Rapid screening of secondary aromatic metabolites in *Populus trichocarpa* leaves

Anne Elizabeth Harman-Ware (✉ anne.ware@nrel.gov)

National Renewable Energy Laboratory <https://orcid.org/0000-0002-7927-9424>

Madhavi Martin

Oak Ridge National Laboratory

Nancy L. Engle

Oak Ridge National Laboratory

Crissa Doeppke

National Renewable Energy Laboratory

Timothy J. Tschaplinski

Oak Ridge National Laboratory

Research

Keywords: pyrolysis-molecular beam mass spectrometry, *Populus trichocarpa*, high-throughput analysis, metabolomics

Posted Date: October 23rd, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-96304/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background High-throughput metabolomics analytical methodology is needed for population-scale studies of bioenergy-relevant feedstocks such as poplar (*Populus* sp.). Here, the authors report the relative abundance of secondary aromatic metabolites in *Populus trichocarpa* leaves rapidly estimated using pyrolysis-molecular beam mass spectrometry (py-MBMS). A training set of leaves was analyzed in conjunction with GC/MS analysis of extracts to determine key spectral features associated with the aromatic metabolites and build PLS models to predict the relative composition of extractable aromatic metabolites in a second test set of leaves. The relative abundance of major secondary aromatic metabolites for the test set of leaves was validated by GC/MS analysis of extracts.

Results The Pearson correlation coefficient for the relative abundance of aromatic metabolites based on ranking between GC/MS analysis and py-MBMS analysis of the test leaf set was 0.86 with $R^2 = 0.76$ using a simplified prediction approach from select ions in MBMS spectra. Metabolites most influential to py-MBMS spectral features in the training set included the following compounds: catechol, salicortin, salicyloyl-coumaroyl-glucoside, a-salicyloylsalicin, tremulacin, and to a lesser degree: benzyl-coumaroyl glucoside, benzoyl-salicyloylsalicin, tremuloidin, salireposide, trichocarpin, salicylic acid, salicyl-salicylic acid-2-O-glucoside, phenethyl-tremuloidin, coumaroyl-tremuloidin, 2,6-cyclohexadiene-1,2-diol, and salicyl-tremuloidin. Ions in py-MBMS spectra with the highest correlation to the abundance of aromatic secondary metabolites, particularly those previously mentioned as determined by GC/MS analysis of extracts, included m/z 68, 71, 77, 91, 94, 105, 107, 108, and 122, and were used to develop the simplified prediction approach without PLS models or *a priori* measurements.

Conclusions The simplified py-MBMS method is capable of rapidly screening leaf tissue for relative abundance of extractable aromatic secondary metabolites to enable prioritization of samples in large populations requiring comprehensive metabolomics that will ultimately inform plant systems biology models and advance the development of optimized biomass feedstocks for renewable fuels and chemicals.

Background

The analysis of metabolites in lignocellulosic biomass is important for the production of sustainable feedstocks that will serve as a renewable source of fuels and chemicals. Metabolite content and composition in biomass are impacted by and provide insights related to genetics, carbon flux, environmental stress responses, and disease resistance. For example, *Populus* species, a promising woody feedstock for the production of renewable materials whose genome has been fully sequenced (1), produce secondary metabolites in leaves that vary genetically and according to stresses and environmental responses (2, 3). A comprehensive understanding of *Populus* leaf metabolomics can provide meaningful insight regarding plant biology and physiology and be leveraged for the domestication and rational design of *Populus* feedstocks to ensure crops with highest biomass yield and optimal composition, as well as disease resistance and environmental stress tolerance.

