High-resolution genetic mapping of allelic variants associated with cell wall chemistry in *Populus*", *BMC Genomics*, 16(24), 2015. doi: 10.1186/s12864-015-1215-z

5. Bryan, A.C., Zhang, J., Guo, J., Ranjan, P., Singan, V., Barry, K., Schmutz, J., Weighill, D., Jacobson, D., Jawdy, S., Tuskan, G.A., Chen, J-C., Muchero, W., "A variable polyglutamine repeat affects subcellular localization and regulatory activity of a *Populus* ANGUSTIFOLIA protein", *G3* 8(8):2631-2641, 2018. doi: 10.1534/g3.118.200188
6. Zhang, J., Li, M., Bryan, A.C., Yoo, C.G., Rottmann, W., Winkeler, K.A., Collins, C.M., Singan, V., Lindquist, E.A., Jawdy, S.S., Gunter, L.E., Engle, N.L., Yang, X., Barry, K., Tschaplinski, T.J., Schmutz, J., Pu, Y., Ragauskas, A.J., Tuskan, G.A., Muchero, W., Chen, J-G., "Overexpression of a serine hydroxymethyltransferase increases biomass production and reduces recalcitrance in the bioenergy crop *Populus*", *Sustainable Energy and Fuels*", 3(1):195-207, 2018. doi: 10.1039/c8se00471d
7. Badmi, R., Payyavula, R.S., Bali, G., Guo, H-B., Jawdy, S.S., Gunter, L., Yang, X., Winkeler, K., Collins, C., Rottmann, W.H., Yee, K., Rodriguez Jr., M., Sykes, R., Decker, S.R., Davis, M.F., Ragauskas, A.J., Tuskan, G., Kalluri, U.C., "A new calmodulin-binding protein expresses in the context of secondary cell wall biosynthesis and impacts biomass properties in *Populus*", *Frontiers in Plant Science*, 9, 2018. doi: 10.3389/fpls.2018.01669
8. Payyavula, R.S., Tschaplinski, T.J., Jawdy, S.S., Sykes, R.W., Tuskan, G.A., Kalluri, U.C., "Metabolic profiling reveals altered sugar and secondary metabolism in response to UGPase overexpression in *Populus*", *BMC Plant Biology*, 14(265), 2014. doi: 10.1186/s12870-014-0265-8

Targeted quantitative trait nucleotide (QTN) stacking and genomic selection

1. Joubert, W., Nance, J., Weighill, D., Jacobson, D., "Parallel accelerated vector similarity calculations for genomics applications", *Parallel Computing*, 75:130-145, 2018. doi: 10.1016/j.parco.2018.03.009
2. Furches, A., Kainer, D., Weighill, D., Large, A., Jones, P., Walker, A.M., Romero, J., Gazolla, J.G.F.M., Joubert, W., Shah, M., Streich, J., Ranjan, P., Schmutz, J., Sreedasyam, A., Macaya-Sanz, D., Zhao, N., Martin, M.Z., Rao, X., Dixon, R.A., DiFazio, S., Tschaplinski, T.J., Chen, J-G., Tuskan, G.A., and Jacobson, D.A., "Finding new cell wall regulatory genes in *Populus trichocarpa* using multiple lines of evidence", *Frontiers in Plant Science*, 2019. doi: 10.3389/fpls.2019.01249
3. Also see Xie et al., 2019 and Xie et al 2018 cited above.