Insights of biomass recalcitrance in *Populus trichocarpa* natural variants for biomass conversion

Chang Geun Yoo,[a,b] Yongil Yang,[a] Yunqiao Pu,*[a] Xianzhi Meng,[c] Wellington Muchero,[a] Kelsey L. Yee,[a] Olivia A. Thompson,[a] Miguel Rodriguez Jr,[a] Garima Bali,[c] Nancy L. Engle,[a] Erika Lindquist,[d] Vasanth Singan,[d] Jeremy Schmutz,[d,e] Stephen P. DiFazio,[f] Timothy J. Tschaplinski,[a] Gerald A. Tuskan,[a] Jin-Gui Chen,[a] Brian Davison,[a] and Arthur J. Ragauskas*[a,b,g]

Notice: This manuscript has been authored by UT-Battelle, LLC under Contract No. DE-AC05-00OR22725 with the U.S. Department of Energy. The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes. The Department of Energy will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan (http://energy.gov/downloads/doe-public-access-plan).

Populus has been investigated as a promising biomass feedstock for alternative fuels and chemicals. Physicochemical characteristics and genomic information of biomass feedstocks are among the essential informaiton that can help not only advance our understanding of biomass recalcitrane but also with its efficient utilization. Herein, recalcitrance of *Populus* natural variants was elucidated in three aspects: (1) sugar release, (2) physicochemical properties, and (3) relative variation of gene expression within poplar natural variants. Sugar release performance of *Populus* natural variants was evaluated with their correlation with biomass recalcitrance-related characteristics. Among the physicochemical properties of poplar, lignin content, lignin molecular weight, lignin S/G ratio, and cellulose accessibility were found to correlate with sugar release. The results demonstrted that lignin content was negatively correlated with sugar release, whereas lignin molecular weight, lignin S/G ratio, and cellulose accessibility were positively associated with poplar sugar release. The trend of differential gene expression of each variant also supports the characterization results and their effects on biomass conversion.

Introduction

Lignocellulosic biomass is a biopolymer mainly composed of polysaccharides (cellulose and hemicellulose) and aromatic polymer (lignin). Sustainability, environmentally-friendly characteristic, and infrastructure-compatibility of the biomass make it a promising feedstock for many applications.[1, 2] Diverse biomass feedstocks, including woody plants, herbaceous plants, and agricultural residues have been tested for producing many bio-based fuels and chemicals. *Populus* is one of the key biofuel feedstock candidates growing in the United States.[3] It has been tested as a model species because of its relatively rapid growth, ease of propagation and genetic manipulation, and a range of available genetic tools.[4, 5] *Populus* is amenable for selective breeding for large-scale plantation establishment and has diverse phenotypes, and its underlying traits can be readily mapped using interspecific hybrids.[4] In addition, its short-rotation and purpose-grown properties can be benefits and economic advantages as a "living inventory" of available biomass.[6] Recently, utilization of *Populus* was expanded to diverse approaches including production of bio-ethanol, furan-based chemicals, and cellulose nanofibers instead of conventional pulp and paper application.[7-10] For instance, the aerogels with cellulose nanofibers from poplar tree showed outstanding physical and mechanical properties such as high porosity, ultra-low density, and high flexibility.[7]

Biomass recalcitrance to chemical or biological conversion is a factor to be overcome for effective utilization of biomass. Natural characteristics of biomass from its structural heterogeneity and complexity of cell-wall constituents were directly and/or indirectly related to the biomass recalcitrance.[11, 12] There have been many efforts to overcome the biomass recalcitrance in different ways such as developing catalysts, applying novel solvents, improving and optimizing process, modifying feedstocks, and others.[13-17] In particular, investigation of the physicochemical properties of biomass feedstock not only elucidate what the biomass recalcitrance factors are, but it also provides clues on how to overcome the physical and chemical barriers for biomass conversion. Diverse approaches have been applied to reveal the underlying basis of biomass recalcitrance with its characteristics. Dutta et al. discussed the influence of rigid cross-linked biomass structure including lignin network and crystalline region of cellulose on chemical and biological decompositions of biomass.[13, 15] Studer et al. reported the effects of lignin content and S/G ratio of *Populus trichocarpa* on its sugar release.[18] The influences of cellulose properties, such as crystallinity and degree of polymerization, were also discussed in the previous studies.[19, 20] Foston and his colleagues analyzed molecular weight and crystallinity of cellulose, monolignol composition, morphology and spatial distribution of biomass components to compare the recalcitrance of *Populus* tension wood and that of normal wood.[21] Besides these properties of biomass, Meng and Ragauskas discussed cellulose accessibility as a factor for biological conversion of biomass.[22] The influences of the aforementioned biomass properties on biomass recalcitrance alone and/or in combination were discussed; however, some of their impacts conflicted depending on the species, types and conditions of pretreatment, and other factors. Also, most of studies were conducted with significantly modified biomass after pretreatment or genetic modification.

Genome-wide transcriptome analysis using next generation sequencing technique has been used to show the gene regulation network for plants.[23] The differential gene expression analysis based on the transcriptome has been

applied to validate different phenotypes of various organs and tissues of plants in response to varying osmotic environment and microbial infection in *Populus*.[24-27] The genes related to lignin biosynthesis and biomass content in sugar cane, switchgrass, *Populus*, and *Eucalyptus* were also identified based on genome wide transcriptome analyses.[28-30] Recently, transcriptome analysis of natural variants was applied to identify the association between single nucleotide polymorphism (SNP) variation in the *Populus* genome and growth traits.[31] However, the analysis of the transcriptome related to biomass recalcitrance properties and sugar release has not been reported yet.

In this study, *Populus trichocarpa* was evaluated as a feedstock for biofuel production by investigating its recalcitrance in combination with transcriptome analysis of natural poplar variants. Recalcitrance of each *P. trichocarpa* natural variant was assessed based on sugar release. The sugar release of *Populus* variants were assessed in terms of diverse physicochemical properties, including chemical composition, degree of polymerization of cellulose and hemicellulose, molecular weight of lignin, cellulose crystallinity, structural information of lignin, and cellulose accessibility. The major biomass recalcitrance-related properties of *Populus* natural variants were also investigated in a correlation with transcriptome RNA-seq analysis.

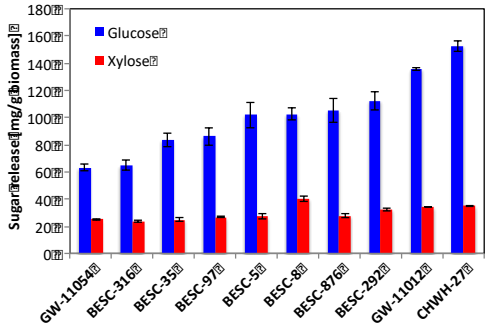


Fig. 1 Glucose and xylose release of *P. trichocarpa* natural variants.

Results and Discussion

To determine the biomass recalcitrance-related properties of *Populus*, diverse characterization methods using HPLC, GPC, 2D ^{13}C-^{1}H HSQC NMR, and solid-state NMR analyses were conducted with selected *Populus* natural variants grown at the same field location (Clatskanie, OR) for 5 years. The poplar natural variants were selected to include a wide range of sugar release as possible. The effects of physiochemical properties on sugar release performance were evaluated with their correlations. Global transcriptome of these variants was also conducted to identify the correlation variation of typical gene expression and physicochemical properties.

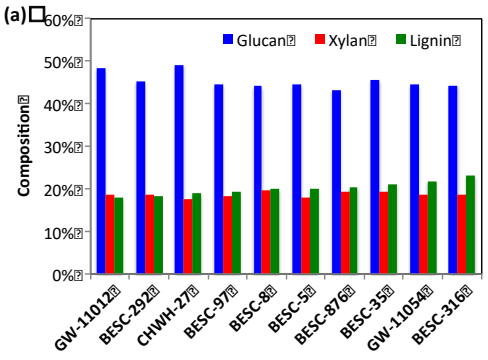
Sugar release of *Populus* natural variants

Cellulose and hemicellulose in biomass are composed of sugars including glucose, xylose, galactose, arabinose, and mannose. The sugars are important intermediates for many biomass-based fuels and chemicals such as bio-ethanol, furan-based chemical, and others. In this study, sugar release was used as an index to evaluate the ease of conversion of each *P. trichocarpa* genotype. Fig. 1 presents glucose and xylose release of each *P. trichocarpa* natural variant. The sugar release results of these natural variants were notably different, ranging from 63 – 153 mg/g biomass for glucose and 24 – 40 mg/g biomass for xylose, respectively. In particular, CHWH-27 and GW-1101 showed more than twice the glucose release compared to GW-1105, which had the lowest value, indicating that selection of the appropriate biomass feedstock is as important as modification of the feedstock and development of conversion methods. Glucose release and xylose release of the *P. trichocarpa* variants showed a positive correlation (R^2=0.50, Pearson coefficient=0.705, *p*-value=0.023) (Fig. S1 in ESI). The results suggest that removal of hemicellulose, which is mainly composed of xylan in *Populus sp.*, may enhance the cellulose accessibility, likely due to its physically surrounding cellulose; therefore, the variants with higher xylose release are likely to have enhanced glucose release.