In order to perform genome-wide association studies (GWAS), develop systems biology models and produce superior biomass feedstocks with many favorable characteristics, large sample populations of varying genotypes and phenotypes must be studied. The measurement of certain phenotypes in large sample populations can be cumbersome if high-throughput methodologies are not available. High-throughput methods using UAV have been developed for the measurement of growth metrics in plants and rapid biomass characterization methods have been established for estimation of lignin and sugar content and enzymatic recalcitrance (4, 5). Comprehensive secondary metabolic profiles of biomass samples are difficult to achieve in high-throughput pipelines due to the number of steps required to prepare, extract and analyze samples. Typically, biomass samples are homogenized, extracted with solvent and undergo chromatography prior to detection, identification, and quantitation of individual metabolite components. While comprehensive metabolomics relies on highly resolved and unbiased methods capable of distinguishing and quantifying specific metabolites, these methods are not necessarily conducive to large sample set analysis in a reasonable timeframe, particularly if there are many metabolites present in the extracts. There have been a number of high-throughput metabolic screening methods developed for plant metabolomics, particularly involving the use of NMR (6) advanced mass spectrometry (7) and colorimetric (8) approaches to analyze extracts. Also, the rapid determination of biomass cell wall polymer content and composition has been achieved using minimal sample preparation followed by analysis using near-infrared spectroscopy or pyrolysis-Molecular Beam Mass Spectrometry (5). However, high-throughput methods used to screen or estimate plant metabolic profiles prior to in-depth metabolite characterization, particularly with minimal sample preparation or without extraction methodology, are lacking. Rapid screening techniques for plant metabolomics could provide insight that would alleviate pressure on comprehensive pipelines by providing down-selection criteria and prioritization of samples for analysis. High-throughput metabolite screening may also provide data that could be used to develop plant biology systems models and advance genomics research.

Here, we report the analysis of *Populus trichocarpa* leaves using pyrolysis-Molecular Beam Mass Spectrometry (py-MBMS) as a screening technique for metabolomics analysis. Two separate sets of leaves were analyzed where one set was used to build spectroscopic strategies for metabolite predictions using GC/MS analysis of extracts and the other set was ranked according to aromatic metabolite composition and validated using GC/MS analysis of extracts. PLS models were also constructed from the two sets of leaves to demonstrate the ability to use py-MBMS data to predict aromatic metabolite content in extract when *a priori* metabolic data is available within an analysis set. Greater than 250 milled and lyophilized whole biomass samples can be analyzed daily by the simplified py-MBMS screening method. Additionally, Py-MBMS simultaneously provided information regarding lignin content and monolignol composition in the leaf samples.

Results

Metabolic profiles of *P. trichocarpa* leaves

Metabolites present in *P. trichocarpa* leaves as determined by GC/MS analysis of extract are detailed in Supplementary Table 1. 176 metabolites were identified and semi-quantified relative to an internal standard by GC/MS analysis of extract from the training leaf set and of those metabolites, 80 were classified to be “aromatic” in nature (consisting of phenolics, benzoates, salicylic acid moieties, etc.). 115 metabolites were identified and semi-quantified in the test set of extracts based on GC/MS analysis, of these 49 were positively identified to be aromatic.

Py-MBMS analysis of *P. trichocarpa* leaves

Py-MBMS analysis of the training set of *P. trichocarpa* genotypes grown in Clatskanie, OR was used to analyze spectral features consistent with biomass composition that would enable estimation of the relative lignin content and syringyl/guaiacyl (S/G) ratio in lignin, as well as the relative abundance of specific metabolites present in the leaves. Table 1 shows the average lignin content and S/G ratio of the leaves determined by summation of mean-normalized ion intensities of *m/z* 120, 124 (G), 137 (G), 138 (G), 150 (G), 152, 154 (S), 164 (G), 167 (S), 168 (S), 178 (G), 180, 181, 182 (S), 194 (S), 208 (S) and 210 (S), where G denotes primarily guaiacyl-derived ions, S denotes primarily syringyl-derived ions, and other ions either derive from other lignin monomers or multiple sources. Lignin content was estimated by using a single point response factor relative to a representative NIST standard of known Klason lignin content using the same ions. Syringyl-to-guaiacyl (S/G) ratios were determined by dividing the sum of S-based ions by the sum of G-based ions using mean-normalized ion intensities.

Table 1
Lignin content and composition of *P. trichocarpa* leaf sets as determined by py-MBMS.

