Bioenerg. Res.
DOI 10.1007/s12155-014-9410-3

Validation of PyMBMS as a High-throughput Screen for Lignin Abundance in Lignocellulosic Biomass of Grasses

Bryan W. Penning · Robert W. Sykes · Nicholas C. Babcock · Christopher K. Dugard · John F. Klimek · David Gamblin · Mark Davis · Timothy R. Filley · Nathan S. Mosier · Clifford F. Weil · Maureen C. McCann · Nicholas C. Carpita

Abstract Pyrolysis molecular-beam mass spectrometry (PyMBMS) was tested as a high-throughput method for relative abundances of guaiacyl and syringyl lignin in lignocellulosic cell-wall materials from stems of a population of maize intermated B73×Mo17 (IBM) recombinant inbred lines. Variations of up to twofold across the population in phenylpropanoid abundance were observed. Several histochemical and quantitative biochemical assays were used to validate the mass spectrometric data for lignin, hydroxycinnamic acids, crystalline cellulose, non-cellulosic glucans, and xylans. We demonstrate PyMBMS to be a valid high-throughput screen suitable for analysis of lignin abundance in large populations of bioenergy grasses. Pentose from xylans and hexose from cellulosic and non-cellulosic glucans also varied substantially across the population, but abundances of diagnostic fragments for these monosaccharides were not well correlated with the abundance of cell-wall polysaccharides.

Keywords Maize · Cell walls · Pyrolysis molecular-beam mass spectrometry · Lignin · *p*-Coumaric acid · Cellulose · Xylan

Abbreviations

PyMBMS	Pyrolysis molecular-beam mass spectrometry
IBM	Intermated B73×Mo17 recombinant inbred population
GAX	Glucuronoarabinoxylan
G-lignin	Guaiacyl-lignin
H-lignin	*p*-Hydroxyphenol-lignin
RIL	Recombinant inbred line
QTL	Quantitative trait locus (loci)
S-lignin	Syringyl-lignin

Introduction

Lignocellulosic biomass, predominantly the cell walls of plants, is a renewable and sustainable source of partially reduced carbon for conversion to biobased liquid transportation fuels. Both annual and perennial C4 energy grasses, such as *Miscanthus*, switchgrass, sorghum, and tropical varieties of maize, are expected to contribute significantly to the portfolio of biomass feedstocks for the production of biofuels [1]. Considerable variation in the abundance of lignin and cellulosic and non-cellulosic polysaccharides is found within a given species. For maize the genetic diversity in the natural population is captured to a large degree in nearly 3,000 inbred

Electronic supplementary material The online version of this article (doi:10.1007/s12155-014-9410-3) contains supplementary material, which is available to authorized users.

B. W. Penning · M. C. McCann · N. C. Carpita
Department of Biological Sciences, Purdue University, 915 West State Street, West Lafayette, IN 47907, USA

B. W. Penning · C. K. Dugard · J. F. Klimek · N. C. Carpita (✉)
Department of Botany & Plant Pathology, Purdue University, 915 West State Street, West Lafayette, IN 47907-2053, USA
e-mail: carpita@purdue.edu

R. W. Sykes · M. Davis
Bioscience Center, National Renewable Energy Laboratory, 15013 Denver West Parkway, Golden, CO 80401, USA

N. C. Babcock · C. F. Weil
Department of Agronomy, Purdue University, 915 West State Street, West Lafayette, IN 47907, USA

D. Gamblin · T. R. Filley
Department of Earth, Atmospheric and Planetary Sciences, Purdue University, 550 Stadium Mall Drive, West Lafayette, IN 47907, USA

N. S. Mosier
Department of Agricultural & Biological Engineering, Purdue University, 225 South University Street, West Lafayette, IN 47907, USA

Published online: 30 January 2014

lines and landraces [2] and subsets of several hundreds of them that represent a bulk of this diversity in more manageable populations [3–5]. Genetic resources include recombinant inbred (RIL) populations [6, 7] and 26 nested association mapping populations of 200 RILs each [8] that present opportunities to determine quantitative trait loci (QTL), the chromosomal regions with high probability to contain gene(s) that contribute to desirable traits. However, use of these valuable genetic resources to identify genes that contribute to carbohydrate and phenylpropanoid abundance and composition requires a reliable, high-throughput screen to quantify cell-wall constituents.