Physicochemical properties of *Populus* natural variants

Physicochemical properties of each *P. trichocarpa* natural variant can directly and/or indirectly provide information about the biomass recalcitrance. Herein, sugar release of *P. trichocarpa* natural variants was linked to diverse characteristics, including chemical composition, degree of polymerization (DP), cellulose crystallinity, lignin molecular weight, structural information of lignin, and cellulose accessibility.

Chemical composition. Chemical composition is a key biomass property, because it not only provides theoretical maximum sugar release, but it also helps to evaluate the biomass recalcitrance. Fig. 2a presents three major components (glucan, xylan and lignin) in *P. trichocarpa*. Glucan and xylan contents of *P. trichocarpa* natural variants ranged from 43.3% (BESC-876) to 49.2% (CHWH-27) and 17.5% (CHWH-27) to 19.6% (BESC-8), respectively. Lignin content also varied from 17.8% (GW-11012) to 23.2% (BESC-316).



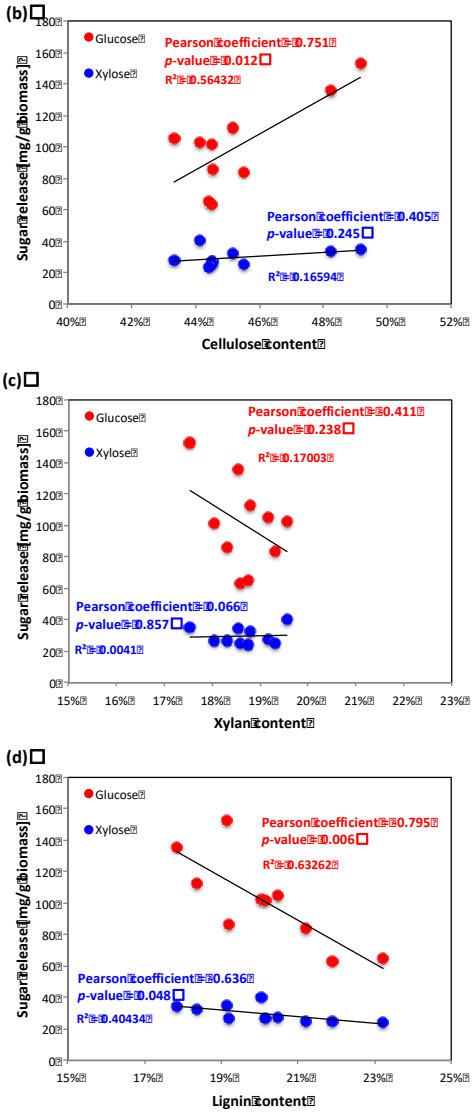


Fig. 2 (a) Chemical composition of *P. trichocarpa* natural variants and (b-d) correlations between each component (cellulose, hemicellulose, and lignin) and sugar release.

Effects of chemical composition on the aforementioned glucose and xylose release were investigated by evaluating the correlation between the sugar release and content of three major components: glucan, xylan, and lignin (Figs. 2b-d). Cellulose (glucan) content of the natural variants showed a correlation with their glucose release (R^2=0.56, Pearson coefficient=0.751, *p*-value=0.012), whereas the correlation with xylose release was not apparent (R^2=0.17, Pearson coefficient=0.405, *p*-value=0.245) (Fig. 2b). Higher glucose release was expected from the variant with lower xylan content, because xylan surrounds cellulose with other hemicelluloses and lignin, thus it physically inhibits the access of enzymes to cellulose fibers.[32, 33] However, a strong correlation was not observed between xylan content and sugar release in this study (Fig. 2c). Also, the xylan content did not significantly affect xylose release.

In addition to the effects of cellulose content, lignin content played a crucial role for both glucose and xylose release.[18] The results showed that lignin had a negative correlation with both glucose (R^2=0.63, Pearson coefficient=-0.795, *p*-value=0.006) and xylose release (R^2=0.40, Pearson coefficient=-0.636, *p*-value=0.048) (Fig. 2d). Recalcitrant effects of lignin, (1) as a physical barrier to the carbohydrate surface and (2) through the non-productive binding to enzymes, have been verified in the previous studies.[34-36] These inhibitory effects can explain these negative correlations of lignin with the sugar release in this study.

Characteristics of cellulose and hemicellulose. Degree of polymerization of cellulose and hemicellulose are other factors possibly affecting biomass recalcitrance. Fig. 3 shows weight-average degree of polymerization (DP_w) and number-average degree of polymerization (DP_n) of cellulose/hemicellulose and the correlation between cellulose DP_w and sugar release. In Fig. 3a, the majority of *P. trichocarpa* natural variants had 4100 – 5400 of cellulose DP_w, whereas some variants had significantly higher DP_w (8844 for BSEC-316) or lower DP_w (2408 and 2505 for CHWH-27 and GW-11012, respectively). The DP_n also varied from 216 to 881, with the similar genotype ranks to that observed for cellulose DP_w. Polydispersity index (PDI) is an indicator for measuring the distribution of molecular mass. Except for BESC-316, samples with relatively low cellulose DP_w (CHWH-27, GW-11012, and BESC-97) had higher PDI than the other genotypes with high cellulose DP_w. High cellulose DP_w samples generally had more uniform size of cellulose in *P. trichocarpa*. Degree of polymerization for hemicellulose in the *P. trichocarpa* samples was also investigated as presented in Fig. 3b. Overall, the variants with high DP_w of hemicellulose also had high DP_n. For example, GW-11054 had the lowest DP_w (328) and DP_n (179) of hemicellulose and BESC-35 had the highest (429 and 299 of DP_w and DP_n, respectively) among the ten natural variants. Also, the variants with high DP of hemicellulose had relatively low PDI.

The correlations between DP of cellulose/hemicellulose and the sugar release are presented in Fig. 3c and d. The glucose release increased as cellulose DP_w decreased. Cellulose DP is associated with the content of the reducing ends, thus the decreasing of cellulose DP provides more activity for exoglucanases and can be interpreted as reduced biomass recalcitrance.[37] However, both cellulose DP_w and hemicellulose DP_w of the natural variants were not significantly correlated with glucose release (Figs. 3c and d).

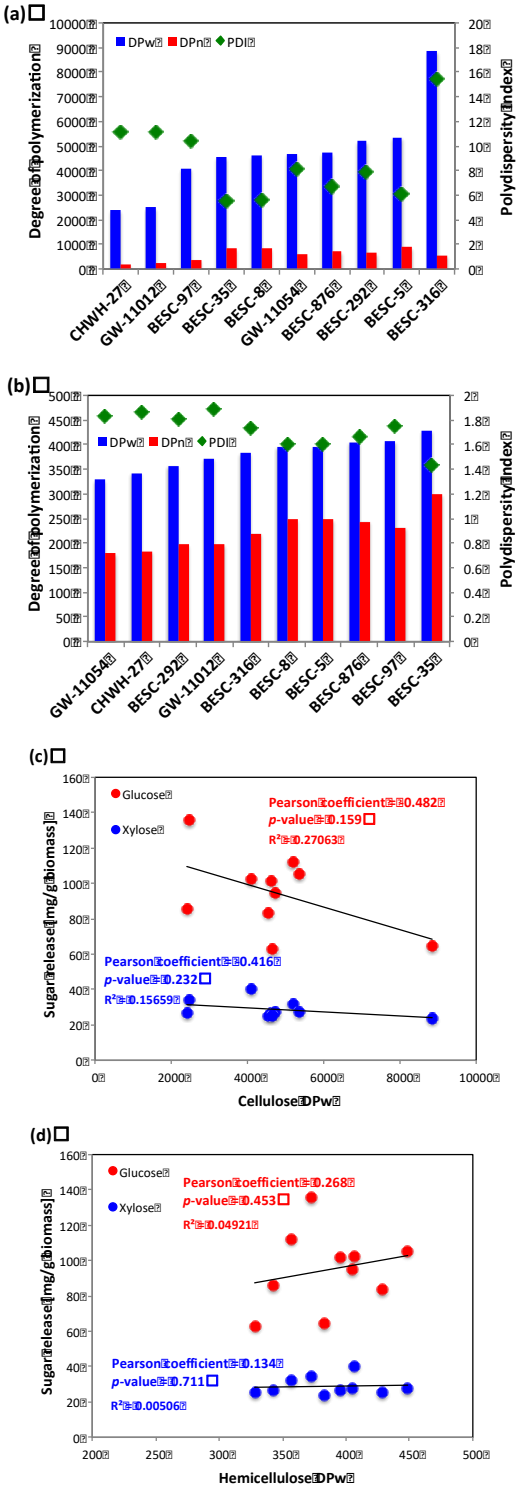


Fig. 3 Degree of polymerization of (a) cellulose and (b) hemicellulose in *P. trichocarpa* natural variants and their correlation with sugar release (c and d).