<i>P. trichocarpa</i> leaf set	Lignin Content (DW %)	S/G
Clatskanie Training (n = 219)	9.7 (± 0.7)	0.6 (± 0.1)
Boardman Test (n = 223)	9.4 (± 0.6)	0.8 (± 0.1)

The leaves in both sets exhibit significantly lower lignin content and S/G as is expected and compared to that typically seen in mature woody stem xylem tissue (9–12). Additionally, the variation in ions and principal component analysis (PCA) show that the lignin-derived ions are not the main source of spectral variance across the sets. Principal component analysis (PCA) loadings for the first principle component plotted in spectral format to demonstrate the variation in the test leaf set arises primarily from ions *m/z* 43, 57, 71, 85, 95, 97, which originate primarily from sugars, as well as *m/z* 77, 91, 94, 105, and 122, derived primarily from aromatics (13, 14), which were generally negatively correlated with the sugar-derived ions (Fig. 1a).

The relative lignin content and S/G determined for the test set of *P. trichocarpa* genotypes grown in Boardman, OR, were found to be similar to those in the training set from Clatskanie, OR (Table 1). Additionally, the majority of spectral variation originates from similar ions derived from aromatics,

including m/z 68, 71, 91, 94, 105, 107, 108, and 122 (Fig. 1b). However, various minor differences from a number of sources (true biomass compositional variation, instrumental drift, etc.) make it difficult to directly compare the two sets and develop PLS modeling approaches that can be used across sets without including several of the same standards, samples, etc. in each set. In order to circumvent this without requiring *a priori* metabolomics by GC/MS, and because these differences may not be the same for any future sets being compared, and also to simplify future analyses, the main source of spectral variation (being ions derived from aromatics) was used as a rough estimate to compare samples within the sets in the same way lignin content has been compared using this method over the last decade (5, 9, 10, 12, 15–17).

Aromatic metabolite predictions using py-MBMS

Pearson correlation coefficients were determined for select ions in the MBMS spectra, lignin content and monolignol ratios (S/G), as well as the total fractional abundance of aromatics determined by GC/MS from the training set of leaves (Table 2). The select ions in the MBMS spectra (Table 2) and the ions chosen to represent the sum of aromatic metabolites were chosen based on the correlation coefficients with aromatic metabolites, the principal component analyses (Fig. 1), and the loadings of the ions in the PLS models constructed using this data (Fig. 2). The PLS models used to predict the relative aromatic composition of extractable metabolites were primarily driven by the ions known to derive from aromatic species (13), primarily m/z 68, 71, 77, 91, 94, 105, 107, 108, and 122, which were also the main source of variation in the spectra of the set and provided a reasonable estimation and validation of using py-MBMS to predict aromatic metabolite composition in poplar leaf extracts. The sum of the aromatic metabolite ion intensities in the MBMS spectra from the training set had a correlation coefficient with the sum of the aromatic metabolites determined by GC/MS of 0.87 and R^2 of 0.76 (Fig. 2c). This simplified method of summing the ion intensities (m/z) 68, 71, 77, 91, 94, 105, 107, 108, and 122 to predict the relative abundance of aromatic metabolites in extract could theoretically be used in place of PLS modeling for ranking samples when limited *a priori* knowledge of metabolite information is available.

While many of the aromatic metabolites had a positive correlation with the sum of the aromatic metabolite intensities from the MBMS, they all differed in value and many of the signature ions produced from each metabolite overlap, making it difficult to differentiate and resolve the specific species (Table 2). Therefore, this method was developed to rapidly estimate the total relative abundance of aromatic metabolites present in poplar leaf extract samples. Further deconvolution and speciation of specific metabolites will be the subject of future investigations.

Table 2

Correlations between select ions from MBMS spectra of *P. trichocarpa* leaves in the training set from Clatskanie, OR and select metabolites as determined by GC/MS as well as lignin content, S/G ratios and the summation of metabolite-derived ions from MBMS spectra and the total aromatic metabolites determined by GC/MS. *metabolite ion intensities summed 68, 71, 77, 91, 94, 105, 107, 108, and 122