PyMBMS is a high-throughput thermochemical technique that yields relative abundance of mass/charge (*m*/*z*) diagnostic fragments of monomeric units of phenylpropanoid-, hydroxycinnamic acid-, and carbohydrate-containing macromolecules of cell walls [9, 10]. PyMBMS has been validated for use to screen lignin abundance in large populations of genetically variant poplar and pine [11, 12]. It was also employed previously to analyze differences in lignocellulosic composition among leaves and stems of switchgrass (*Panicum virgatum* L. cv. Alamo) grown under field, greenhouse, and growth chamber conditions [13].

The PyMBMS profiles of grasses and woody species differ, consistent with a broad distinction between dicot species and grasses with respect to cell-wall composition. Application of PyMBMS in a genetic screen for biomass quality traits in grasses requires a validation of the association of putative diagnostic fragments with the molecular composition, the special plant cell wall of these species. In 'type II' cell walls of grasses, glucuronoarabinoxylans (GAXs) are the major cross-linking glycans that interlace cellulose microfibrils in primary and secondary walls [14, 15]. Small amounts of xyloglucans and glucomannans also form the cross-linking matrix. Pectic polysaccharides, homogalacturonan, and branched rhamnogalacturonan I, comprise a much smaller proportion of the wall compared to dicots and monocots with 'type I' walls [14]. Hydroxycinnamic acids, such as ferulic and *p*-coumaric acid, are esterified to arabinosyl units of GAX and connect this polysaccharide matrix to networks of phenylpropanoids [15]. With the exception of a high abundance of hydroxycinnamic acids in commelinid monocots, the lignins of all angiosperm species comprise guaiacyl (G), syringyl (S), and *p*-hydroxyphenyl (H) units in widely different proportions [16, 17].

Because of the broad differences in composition between phenolic-rich primary and secondary walls in grasses compared to dicotyledonous species, we evaluated PyMBMS for carbohydrate and aromatic abundance in the maize IBM population. While the parents used in the original cross, Mo17 and B73, differ in composition, the RILs derived from intermating these parental lines show much greater compositional variation, a phenomenon termed 'transgressive segregation'. We found that, with few exceptions, assignment of the *m*/*z* deriving from specific monolignol components is consistent with those previously reported for woody plants but the relative proportions can vary substantially [9, 10, 12]. We validated the PyMBMS data for phenylpropanoid abundance by several histochemical and biochemical techniques, establishing cellulose content, non-cellulosic sugar distributions, saponifiable hydroxycinnamic acids, and monolignol distribution. These results demonstrate PyMBMS to be a valid high-throughput screen suitable for the analysis of the composition of lignin in large populations of bioenergy grasses, but complementary protocols need to be implemented to survey in medium- and high-throughput, the hydroxycinnamic acids, crystalline cellulose, non-cellulosic glucose, and xylose constituents of grass biomass.

Materials and Methods

Maize Stover Collection and Processing

The maize Mo17 and B73 parents and the IBM population were grown at the University of Illinois field station (Champaign–Urbana, IL). An association panel of 282 inbreds and landraces was grown at the Purdue University Agricultural Center for Research and Education. Total weekly growing degree days and precipitation were gathered from the Illinois State Water Survey (http://www.isws.illinois.edu/atmos/statecli/cuweather/). Growing degree days were calculated as the average of the daily maximum, up to 86 °F, and minimum temperatures minus the maize base growth temperature of 50 °F.

The lower 70 cm of field-dried maize stover was collected after ear harvest and placed in 40-°C air dryers for 1–2 weeks. After drying, approximately 5-cm sections of the internode with brace roots of five plants were cut, measured, and weighed. The samples were ground to less than 0.5-mm particle size in a Wiley Mill with a 40-mesh screen. About 3 g of ground stover was placed into a 15-mL Falcon tube and washed twice with 10 mL of 50 % (*v*/*v*) ethanol in water at 70 °C to remove water-soluble material. The ethanol was removed by five washes with 10-mL distilled water at an ambient temperature. Samples were centrifuged in a swinging bucket rotor at 2,500 RPM for 5 min, and the liquid was decanted. In between centrifuge steps, the cell-wall samples were resuspended in each washing step by vortex mixing. After the final wash, the cell-wall material was resuspended in water and freeze-dried.