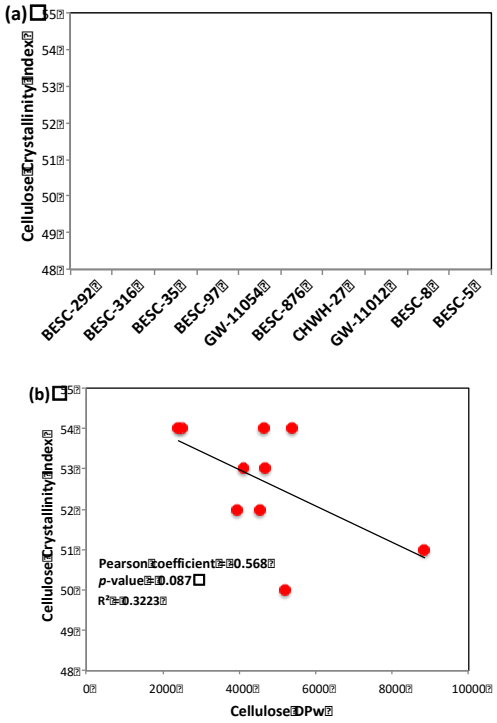


Fig. 4 (a) Crystallinity of cellulose and (b) its correlation with cellulose DP_w.

Crystallinity of cellulose is another characteristic of plant cell walls. Fig. 4a presents crystallinity index (CrI) of the natural variants ranging from 50 to 54. It is well known that hydrolysis of crystalline cellulose is more difficult than that of amorphous cellulose.[19] However, there was no significant correlation observed between cellulose CrI of *P. trichocarpa* and its sugar release (Fig. S2 in ESI). This result is similar to that reported previously;[38] thus cellulose crystallinity as an index of biomass recalcitrance is still open to debate. Still, *P. trichocarpa* variants with lower cellulose DP had more crystalline cellulose in the cell walls (Fig. 4b).

Characteristics of lignin. Physicochemical properties of lignin including molecular weight and structural information are representative factors to understand biomass characteristics. Fig. 5a presents lignin molecular weights and PDI for *P. trichocarpa* natural variants. The weight-average molecular weight (M_w) and number-average molecular weight (M_n) of the natural variants ranged from 9,700 (BESC-316) to 11,700 (BESC-97) and from 2,860 (BESC-316) to 4,370 (BESC-97), respectively, which are similar to the results of other poplar natural variants in previous study.[38] The M_w was positively correlated with glucose release (R^2=0.41, Pearson coefficient=0.644, *p*-value=0.044) (Fig. 5b), suggesting that the variants with higher M_w might be converted more easily. Meng et al. reported that poplar with lower lignin molecular weights had higher glucose release.[38] One possible reason for this opposite result may be due to biomass pretreatment, since the sugar release in the study by Meng et al. was with pretreated poplar, whereas the sugar release in this study was from untreated biomass. In

contrast, M_n and PDI of lignin were not significantly correlated with sugar release.

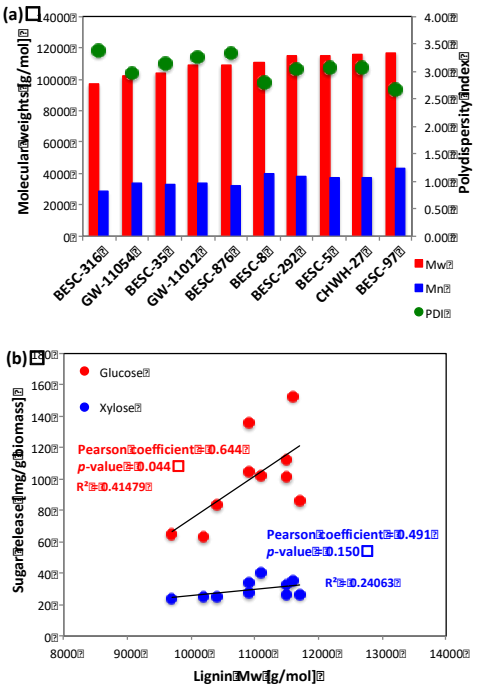


Fig. 5 Lignin molecular weights and PDI of *P. trichocarpa* natural variants and correlation of M_w with sugar release.

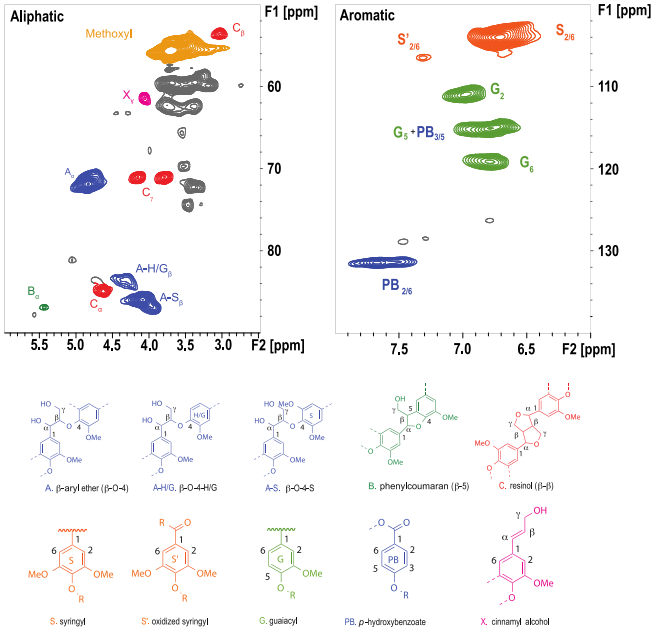


Fig. 6 Aliphatic and aromatic regions of 2D HSQC NMR spectra of a *P. trichocarpa* natural variant (BESC-8).

In addition to lignin content and molecular weights, structural information of lignin, such as lignin S/G ratio, contents of hydroxycinnamates, and contents of lignin inter-unit linkages are important characteristics of biomass. Nuclear magnetic resonance (NMR) analysis was conducted to measure the structural information of lignin in each *P. trichocarpa* natural variant. Fig. 6 presents aliphatic and aromatic regions of *P. trichocarpa variant* (BESC-8). Lignin subunits (syringyl (S) and guaiacyl (G) units), hydroxycinnamates (*p*-hydroxybenzoate (PB)), and dominant lignin inter-unit linkages (β-aryl ether (β-*O*-4), phenylcoumaran (β-5), and resinols (β–β)) were identified and semi-quantitatively analyzed. The semi-quantitative analysis for S, G, and PB units was conducted by integrating the peaks of $S_{2/6}$, α-oxidized $S_{2/6}$, G_2, and $PB_{2/6}$ at δ_C/δ_H 103.9/6.70, δ_C/δ_H 106.5/7.31, δ_C/δ_H 110.9/6.98, and δ_C/δ_H 131.4/7.67, respectively. For the lignin inter-unit linkages, α position of β-*O*-4, β-5, and β–β at δ_C/δ_H 71.9/4.87, δ_C/δ_H 86.9/5.47, and δ_C/δ_H 84.9/4.65 were used. Table 1 presents the information for lignin subunits, hydroxycinnamates, inter-unit linkages of the *P. trichocarpa* natural variants. BESC-316 had the lowest S/G ratio (1.67) and CHWH-27 had the highest (3.88). Lignin S/G ratio has been considered as an indicator to evaluate biomass recalcitrance.[18, 37] To verify the effect of lignin S/G ratio on biomass recalcitrance, the correlation between lignin S/G ratio and sugar release were determined (Fig. 7a). The variants with higher lignin S/G ratio had greater glucose release, and lignin S/G ratio was significantly correlated with glucose release (R^2=0.52, Pearson coefficient=0.722, *p*-value=0.018). Studer et al. suggested that the sugar release was related to both lignin content and lignin S/G ratio.[18] They reported that both glucose and xylose release were correlated with lignin S/G ratio. High sugar release of the *P. trichocarpa* with high lignin S/G ratio was explained by the greater linearity of S-rich lignin with less cross-linking compared to G-rich lignin. A positive correlation (R^2=0.38, Pearson coefficient=0.613, *p*-value=0.059) was observed between lignin S/G ratio and M_w in this study (Fig. 7b). This observation is inconsistent with the previous studies which explained the lower molecular weight of lignin with the higher occurrence of β–β linkage in S-rich lignin.[18, 39] However, the β–β linkage was not the dominant lignin inter-unit linkage in the *P. trichocarpa* natural variants; therefore, the positive correlation between lignin S/G ratio and M_w can be explained by the higher molecular weight of S unit with one more methoxyl group compared to that of G unit. The PB unit, another monolignol conjugate, is involved in acylation of γ-hydroxyl of lignin side-chains likely due to the *p*-hydroxybenzoyl-CoA enzyme involvement in lignification in poplar.[40] It mostly exists as a free-phenolic pendant on lignin and can easily be removed as a single compound. Herein, the PB content in the natural variants ranged from 4 to 20% of the total lignin subunits (=S+G) and was negatively correlated with lignin S/G ratio (R^2=0.44, Pearson coefficient=-0.663, *p*-value=0.037) (Fig. 7c). Also, the PB content was negatively correlated with glucose release (R^2=0.39, Pearson coefficient=-0.628, *p*-value=0.052) (Fig. 7d).