Trait	Lignin content	S/G	Sum metabolite ions in MBMS*	Aromatics by GC/MS
m/z 64	-0.34	0.02	0.68	0.65
m/z 65	-0.38	0.32	0.74	0.69
m/z 66	-0.54	0.32	0.87	0.80
m/z 68	-0.49	0.30	0.71	0.59
m/z 77	-0.64	0.38	0.98	0.85
m/z 78	-0.43	0.23	0.72	0.66
m/z 91	-0.39	0.25	0.68	0.59
m/z 92	-0.32	0.25	0.61	0.53
m/z 94	-0.55	0.35	0.93	0.85
m/z 105	-0.66	0.35	0.98	0.85
m/z 106	-0.52	0.40	0.77	0.68
m/z 107	-0.59	0.36	0.96	0.82
m/z 108	-0.65	0.45	0.95	0.84
m/z 110	-0.26	-0.06	0.61	0.60
m/z 122	-0.65	0.39	0.98	0.86
m/z 181	-0.19	0.38	0.59	0.55
2-O-salicyloylsalicin	-0.36	0.26	0.52	0.63
2,6-cyclohexadiene-1,2-diol	-0.45	0.29	0.69	0.55
6-hydroxy-2-cyclohexenone alcohol	-0.52	0.38	0.80	0.71
α -salicyloylsalicin	-0.55	0.40	0.87	0.88
benzoyl-gentisyl alcohol	-0.34	0.29	0.54	0.50
benzoyl-salicyloylsalicin	-0.44	0.33	0.62	0.72
benzyl-coumaroyl-glucoside	-0.53	0.31	0.76	0.74
catechol	-0.53	0.36	0.80	0.77

Trait	Lignin content	S/G	Sum metabolite ions in MBMS*	Aromatics by GC/MS
coumaroyl-tremuloidin	-0.45	0.25	0.67	0.65
phenethyl-tremuloidin	-0.45	0.23	0.65	0.63
salicortin	-0.50	0.33	0.83	0.81
salicyl-salicylic acid-2-O-glucoside	-0.40	0.30	0.66	0.75
salicylic acid	-0.45	0.25	0.68	0.61
salicyloyl-coumaroyl-glucoside	-0.58	0.37	0.85	0.86
salicyltremuloidin	-0.45	0.30	0.68	0.83
salireposide	-0.42	0.26	0.68	0.67
Sum metabolite ions in MBMS*	-0.66	0.39	1.00	0.87
Aromatics by GC/MS	-0.54	0.32	0.87	1.00
tremulacin	-0.52	0.38	0.77	0.85
tremuloidin	-0.40	0.24	0.56	0.55
trichocarpin	-0.43	0.32	0.68	0.67

The total aromatic fraction of secondary metabolites in the test set from Boardman, OR, were predicted based on the simplified py-MBMS ion summation method and validated by GC/MS analysis of the extract. After GC/MS metabolic profiles were generated, PLS models of the test set were also built to further validate the use of the simplified ion method. There is a correlation between the total abundance of the aromatic metabolites as determined by GC/MS and the sum of the aromatic metabolite ions determined by MBMS (Fig. 3a), with an R^2 of 0.78 and Pearson correlation coefficient of 0.88 (Table 3) for the second test set of leaves, indicating that this simplified method is reasonably capable of predicting the relative abundance of aromatics in poplar leaf extracts across data sets.

Additionally, similarities in correlations between the two data sets for py-MBMS ions, metabolites and other compositional features include positive correlations between aromatic-derived ions, such as 105 and 122, with many salicylate metabolites (Table 3). The value of correlation coefficients for different traits differed to some degree between sets though, due, in part, to the differences in the actual metabolites detected and analyzed by GC/MS in the two sets. While these differences and metabolites themselves could not be speciated amongst these two sets, it was still possible to reasonably predict total aromatic metabolites in the two sets using the simplified py-MBMS ion summation method as

demonstrated in Fig. 3a. Additionally, the relative abundance of aromatic metabolites based on ranking GC/MS abundance reasonably correlates with the ranking determined by py-MBMS (Fig. 3b).

The PLS model for the test set from Boardman, OR (Fig. 4) was driven by similar aromatic-derived ions as the training set, but was different enough that the PLS models developed by one set could not be used to accurately predict aromatics in the other set. Interestingly, the PLS model for the test performed better than the model developed for the training set. These results further validate that PLS models would be similar but not directly translatable between different data sets and hence the need for the simplified ion intensity summation approach for screening purposes.