PyMBMS

PyMBMS was carried out at NREL using a modified protocol from Sykes et al. [18]. Briefly, about 4 mg of freeze-dried

stover was added to an 80-μL stainless steel cup and pyrolyzed at 500 °C with a 0.9-L min^{-1} helium flow at standard temperature and pressure. Data acquisition occurs over a 90-s period with full pyrolysis product evolution occurring within the first 30 s. A Merlin data acquisition system with 17-eV electron ionization was used to gather mass spectral data from *m/z* of 30–450 using a 0.5-s scan rate. Spectra are reported as an average total ion count of the 60 spectra obtained during the 30-s product evolution period for each sample.

Statistical Analyses

Pearson's correlations [19] were carried out using the SAS (Version 9.2; SAS Institute Inc.) *proc corr* procedure and a matrix histogram plot. Data generated were normalized to correct for skewing as a result of instrument changes between pyrolysis runs in 2009. Normality was tested in SAS using the *proc univariate* procedure using the Shapiro–Wilk test [20], as appropriate for the sample sizes tested.

Phloroglucinol Staining

Phloroglucinol staining was carried out using a modified protocol from Sindhu et al. [21]. Fresh maize stems were frozen in polyethylene glycol, frozen at −80 °C for 30 to 45 min and cut in 100-μm thick sections using a Microm HM550 microtome. Cut slices were thawed and stained with Wiesner's solution (95:5 (*v*/*v*) of 2 % phloroglucinol in ethanol (*w*/*v*):50 % HCl in water (*v*/*v*)) for 2 min and then washed briefly in the same solution without phloroglucinol. Images were captured using Spot Advanced software version 4.1 with a Spot Insight 4 camera attached to a Nikon SMZ21500 microscope at magnifications listed in the figure.

Klason Lignin

Klason lignin determinations were performed as described by Theander and Westerlund [22] as modified [23]. Briefly, 50± 0.2 mg of cell-wall material was suspended in ice-cold 12 M (mol dm^{-3}) H_2SO_4 and incubated at ice temperature for 30 min, then diluted to 1 M with water and incubated at 120 °C to hydrolyze cellulosic and non-cellulosic polysaccharides. Insoluble lignin residue was filtered with pre-weighed glass-fiber filters and oven-dried at 105 °C for 24 h and weighed. Samples on filters were heated in a muffle furnace at 500 °C for 5 h, cooled and the filters reweighed to obtain lignin weights.

Saponifiable Hydroxycinnamic Acids

To determine total saponifiable hydroxycinnamates and other phenolic substances, 10±0.2 mg of isolated walls was suspended in 1 mL of 1 M NaOH containing 10 nmol of 3,4,5-trimethyl-*trans*-cinnamic acid (internal standard) and incubated for 30 min at 42 °C. After cooling, 1 mL of 3 M HCl was added to give 1 M in excess of NaOH, and the phenolic substances were partitioned into ethyl acetate. The ethyl acetate was evaporated and the residues resuspended in 1 mL of 50 % (*v*/*v*) methanol in water. The phenolic substances were separated in a 3.0×75-mm reverse-phase column of Microsorb-MV C18 (Rainin) in a gradient of 2.0 % (*v*/*v*) formic acid in water to 25 % acetonitrile in 0.1 % (*v*/*v*) formic acid in water at 0.9 mL min^{-1}. Ferulic acid and *p*-coumaric acid were quantified by UV absorbance at 330 nm based on standards treated similarly [24].