Table 1. Quantitative information of lignin subunit, hydroxycinnamates, and inter-unit linkages in *P. trichocarpa* natural variants.

Name	Content of aromatics [%]	Content of lignin inter-unit linkages [%]

	S	G	PB	S/G ratio	β-*O*-4	β-5	β-β
BESC-316	63	38	13	1.67	56	2.8	5.7
GW-11054	67	33	20	2.05	59	1.3	3.4
BESC-292	70	30	9	2.36	60	1.9	3.9
BESC-97	73	27	13	2.64	59	1.1	3.7
BESC-876	74	26	9	2.85	56	1.1	4.7
GW-11012	75	25	7	3.00	58	0.6	4.0
BESC-8	75	25	4	3.01	59	1.6	5.8
BESC-35	76	24	12	3.25	57	1.8	5.2
BESC-5	77	22	1	3.47	59	1.5	7.6
CHWH-27	80	20	7	3.88	61	0.5	4.6

Note. Content (%) expressed as a fraction of S + G.

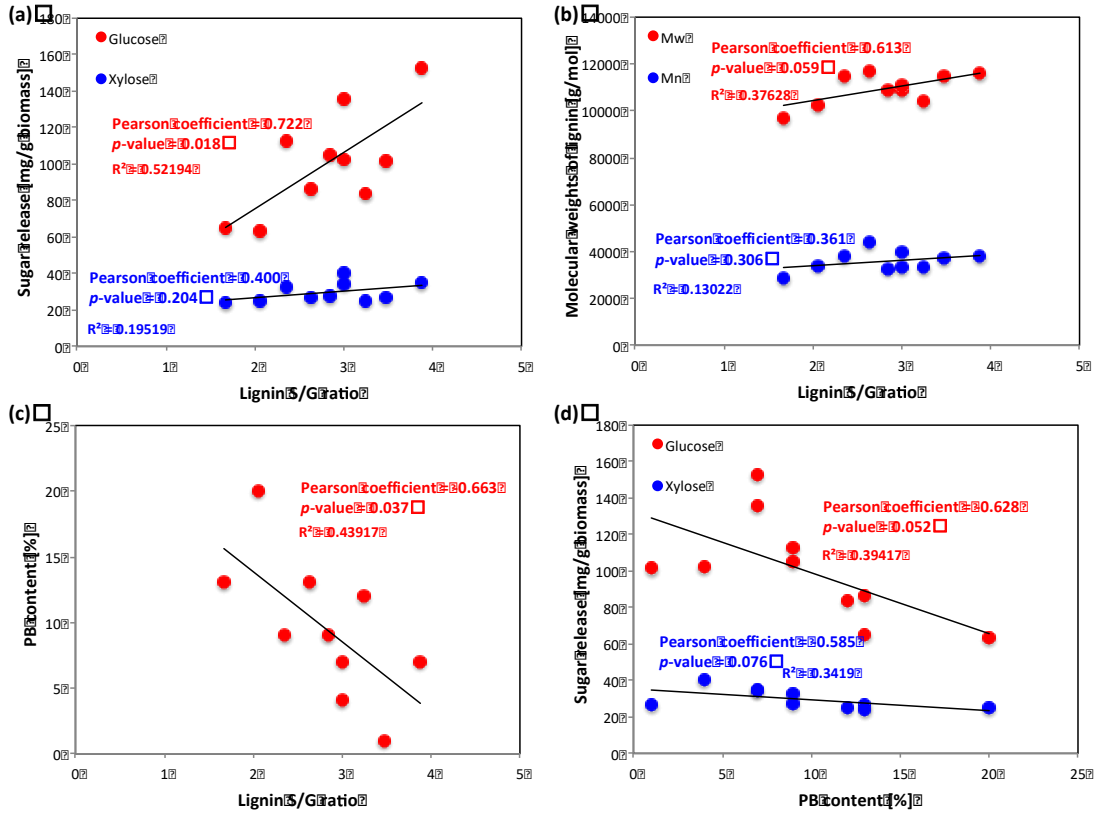


Fig. 7 Correlations of lignin S/G ratio of *P. trichocarpa* natural variants with (a) sugar release, (b) lignin molecular weight, and (c) PB content; (d) correlation between PB content and sugar release.

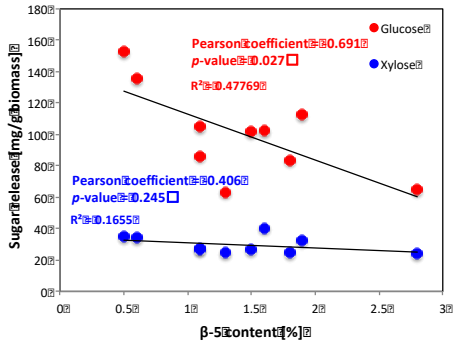


Fig. 8 Correlations between β-5 content and sugar release.

Aryl ether bonds and carbon-carbon bonds are major lignin linkages in the plant cell walls. The β-*O*-4 was the largest linkages (56 – 60%) followed by β-β (3.4 – 7.6%) and β-5 (0.5 – 2.8%). The β-*O*-4 and β–β linkages did not showed a meaningful correlation with lignin S/G ratio, while β-5 had a strong correlation with the S/G ratio (Fig. S3 and S4 in ESI). Instead, the β-5 linkage, which is majorly involved in G unit bonding, was negatively correlated with glucose release (R^2=0.48, Pearson coefficient=-0.691, *p*-value=0.027) (Fig. 8). A clear correlation between other linkages (β-*O*-4 or β-β) and sugar release was not observed (Figs. S5 and S6 in ESI).

Cellulose accessibility. Accessibility of biomass is a good indication of how well the feedstock would be hydrolyzed by enzymes.[41] In this study, accessibility of each *P. trichocarpa* natural variant was evaluated by a modified Simon's staining method using two different dyes. Simon's staining is a relatively fast and accurate method to estimate accessibility of biomass by measuring the

interior and exterior surface area of cellulose without the unexpected, irreversible pore collapse, known as hornification, caused by pre-drying.[41] Direct orange (DO) dye has a 5 – 36 nm molecular diameter with relatively high cellulose affinity for hydroxyl groups on the cellulose surface, so the DO dye only populates the larger pores which cellulase (~5.1 nm) can access. Direct blue (DB) dye has ~1 nm molecular diameter with a molecular weight of 992.8 g/mol; therefore, it penetrates the smaller pores of the substrate. The ratio of adsorbed DO and DB is also used to estimate the relative pore size distribution.[42] As Fig. 9a presents, the maximum amount of the total dyes adsorbed varied from 26.5 mg/g biomass (BESC-316) to 46.5 mg/g biomass (BESC-292). The measured total dye adsorption of *P. trichocarpa* samples was well correlated with their sugar release (Fig. 9b). In particular, glucose release was positively correlated (R^2=0.55, Pearson coefficient=0.745, *p*-value=0.013) with the total dye adsorption (accessibility of biomass). Individual DO and DB was also correlated with glucose release in this study (Fig. S7 and S8 in ESI).