Table 3

Correlations between select ions from MBMS spectra of *P. trichocarpa* leaves in the test set from Boardman, OR and select metabolites as determined by GC/MS as well as lignin content, S/G ratios and the summation of metabolite-derived ions from MBMS spectra and the total aromatic metabolites determined by GC/MS. *metabolite ion intensities summed 68, 71, 77, 91, 94, 105, 107, 108, and 122

Trait	Lignin Content	S/G	Sum metabolite ions in MBMS*	Aromatics by GC/MS
m/z 122	-0.15	0.10	0.96	0.87
m/z 108	-0.05	0.06	0.86	0.87
m/z 107	-0.03	0.00	0.93	0.87
m/z 106	-0.10	0.11	0.88	0.82
m/z 105	-0.15	0.08	0.95	0.81
m/z 94	-0.04	0.02	0.74	0.69
m/z 91	-0.04	-0.01	0.54	0.51
m/z 78	-0.07	0.01	0.79	0.68
m/z 77	-0.12	0.07	0.97	0.87
m/z 68	-0.07	-0.05	0.74	0.58
m/z 66	-0.07	0.01	0.72	0.63
1,2,3-benzenetriol	-0.11	0.06	0.43	0.50
1,2,4-benzenetriol	-0.09	0.05	0.55	0.64
2,5-dihydroxybenzoic acid-2-O-glucoside	-0.01	0.00	0.48	0.51
Aromatics by GC/MS	-0.13	0.09	0.88	1.00
benzoyl-salicyloylsalicin	-0.12	0.06	0.67	0.49
catechol	-0.03	0.02	0.52	0.62
phenol	-0.06	-0.01	0.64	0.68
salicin	-0.01	0.03	0.61	0.72
salicyl alcohol	-0.09	0.07	0.36	0.41
salicyl-coumaroyl-glucoside	0.04	-0.11	0.70	0.62
salicylic acid	-0.04	0.00	0.63	0.44
salicyltremuloidin	-0.10	0.05	0.68	0.46
Sum metabolite ions in MBMS*	-0.13	0.05	1.00	0.88

Trait	Lignin Content	S/G	Sum metabolite ions in MBMS*	Aromatics by GC/MS
tremulacin	-0.12	0.03	0.63	0.47
tremuloidin	-0.18	0.16	0.55	0.78
trichocarpin	-0.06	-0.01	0.49	0.50
trichocarpin conjugate	-0.11	0.07	0.71	0.69
trichocarpinene	-0.09	0.00	0.50	0.48

Discussion

The variation in sugars vs aromatics in poplar leaves is different, but parallel to that seen in woody tissue where sugar and lignin (also aromatic, primarily phenolic) abundances are negatively correlated and the main drivers of compositional variation (9, 10). The leaf aromatics can be distinguished from lignin-derived phenolics because higher molecular weight, known lignin-derived ions, such as m/z 124, 137, 154, 167, 180, 210 (5) are otherwise not strongly correlated and/or are negatively correlated with the aromatic ions derived from leaves, and were, therefore, m/z 68, 71, 77, 91, 94, 105, 107, 108, and 122 were considered as primarily originating from secondary aromatic metabolites. The use of m/z 68, 71, 77, 91, 94, 105, 107, 108, and 122 to estimate the relative abundance or ranking of aromatic components in samples across two different sets of almost 500 poplar leaf samples demonstrates a simple and robust method to screen samples for rapid secondary metabolite insight and for prioritization of samples for in-depth metabolomics pipelines.

Conclusions

We have reported a rapid technique that can be used to analyze and screen biomass tissue for the relative abundance total aromatic metabolites. Error associated with this method could result from the relative abundance of different aromatic metabolites, and further studies and computational advances could potentially enable deconvolution of spectral features to provide resolved abundances of specific metabolites. If *a priori* metabolic information is available for standards or a range of samples within given data sets, PLS models specific to that data set could also be used to predict relative aromatic metabolite abundances. These methods could be used to screen large populations of biomass for aromatic metabolites to inform GWAS analyses, quantitative trait loci mapping, and development of metabolomic and systems biology models. The metabolite information obtained may also be used to inform sustainability metrics and gene x environment interactions in biomass studies. Rapid metabolomic analyses would ultimately enable the rational design of sustainable biomass used for biorefinery and bioenergy applications.