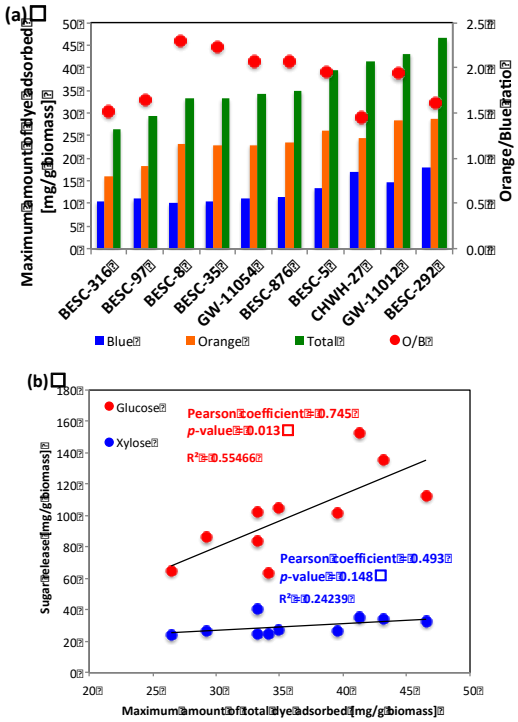


Fig. 9 (a) Cellulose accessibility of *P. trichocarpa* natural variants and (b) its correlation with sugar release.

The accessibility was also correlated with the aforementioned physicochemical properties of *Populus* samples. For example, lignin content was well correlated with the adsorption of both DO (R^2=0.48, Pearson coefficient=-0.694, *p*-value=0.026) and DB (R^2=0.47, Pearson coefficient=-0.685, *p*-value=0.029) dyes (Fig. 10a), while xylan content was only correlated with the DB adsorption (R^2=0.32, Pearson coefficient=-0.564, *p*-value=0.089) (Fig. 10b). It indicated that lignin content inhibited the accessibility of *Populus* natural variants by reducing both pore size and cellulose affinity. Cellulose content showed a positive correlation only with DB (R^2=0.41, Pearson coefficient=0.638, *p*-value=0.047) as presented in Fig. 10c. As Fig. 10d presents, the *Populus* variants with higher lignin M_w had more adsorption of DB (R^2=0.31, Pearson coefficient=0.559, *p*-value=0.093), but other properties including lignin S/G ratio, cellulose DP, and cellulose CrI did not show notable correlations. (Figs. S9-11 in ESI). The *Populus* variants with a better accessibility of cellulose resulted in higher sugar release. We also observed that two lignin-related properties, lignin content and lignin M_w were involved in cellulose accessibility of the variants.

Correlations between physicochemical properties of *P. trichocarpa* natural variants and sugar releases are summarized in Table 2 with their statistical significance. Several properties of the natural variants showed significant correlations with glucose and/or xylose releases. In brief, cellulose content, cellulose accessibility, lignin S/G ratio, and lignin M_w showed positive correlations with glucose release (*p*-value<0.05). Lignin content (*p*-value<0.05) and PB content (*p*-value<0.10) had negative correlations with both glucose and xylose releases. Also, the content of phenylcoumaran showed a negative correlation with glucose release (*p*-value<0.05). Since each variant has multiple factor variations, some of the combinational and offset effects by multiple factors were not evaluated in this study. At this point, our findings suggest that besides cellulose content and accessibility, lignin-related properties including lignin content, lignin S/G ratio and lignin M_w were dominant biomass recalcitrance factors in *P. trichocarpa* natural variants.

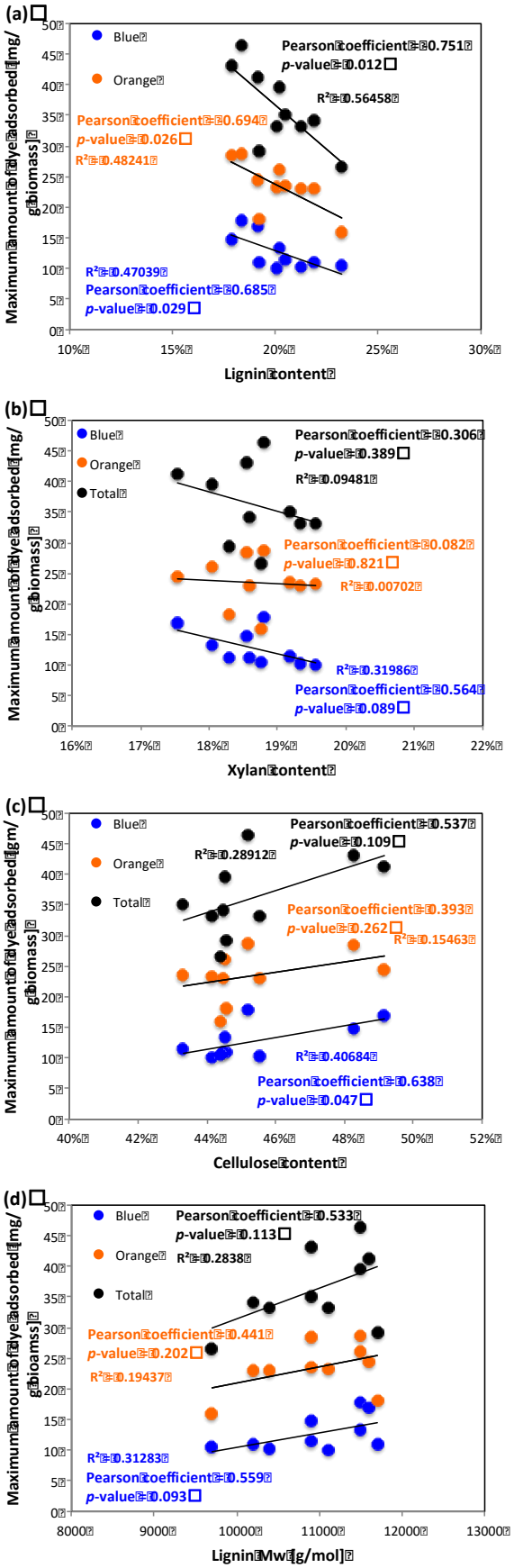


Fig. 10 Correlation of cellulose accessibility with (a) lignin content, (b) xylan content, (c) cellulose content, and (d) lignin M_w.

Table 2. Summary of correlations between physicochemical properties of *P. trichocarpa* natural variants and their sugar releases.

Properties	Glucose release		Xylose release	
	Coefficient	*p*-value	Coefficient	*p*-value
Cellulose content	+	<0.05	+	>0.10
Xylan content	-	>0.10	-	>0.10
Lignin content	-	<0.05	-	<0.05
Cellulose DP	-	>0.10	-	>0.10
Cellulose CrI	+	>0.10	-	>0.10
Cellulose accessibility	+	<0.05	+	>0.10
Hemicellulose DP	+	>0.10	+	>0.10
Lignin S/G ratio	+	<0.05	+	>0.10
Lignin Mw	+	<0.05	+	>0.10
PB content	-	<0.10	-	<0.10
β-aryl ether content	+	>0.10	+	>0.10
Phenylcoumaran conten	-	<0.05	-	>0.10
Resinols content	-	>0.10	-	>0.10

Gene expression analysis of *Populus* natural variants

To identify the correlation between variation of gene expression and aforementioned characterization results, the global transcriptome of *P. trichocarpa* natural variants grown in the field was analyzed. A subset of the RNA-seq data from five natural variants was selected according to the sugar release results. GW-11054 was selected as a representative genotype of *P. trichocarpa* with the lowest glucose release. BESC-5 and BESC-8 were designated as representatives of the variants with average glucose release, and GW-11012 and CHWH-27 were chosen for representing variants with high glucose release. A differentially expressed gene (DEG) was determined by comparison of normalized count number against that of BESC-8 for each gene. A total of 38 genes were differentially and significantly expressed over five transcriptomes with valuable read count ≥ 1 ($p < 0.1$). Among 38 genes, 30 genes were detected in all genotypes, whereas 8 genes were not expressed in at least one genotype (Fig. S12 and Table S1 in ESI).

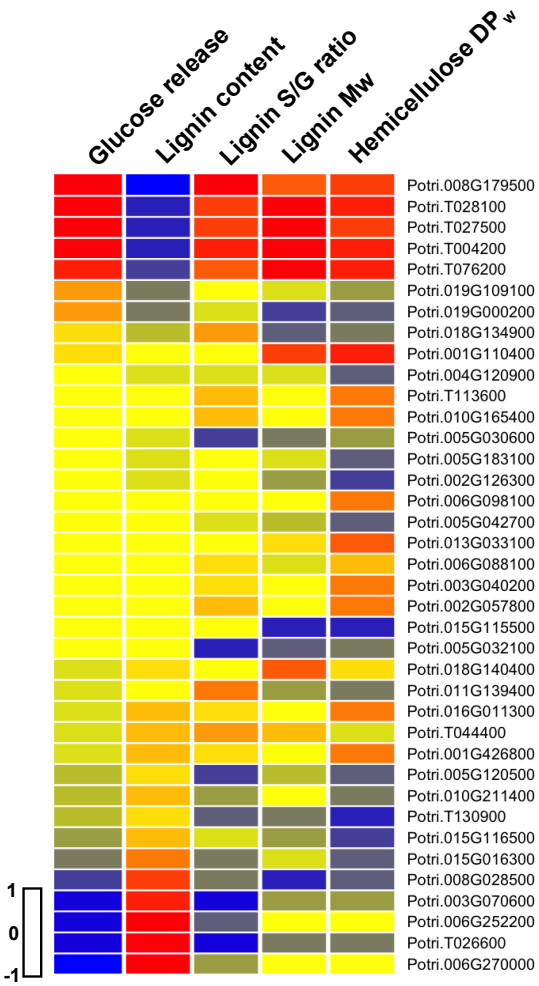


Fig. 11 Correlation heatmap of DEG with selected biomass properties and glucose release of *P. trichocarpa* natural variants.

Based on the DEG results relative to the BESC-8 genotype, the correlations of DEG and the selected properties of *P. trichocarpa,* including lignin content, lignin S/G ratio, lignin M_w, glucose release, and hemicellulose DP_w were determined. A range of 5 to 10 genes was found to correlate with high correlation coefficient against each correlation test ($|r^2| > 0.6$, Fig. 11, Table S1). Relationship of correlations for each property was also tested and shown in Fig. 12. A negative correlation with a high regression value (R^2=0.98) was observed between DEGs with lignin content and glucose release. On the contrary, the correlation of DEGs with lignin S/G ratio and glucose release were positively associated. These results were consistent with the aforementioned correlation trends between biomass properties and sugar release (Figs. 2d, 3c, 3d, 5b and 7a). In particular, NB-ARC domain-containing protein and zinc binding dehydrogenase family, which were reported as lignin biosynthesis-related genes, were identified as the genes affecting both correlations (Fig. 11).[43, 44] However, the coefficient of the correlation between DEG and lignin M_w against the correlation of DEG and sugar release was not as high as others. The correlation between DEG and hemicellulose DP_w versus the correlation of DEG with sugar release was also not significant, which was consistent with the characterization results.

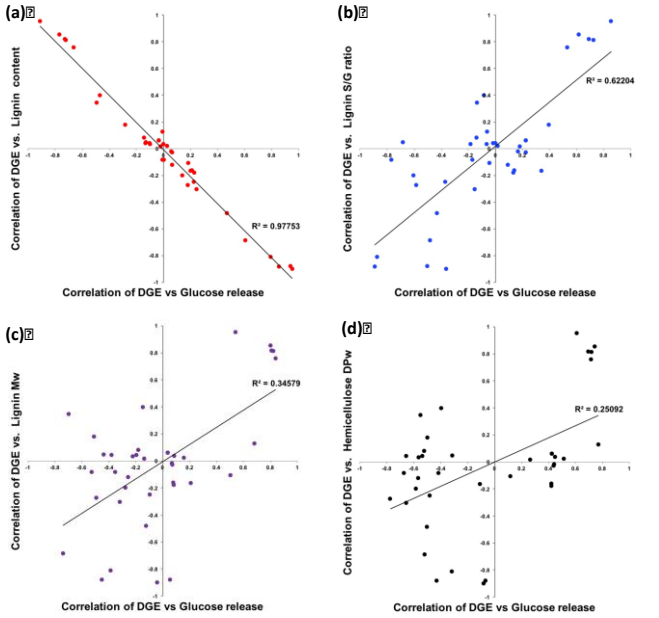


Fig. 12 Linear regression of correlation of DEG with the selected properties of *P. trichocarpa* natural variants and correlation of DEG with glucose release.

Conclusions

Although several of the biomass recalcitrance-related factors have been explored in the previous studies, few of them have integrated all the cell wall structural components together along with the plant genetics of poplar and recalcitrance. The results of this study present the associations between physicochemical properties of *Populus trichocarpa* natural variants and their sugar release performance. Specifically, lignin content, lignin S/G ratio, lignin M_w, cellulose content, and cellulose accessibility were found to influence sugar release. Lignin content negatively affected the sugar release of *P. trichocarpa* genotypes, whereas lignin S/G ratio, lignin M_w, cellulose content, and cellulose accessibility were positively correlated with conversion performance. Associations of cellulose accessibility with lignin content and lignin M_w were also observed. On the contrary, cellulose DP, hemicellulose DP, cellulose crystallinity, and others were not correlated with sugar release. In addition, these correlations were evaluated and supported by differentially expressed gene information determined from the transcriptome analysis. The DGE results from natural variants suggested potential target gene groups correlated with the recalcitrance factors. These results provide valuable information for the future biomass studies to understand the novel genes and its regulatory mechanism in

forest genetics and apply to tree breeding for enhancing biomass utilization.

Experimental

Materials

Populus trichocarpa natural variants were collected from trees that had been growing in the field for five years at a common garden at Clatskanie, Oregon. After harvesting, the stem samples were air-dried, debarked, chipped, and Wiley milled to 0.42 mm size. The plant growth conditions were described in the previous study.[45]

Sugar release test

Glucose and xylose release of the *P. trichocarpa* natural variants by enzymatic hydrolysis were determined to evaluate the conversion of each variant. The reaction conditions were determined according to the previous study.[46] The enzymatic hydrolysis was conducted in an incubator shaker at 50 °C with a 5% (*w/v*) solid loading for 5 days. Cellulase mix Cellic® Ctec2 was loaded at 24 FPU/g glucan with β-glucosidase (Novozymes 188; 25% volume ratio to Ctec2), hemicellulases Cellic® Htec2 (20% volume ratio to Ctec2), and streptomycin (0.063 mg/mL final concentration). After the hydrolysis, the hydrolysate of each sample was filtered using a 0.2μm filter and acidified to pH 2 with 2 M H_2SO_4. The acidified hydrolysate was analyzed using high performance liquid chromatography (Hitachi High Technologies America Inc., Irving, TX) equipped with a refractive index (RI) detector (model L-2490) and Aminex™ HPX-87H column (Bio-Rad Laboratories Inc. Hercules, CA) at 60 °C and at a flow rate of 0.5 mL/min with 5.0 mM H_2SO_4 as a mobile phase. Glucose and xylose standards were used for the calibration.

Chemical composition analysis

Prior to the chemical composition analysis, biomass was extracted with dichloromethane using a Soxhlet extraction apparatus for 8 hours. The extractives-free biomass was air-dried and stored in a refrigerator for the following analysis. Cellulose, hemicellulose, and lignin in each biomass sample were analyzed by a two-step acid hydrolysis method according to the literature.[47] Briefly, the extractives-free samples were hydrolyzed at 30 °C in 72% sulfuric acid for 1 hour, and then autoclaved at 121 °C for another 1 hour after diluting to 4% sulfuric acid with deionized (DI) water. Carbohydrate fractions were measured by a high performance anion exchange chromatography system with pulsed amperometric detector (HPAEC-PAD, Dionex ICS-3000, Dionex Corp., USA), CarboPac PA1 column (2×250 mm, Dionex), and CarboPac PA1 guard column (2×50 mm, Dionex). Moisture content was measured by a halogen moisture analyzer.

Degree of polymerizations of cellulose and hemicellulose analyses

Cellulose and hemicellulose were isolated from the extractives-free *P. trichocarpa* samples as described previously.[48, 49] Samples was delignified with peracetic acid (5% solid loading) at 25 °C for 24 hours. The recovered holocellulose samples were vacuum-dried at 40 °C for 12 hours. Cellulose was recovered through a two-step alkali extraction. In brief, the holocellulose was extracted with 5 mL of 17.5% sodium hydroxide (NaOH) at 25 °C for 2 hours, and then the mixture was reduced to 8.75% NaOH solution by adding DI water and stirred at 25 °C for another 2 hours. The alkaline slurry was filtrated and rinsed with 1% acetic acid and excess of DI water. Finally, the solid fraction, called α-cellulose, was air-dried for cellulose DP analysis. Cellulose DP was analyzed by an Agilent gel permeation chromatography (GPC) SECurity 1200 system with four Waters Styragel columns (HR1, HR2, HR4, and HR5), a RI detector, and a UV detector. Prior to the GPC analysis, the isolated α-cellulose was derivatized by tricarbanilation with phenyl isocyanate, as described in the previous study.[49] Mobile phase for the analysis was THF and polystyrene standards were used for calibration. Data collection and processing were conducted by Polymer Standards Service WinGPC Unity software. DP_w and DP_n of cellulose were calculated with the molecular weight of the tricarbanilated cellulose repeating unit (519 g/mol).