Methods

Collection and Preparation of Leaf Samples

Leaves from 851 *P. trichocarpa* genotypes established in a common garden at Clatskanie, OR in 2009 (18, 19) were sampled over a 3-d period in July 2012, as described elsewhere (20), from which 219 samples were selected for inclusion in the present study. Additionally, 60 *P. trichocarpa* genotypes growing in a drought stress trial at Boardman, OR were sampled on July 11 and 12, 2018, including 2 replicate trees per genotype in a well-irrigated plot (44.7 cm in each the 2017 and 2018 growing seasons) and 2 replicate trees growing in a drought stress plot receiving 60% (26.8 cm) the irrigation level of the well-irrigated plot. Trees were established in the spring of 2016, with the irrigation manipulation applied during the second (2017) and third (2018) growing seasons. Similar to the leaves sampled at the Clatskanie, OR, a fully expanded leaf of leaf plastochron index 9 ± 1 was rapidly collected, placed on dry ice, shipped back to Oak Ridge National Laboratory, stored at -80°C until being processed for metabolite analyses.

Extraction and Analysis of Leaf Metabolites by GC/MS

For leaf samples collected at Clatskanie, OR, metabolite extraction and analysis were prepared as described previously (21). For leaf samples collected from genotypes at Boardman, OR, metabolite extraction and analysis were performed based on protocols also described previously (22). Briefly, lyophilized samples were ground and extracted twice overnight with 2.5 mL portions of 80% ethanol. Sorbitol ($75\ \mu\text{L}$; $1\ \text{mg mL}^{-1}$) was added before extraction as an internal standard. Following centrifugation, a $500\ \mu\text{L}$ aliquot of the combined extract was dried in a stream of nitrogen and silylated by addition of $500\ \mu\text{L}$ of acetonitrile (TS-20062; ThermoFisher) and $500\ \mu\text{L}$ of N-methyl-N-trimethylsilyltrifluoroacetamide (MSTFA) with 1% trimethylchlorosilane (TMCS) (TS-48915; ThermoFisher) and heating for 1 h at 70°C to generate trimethylsilyl (TMS) derivatives. After 2 days, a 0.1 or $1\ \mu\text{L}$ aliquot was injected into an Agilent Technologies Inc. (Santa Clara, CA) 7890A gas chromatograph coupled to a 5975C inert XL mass spectrometer configured (23), and aromatic metabolites were identified, quantified, and normalized as previously described (23).

The fraction of the aromatic metabolites in the DW extract of leaves was obtained by summing the abundance of each aromatic metabolite ($\mu\text{g/g}$ DW internal standard equivalent) and dividing by the total sum of the metabolites accounted for (DW internal standard equivalent) in each sample.

Pyrolysis-MBMS

A Frontier PY2020 unit pyrolyzed 4 mg of cryo-milled biomass at 500°C for 30 s in $80\ \mu\text{L}$ deactivated stainless steel cups. Each biomass sample was analyzed in duplicate. An Extrel Super-Sonic MBMS Model Max 1000 was used to collect mass spectra (m/z 30–450 at 17 eV) which was processed using Merlin Automation software (V3). Spectral ion intensities were normalized based on summation of total ion intensities (TIC) and spectra were subsequently analyzed using The Unscrambler X v10 (Camo, Trondheim, Norway) and R Studio (ref).

The training sample set of fully expanded leaves from *P. trichocarpa* genotypes was collected in the summer of 2012 from trees of a GWAS population established in a common garden at Clatskanie, OR in 2009. For this study, a total of 219 samples from individual genotypes were selected that varied in total and specific aromatic metabolite concentrations based on GC/MS analysis of foliar extracts. The test set consisted of 223 *P. trichocarpa* similar leaf samples collected from trees growing under different irrigation conditions at Boardman, OR in the summer of 2018 in their second year. With Clatskanie, OR, being a mesic site in the Columbia River Delta with a high water table, trees were grown without supplemental irrigation. With Boardman, OR, typically only receiving approx. 18 cm of rainfall annually, trees were supplemented with irrigation water. Surprisingly, there were no significant differences in the total metabolite concentrations between drought-stressed and well-watered trees within a given genotype (data not shown). Based on the analysis of the training set, it was determined that ions (m/z) 68, 71, 77, 91, 94, 105, 107, and 122 were highly correlated to the abundance of aromatic metabolites and were used to predict the relative abundance of total aromatic metabolites that would have been present in the extracts of the test set. Partial Least Squares (PLS) regression models constructed from the GC/MS and py-MBMS data (restricted to m/z 50–250) of the two different sets were generated using the Unscrambler X V.10.5 (Camo Software). Data were mean centered with complete cross validation using the NIPALS algorithm.

Declarations

Funding

This research was supported by the U. S. Department of Energy (DOE), Office of Energy Efficiency and Renewable Energy (EERE), Bioenergy Technologies Office (BETO), under Award No. DE-AC36-08GO28308 with the National Renewable Energy Laboratory.

Funding was also provided by the Center for Bioenergy Innovation (CBI), from the U.S. Department of Energy Bioenergy Research Centers supported by the Office of Biological and Environmental Research in the DOE Office of Science. This manuscript has been authored or coauthored by UT-Battelle, LLC under Contract No. DE-AC05-00OR22725 with the U.S. Department of Energy. The publisher, by accepting the article for publication, acknowledges that the U. S. Government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this work, or allow others to do so, for U.S. Government purposes. The views expressed in the article do not necessarily represent the views of the U.S. Department of Energy or the United States Government.

Acknowledgements

The authors would like to thank Sean Cornell and Andrew Grewe for assistance with sample preparation, and to Brian Stanton, Katherine Haiby, and Richard Shuren of Greenwood Resources, Portland, OR, for establishment and maintenance field sites. Also, thanks to Jonathan Cumming and Stephen P. DiFazio for experimental assistance and design, management and insight.

Data availability

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>). Data not presented in the manuscript is available upon request from the coauthors.

Author Contributions

AEW performed experiments, analyzed data, wrote text, constructed figures, and edited manuscript, MM provided data analysis, NLE performed experiments, collected and prepared samples, provided text and editing of manuscript, CD performed experiments, TJT performed experiments, collected and prepared samples, provided text and editing of manuscript as well as oversight and poplar field experimental design.

Ethics approval and consent to participate: Not applicable

Consent for publication: Not applicable

Competing Interests: The authors declare no competing interests

References

1. Tuskan GA, Difazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, et al. The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *science*. 2006;313(5793):1596–604.
2. Tschaplinski TJ, Abraham PE, Jawdy SS, Gunter LE, Martin MZ, Engle NL, et al. The nature of the progression of drought stress drives differential metabolomic responses in *Populus deltoides*. *Ann Bot*. 2019;124(4):617–26.
3. Tsai CJ, Guo W, Babst B, Nyamdari B, Yuan Y, Payyavula R, et al. Salicylate metabolism in *Populus*. *BMC Proceedings*. 2011;5(7):19.
4. Ostos-Garrido FJ, de Castro AI, Torres-Sánchez J, Pistón F, Peña JM. High-Throughput Phenotyping of Bioethanol Potential in Cereals Using UAV-Based Multi-Spectral Imagery. *Front Plant Sci*. 2019;10:948.
5. Decker SR, Harman-Ware AE, Happs RM, Wolfrum EJ, Tuskan GA, Kainer D, et al. High Throughput Screening Technologies in Biomass Characterization. *Frontiers in Energy Research*. 2018;6.
6. Corol ID, Harflett C, Beale HM, Ward LJ. An Efficient High Throughput Metabotyping Platform for Screening of Biomass Willows. *Metabolites*. 2014;4(4).
7. Hall R, Beale M, Fiehn O, Hardy N, Sumner L, Bino R. Plant Metabolomics. *Plant Cell*. 2002;14(7):1437.
8. Ainsworth EA, Gillespie KM. Estimation of total phenolic content and other oxidation substrates in plant tissues using Folin–Ciocalteu reagent. *Nat Protoc*. 2007;2(4):875–7.