The liquid fraction was adjusted to pH 6 – 7 with anhydrous acetic acid, and mixed with three times the volume of ethanol (200 proof) to precipitate hemicellulose. The precipitated hemicellulose was centrifuged and recovered by freeze-drying. The analysis for hemicellulose DP was conducted with an Agilent GPC SECurity 1200 system with three columns (Ultrahydrogel 120, 250, and 500, Waters Inc.). The samples were dissolved in 0.2 M NaOH/0.1 M sodium acetate (pH 13.0) and analyzed. Since major composition in the hemicellulose was xylan, the molecular weight of xylan repeating unit (132 g/mol) was used for the hemicellulose DP calculation.

Cellulose crystallinity test

Cellulose crystallinity was measured by solid-state NMR analysis. For the analysis, cellulose was isolated using 2.5M HCl, as described previously.[21] The isolated cellulose was kept in a sealed container to maintain the moisture content higher than 30%. The cellulose was packed into a 4 mm cylindrical ceramic MAS rotor. Cross-polarization magic angle spinning (CP/MAS) NMR analysis was performed using a Bruker Avance III 400 MHz spectrometer operating at a frequency of 100.59 MHz for ^{13}C in a Bruker double-resonance MAS probe at spinning speed of 10 kHz. A 5 μs (90 °) proton pulse, 1.5 ms contact pulse, 4 s recycle delay, and 4-8k scans were applied for CP/MAS experiments.

Lignin molecular weight analysis

Cellulolytic enzyme lignin (CEL) was isolated from *P. trichocarpa* samples according to the procedure in our previous study.[38] In brief, the extractives-free samples were ball-milled using Retsch PM 100 at 580 rpm for 2 hours. The ball-milled samples were hydrolyzed at 50 °C for 48 hours using the CTec2 enzyme in acetate buffer (pH 4.8) solution. The enzymatic hydrolysis was repeated with fresh enzyme mixture under the same conditions. The solid residues were air-dried and extracted with 96% dioxane/water mixture for 48 hours. The dioxane-extracted lignin fraction, named CEL, was recovered by freeze-drying for the lignin characterization.

Lignin molecular weights were analyzed by the Agilent GPC SECurity 1200 system with four Waters Styragel columns (HR1, HR2, HR4, and HR5), a refractive index (RI) detector, and a UV detector. Prior to the analysis, the isolated lignin was acetylated using acetic anhydride/pyridine mixture (1:1, *v/v*) at 25 °C for 24 hours. Trace of acetic acid and pyridine were removed from the acetylated lignin using a rotary evaporator with ethanol at 40 °C under vacuum condition. THF was used as a mobile phase and polystyrene standards were used for the calibration.

Heteronuclear single quantum coherence (HSQC) NMR analysis for structural information of lignin

The isolated lignin from each *P. trichocarpa* natural variant was dissolved in a 5 mm NMR tube with DMSO-d_6. Two-dimensional (2D) ^{1}H-^{13}C HSQC NMR experiment was conducted at 298 K using a Bruker Avance III 400 MHz spectroscopy equipped with a 5 mm Broadband Observe probe with the following conditions: spectral width of 11 ppm in F2 (1H) with 2048 data points and 190 ppm in F1 (13C) with 256 data points; 128 scans (NS) and 1 s interscan delay (D1). Bruker's TopSpin 3.5 software was used for volume integration of contours in HSQC spectra.

Simon's staining analysis for cellulose accessibility

Direct Orange (Pontamine Fast Orange 6RN) and Direct Blue (Pontamine Fast Sky Blue 6BX) dyes were purchased from Pylam Products Co. Inc. (Garden City, NY). The blue dye was used as received. In terms of orange dye, an ultrafiltration was performed to remove the low molecular weight fraction because it has been suggested that only the high molecular weight fraction of orange dye is responsible for the increased affinity for cellulose binding, whereas the low molecular weight fraction has a similar affinity for cellulose binding as the blue dye. It was done by filtering a 1% solution of orange dye through a 100 K membrane using an Amicon ultrafiltration apparatus (Amicon Inc., Beverly, MA) under 200 kPa nitrogen gas pressure. The concentration of orange dye after ultrafiltration was determined by placing 1.00 mL of solution in a 50 °C oven for a week followed by measuring the solid residue after that.

Biomass samples (~100 mg) were weighted into five centrifuge tubes together with 1 mL of phosphate buffered saline solution. A set of tubes containing a 1:1 mixture of blue and orange dye at increasing concentrations were prepared by adding dyes in a series of increasing volumes. DI water was then added to each tube to make up the final volume to 10 mL. Samples were then incubated at 70 °C for ~6 h with shaking at 200 rpm. After that, the absorbance of the supernatant solution was obtained on a Lambda 35 UV/Vis spectrophotometer at 455 and 624 nm, representing the wavelength of maximum absorbance for orange and blue dye, respectively. The amount of each dye adsorbed by biomass was calculated by the differences between the initial concentration and the supernatant concentration which was determined by solving two Lambert-Beer law equations simultaneously. The maximum amount of dye adsorbed was finally calculated using the Langmuir adsorption equation.

Statistical analysis

Correlations between each factor and sugar releases were analyzed by Microsoft Excel 2011 and XLSTAT. A Pearson correlation coefficient and *p*-value were used to evaluate the significance of each correlation.

Transcriptome analysis

Transcriptome data was prepared by the Illumina RNA-seq technique.[23] Total RNA was extracted from the developing xylem just below bark of matured stem of the *P. trichocarpa* genotypes. Total RNA extraction was performed as described in the previous study.[50] A total of 4 µg of total RNA was subjected to generate cDNA library for Illumina sequencing.

The RNA-seq library was generated by standard method as described in Illumina sequencing sample preparation protocol (San Diego, CA, USA). The sequencing procedure was performed using an illumina Hi-seqII. Raw reads were evaluated by BBDuk (k-mer =25) in the JGI pipeline and then reads were trimmed below 25 base length. Filtered reads were aligned to the *Populus trichocarpa v 3.0* annotation gff3 file using HISAT version 0.1.4-beta.[51] Raw count was generated by featureCounts.[52] The differential expression gene analysis (fold change determination) was performed by Bioconductor package of DESeq2, count-based statistical method.[53] BESC-8 normalized count was used as a base count to calculate fold change of other transcriptomes.

For the case of absolutely expressed or not expressed genes in BESC-8, we used original count number for further analysis. The correlation coefficient between DEG and physicochemical property was calculated with log2 scaled fold change value by "cor" function with method option of Person's product-moment correlation in R package (R). A heatmap was generated by the gplots package in R package (R).[54] The relationship of correlations was determined by the linear regression analysis integrated function in Microsoft Excel (Redmond, WA, USA). All figures were drawn under the same data set and results using Microsoft Excel.

Acknowledgements

This manuscript has been authored by UT-Battelle, LLC under Contract No. DE-AC05-00OR22725 with the U.S. Department of Energy. This study was supported and performed as part of the BioEnergy Science Center (BESC). The BioEnergy Science Center is a U.S Department of Energy Bioenergy Research Center supported by the Office of Biological and Environmental Research in the DOE Office of Science.

Notes and references

1. J. A. Melero, J. Iglesias and A. Garcia, *Energy & Environmental Science*, 2012, **5**, 7393-7420.
2. A. Perimenis, H. Walimwipi, S. Zinoviev, F. Müller-Langer and S. Miertus, *Energy Policy*, 2011, **39**, 1782-1793.
3. P. Sannigrahi, A. J. Ragauskas and G. A. Tuskan, *Biofuels, Bioproducts and Biorefining*, 2010, **4**, 209-226.