9. Sykes R, Kodrzycki B, Tuskan G, Foutz K, Davis M. Within tree variability of lignin composition in *Populus*. *Wood Sci Technol*. 2008;42(8):649–61.
10. Sykes R, Yung M, Novaes E, Kirst M, Peter G, Davis M. High-Throughput Screening of Plant Cell-Wall Composition Using Pyrolysis Molecular Beam Mass Spectroscopy. In: Mielenz JR, editor. *Biofuels: Methods and Protocols*. Totowa: Humana Press; 2009. pp. 169–83.
11. Yin T, Zhang X, Gunter L, Priya R, Sykes R, Davis M, et al. Differential Detection of Genetic Loci Underlying Stem and Root Lignin Content in *Populus*. *PLOS ONE*. 2010;5(11):e14021.
12. Studer MH, DeMartini JD, Davis MF, Sykes RW, Davison B, Keller M, et al. Lignin content in natural *Populus* variants affects sugar release. *Proceedings of the National Academy of Sciences*. 2011;108(15):6300-5.
13. Evans RJ, Milne TA. Molecular characterization of the pyrolysis of biomass. *Energy Fuels*. 1987;1(2):123–37.
14. Faix O, Meier D, Fortmann I. Pyrolysis-gas chromatography-mass spectrometry of two trimeric lignin model compounds with alkyl-aryl ether structure. *J Anal Appl Pyrol*. 1988;14(2):135–48.
15. Decker SR, Sykes RW, Turner GB, Lupoi JS, Doepkke C, Tucker MP, et al. High-throughput Screening of Recalcitrance Variations in Lignocellulosic Biomass: Total Lignin, Lignin Monomers, and Enzymatic Sugar Release. *JoVE*. 2015(103):e53163.
16. Mann DGJ, Labbé N, Sykes RW, Gracom K, Kline L, Swamidoss IM, et al. Rapid assessment of lignin content and structure in switchgrass (*Panicum virgatum* L.) grown under different environmental conditions. *BioEnergy Research*. 2009;2(4):246–56.
17. Penning BW, Sykes RW, Babcock NC, Dugard CK, Klimek JF, Gamblin D, et al. Validation of PyMBMS as a High-throughput Screen for Lignin Abundance in Lignocellulosic Biomass of Grasses. *BioEnergy Research*. 2014;7(3):899–908.
18. Evans LM, Slavov GT, Rodgers-Melnick E, Martin J, Ranjan P, Muchero W, et al. Population genomics of *Populus trichocarpa* identifies signatures of selection and adaptive trait associations. *Nat Genet*. 2014;46(10):1089–96.
19. Slavov GT, DiFazio SP, Martin J, Schackwitz W, Muchero W, Rodgers-Melnick E, et al. Genome resequencing reveals multiscale geographic structure and extensive linkage disequilibrium in the forest tree *Populus trichocarpa*. *New Phytol*. 2012;196(3):713–25.
20. Weighill D, Jones P, Shah M, Ranjan P, Muchero W, Schmutz J, et al. Pleiotropic and Epistatic Network-Based Discovery: Integrated Networks for Target Gene Discovery. *Frontiers in Energy Research*. 2018;6(30).
21. Weighill D, Jones P, Bleker C, Ranjan P, Shah M, Zhao N, et al. Multi-Phenotype Association Decomposition: Unraveling Complex Gene-Phenotype Relationships. *Front Genet*. 2019;10:417.
22. Abraham PE, Yin H, Borland AM, Weighill D, Lim SD, De Paoli HC, et al. Transcript, protein and metabolite temporal dynamics in the CAM plant *Agave*. *Nature Plants*. 2016;2(12):16178.
23. Tschaplinski TJ, Standaert RF, Engle NL, Martin MZ, Sangha AK, Parks JM, et al. Down-regulation of the caffeic acid O-methyltransferase gene in switchgrass reveals a novel monolignol analog.

Figures

Figure 1

(caption not included with figure)

Figure 2

(caption not included with figure)

Figure 3

(caption not included with figure)

Figure 4

(caption not included with figure)

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [HTPmetpopleavesSuppTable1docx.docx](#)