4. G. A. Tuskan, S. Difazio, S. Jansson, J. Bohlmann, I. Grigoriev, U. Hellsten, N. Putnam, S. Ralph, S. Rombauts and A. Salamov, *Science*, 2006, **313**, 1596-1604.
5. G. Tuskan, S. DiFazio and T. Teichmann, *Plant Biology*, 2004, **7**, 2-4.
6. M. Hinchee, W. Rottmann, L. Mullinax, C. Zhang, S. Chang, M. Cunningham, L. Pearson and N. Nehra, in *Biofuels*, Springer, 2011, pp. 139-156.
7. W. Chen, H. Yu, Q. Li, Y. Liu and J. Li, *Soft Matter*, 2011, **7**, 10360-10368.
8. J. Li, J. Li, D. Feng, J. Zhao, J. Sun and D. Li, *RSC Advances*, 2017, **7**, 28889-28897.
9. J. Zhang, F. Gu, J. Zhu and R. S. Zalesny, *Bioresource Technology*, 2015, **186**, 223-231.
10. C. G. Yoo, S. Zhang and X. Pan, *RSC Advances*, 2017, **7**, 300-308.
11. M. E. Himmel, S.-Y. Ding, D. K. Johnson, W. S. Adney, M. R. Nimlos, J. W. Brady and T. D. Foust, *Science*, 2007, **315**, 804-807.
12. S. P. Chundawat, G. T. Beckham, M. E. Himmel and B. E. Dale, *Annual Review of Chemical and Biomolecular Engineering*, 2011, **2**, 121-145.
13. S. Dutta, K. C.-W. Wu and B. Saha, *Catalysis Science & Technology*, 2014, **4**, 3785-3799.
14. L. Shuai, M. T. Amiri, Y. M. Questell-Santiago, F. Héroguel, Y. Li, H. Kim, R. Meilan, C. Chapple, J. Ralph and J. S. Luterbacher, *Science*, 2016, **354**, 329-333.
15. S. Dutta and K. C.-W. Wu, *Green Chemistry*, 2014, **16**, 4615-4626.
16. H. Shen, C. R. Poovaiah, A. Ziebell, T. J. Tschaplinski, S. Pattathil, E. Gjersing, N. L. Engle, R. Katahira, Y. Pu and R. Sykes, *Biotechnology for Biofuels*, 2013, **6**, 71.
17. J. S. Luterbacher, J. M. Rand, D. M. Alonso, J. Han, J. T. Youngquist, C. T. Maravelias, B. F. Pfleger and J. A. Dumesic, *Science*, 2014, **343**, 277-280.
18. M. H. Studer, J. D. DeMartini, M. F. Davis, R. W. Sykes, B. Davison, M. Keller, G. A. Tuskan and C. E. Wyman, *Proceedings of the National Academy of Sciences*, 2011, **108**, 6300-6305.
19. S. Park, J. O. Baker, M. E. Himmel, P. A. Parilla and D. K. Johnson, *Biotechnology for Biofuels*, 2010, **3**, 10.
20. C. Cateto, G. Hu and A. Ragauskas, *Energy & Environmental Science*, 2011, **4**, 1516-1521.
21. M. Foston, C. A. Hubbell, R. Samuel, S. Jung, H. Fan, S.-Y. Ding, Y. Zeng, S. Jawdy, M. Davis and R. Sykes, *Energy & Environmental Science*, 2011, **4**, 4962-4971.
22. X. Meng and A. J. Ragauskas, *Current Opinion in Biotechnology*, 2014, **27**, 150-158.
23. S. Goodwin, J. D. McPherson and W. R. McCombie, *Nature Reviews Genetics*, 2016, **17**, 333-351.
24. A. J. Foster, G. Pelletier, P. Tanguay and A. Séguin, *PloS ONE*, 2015, **10**, e0138162.
25. R. M. Cossu, T. Giordani, A. Cavallini and L. Natali, *Tree Genetics & Genomes*, 2014, **10**, 53-66.
26. H. Bao, E. Li, S. D. Mansfield, Q. C. Cronk, Y. A. El-Kassaby and C. J. Douglas, *BMC Genomics*, 2013, **14**, 359.
27. M. Chen, H. Bao, Q. Wu and Y. Wang, *International Journal of Molecular Sciences*, 2015, **16**, 13937-13958.
28. R. Vicentini, A. Bottcher, M. dos Santos Brito, A. B. dos Santos, S. Creste, M. G. de Andrade Landell, I. Cesarino and P. Mazzafera, *PloS ONE*, 2015, **10**, e0134909.
29. C. A. Hefer, E. Mizrachi, A. A. Myburg, C. J. Douglas and S. D. Mansfield, *New Phytologist*, 2015, **206**, 1391-1405.
30. E. Meyer, T. L. Logan and T. E. Juenger, *The Plant Journal*, 2012, **70**, 879-890.
31. Q. Du, C. Gong, Q. Wang, D. Zhou, H. Yang, W. Pan, B. Li and D. Zhang, *New Phytologist*, 2016, **209**, 1067-1082.
32. M. J. Taherzadeh and K. Karimi, *International Journal of Molecular Sciences*, 2008, **9**, 1621-1651.
33. T.-C. Hsu, G.-L. Guo, W.-H. Chen and W.-S. Hwang, *Bioresource Technology*, 2010, **101**, 4907-4913.
34. C. G. Yoo, M. Li, X. Meng, Y. Pu and A. J. Ragauskas, *Green Chemistry*, 2017.
35. X. Pan, *Journal of Biobased Materials and Bioenergy*, 2008, **2**, 25-32.
36. S. Sun, Y. Huang, R. Sun and M. Tu, *Green Chemistry*, 2016, **18**, 4276-4286.
37. Y. Pu, F. Hu, F. Huang, B. H. Davison and A. J. Ragauskas, *Biotechnology for biofuels*, 2013, **6**, 15.
38. X. Meng, Y. Pu, C. G. Yoo, M. Li, G. Bali, D. Y. Park, E. Gjersing, M. F. Davis, W. Muchero and G. A. Tuskan, *ChemSusChem*, 2017, **10**, 139-150.
39. T. Kishimoto, W. Chiba, K. Saito, K. Fukushima, Y. Uraki and M. Ubukata, *Journal of Agricultural and Food Chemistry*, 2009, **58**, 895-901.
40. F. Lu, S. D. Karlen, M. Regner, H. Kim, S. A. Ralph, R.-C. Sun, K.-i. Kuroda, M. A. Augustin, R. Mawson and H. Sabarez, *BioEnergy Research*, 2015, **8**, 934-952.
41. R. P. Chandra and J. N. Saddler, *Industrial Biotechnology*, 2012, **8**, 230-237.
42. X. Meng, M. Foston, J. Leisen, J. DeMartini, C. E. Wyman and A. J. Ragauskas, *Bioresource Technology*, 2013, **144**, 467-476.
43. C. Bukh, P. H. Nord-Larsen and S. K. Rasmussen, *Journal of Experimental Botany*, 2012, **63**, 6223-6236.
44. Y. Kawano, T. Kaneko-Kawano and K. Shimamoto, *Frontiers in Plant Science*, 2014, **5**, 522.
45. W. Muchero, J. Guo, S. P. DiFazio, J.-G. Chen, P. Ranjan, G. T. Slavov, L. E. Gunter, S. Jawdy, A. C. Bryan and R. Sykes, *BMC Genomics*, 2015, **16**, 24.
46. A. Dumitrache, J. Natzke, M. Rodriguez, K. L. Yee, O. A. Thompson, C. R. Poovaiah, H. Shen, M. Mazarei, H. L. Baxter and C. Fu, *Plant Biotechnology Journal*, 2017, **15**, 688-697.
47. A. Sluiter, B. Hames, R. Ruiz, C. Scarlata, J. Sluiter, D. Templeton and D. Crocker, *Determination of structural carbohydrates and lignin in biomass: Laboratory Analytical Procedure (LAP) (Revised August 2012). Issue Date: 4/25/2008*, NREL/TP-510-42618.
48. R. Kumar, F. Hu, C. A. Hubbell, A. J. Ragauskas and C. E. Wyman, *Bioresource Technology*, 2013, **130**, 372-381.
49. M. Li, Y. Pu, C. G. Yoo, E. Gjersing, S. R. Decker, C. Doeppke, T. Shollenberger, T. J. Tschaplinski, N. L. Engle and R. W. Sykes, *Biotechnology for Biofuels*, 2017, **10**, 12.
50. Y. Yang, C. G. Yoo, H.-B. Guo, W. Rottmann, K. A. Winkeler, C. M. Collins, L. E. Gunter, S. S. Jawdy, X. Yang and H. Guo, *Biotechnology for Biofuels*, 2017, **10**, 74.
51. D. Kim, B. Langmead and S. L. Salzberg, *Nature Methods*, 2015, **12**, 357-360.
52. Y. Liao, G. K. Smyth and W. Shi, *Bioinformatics*, 2014, **30**, 923-930.
53. M. I. Love, W. Huber and S. Anders, *Genome Biology*, 2014, **15**, 550.

54. R Core Team. R: A language and environment for statistical computing. R foundation for statistical computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-projcet.org/.