Copper Oxide Oxidation Determination of Lignin Monomers

Alkaline cupric-oxide (CuO) oxidation followed by GC/MS was used to extract and quantify lignin-derived phenols [25]. The CuO extractions utilized Monel reaction vessels (Prime Focus, Inc. Seattle, WA, USA) and followed the original method of Hedges and Ertel [26] with slight modifications [27]. Ethyl vanillin and DL-12 hydroxystearic acid were added as internal quantitation standards. Lignin-derived phenols were derivatized as trimethylsilane (TMS) derivatives of vanillyl [G]-based (i.e., vanillin, acetovanillone, and vanillic acid), syringyl [S]-based (i.e., syringaldehyde, acetosyringone, and syringic acid), and cinnamyl (Ci)-based (i.e., *p*-hydroxycinnamic and ferulic acids) monomers and quantified using extracted-ion internal calibration curves.

Duplicate CuO analyses were performed for all samples. A reference sample of peach leaf (NIST 1547) was used in each batch of analyses to monitor the consistency of the reaction and GC performance. Mean standard reproducibility for the analytical method was 2–5 % for lignin phenols. Derivatives were separated on a 0.25-mm×30-m column of RTS-5 MS (Restek, Bellefonte, PA). After an initial temp of 60 °C, the oven ramped to 300 °C at 7 °C min^{-1}, with a hold time of 7 min at the upper temp. Derivatives were identified and quantified by electron-impact mass spectrometry using extracted-ion calibration curves. The amounts of G-, S-, and H-lignin-derived phenols and hydroxycinnamate concentration were quantified as a percent of cell-wall mass [25].

Carbohydrate Analysis

Samples (5±0.2 mg) of isolated cell walls were hydrolyzed in 1 mL of 2-M trifluoroacetic acid (TFA) containing 0.5 μmol of *myo*-inositol (internal standard) at 120 °C for 90 min in a conical centrifuge tube, cooled to ambient temperature, and undigested cellulose and other material were pelleted by centrifugation. The supernatant liquid was transferred to a fresh tube, and after the addition of 1.0 mL of liquefied *tert*-butyl alcohol, the supernatants were evaporated in a stream of nitrogen. The cellulosic pellet was washed several times with

water and suspended in 1.0 mL of water. Glucose equivalents were determined by phenol-sulfuric assay [28]. Crystalline cellulose was also measured by acetic-nitric digestion [29].

Hydrolyzed sugars were dissolved in water, and samples were separated isocratically in 20 mM NaOH on a CarboPacPA-10 anion-exchange column (0.4×25 cm; Dionex) and detected and quantified by pulsed amperometric detection. Alditol acetate derivatives were prepared from a subset of these samples as described previously [30]. Derivatives were separated by gas–liquid chromatography (GLC) on a 0.25-mm×30-m column of SP-2330 (Supelco, Bellefonte, PA). Temperature was held at 80 °C during injection, then ramped quickly to 170 °C at 25 °C min^{-1}, and then to 240 °C at 5 °C min^{-1} with a 10-min hold at the upper temperature. Helium flow was 1 mL min^{-1} with splitless injection. The electron impact mass spectrometry (EIMS) was performed with a Hewlett–Packard MSD at 70 eV and a source temperature of 250 °C.

Results

Parental lines of maize B73 and Mo17 were derived from the stiff-stalk and non-stiff-stalk groups, respectively [3]. Our intermated B73×Mo17 (IBM) recombinant inbred test population consisted of 264 lines and the B73 and Mo17 parents, all grown at the University of Illinois agronomy field station over three growing seasons. Senescent stems of internodes 3 and 4 above the uppermost prop roots were milled to pass a 40-mesh screen and washed extensively with water and 50 % (*v*/*v*) ethanol in water (60 °C) to remove small molecules that would interfere with biomass analysis. The insoluble material was then freeze-dried for analysis.

As assigned previously for poplar and other materials (Table S1) [9, 11], PyMBMS on 1-mg samples of this material gave spectra with fragment ions of between *m*/*z* 55 and 210 for derivatives of pentose (*m*/*z* 114), hexose (*m*/*z* 126), *p*-coumaric acid (*m*/*z* 147 and *m*/*z* 164) and related fragments (4-vinyl phenol, *m*/*z* 120), ferulic acid (*m*/*z* 194, shared with S-lignin), G-lignin (*m*/*z* 124, *m*/*z* 137, *m*/*z* 138, *m*/*z* 151, and *m*/*z* 178), and S-lignin (*m*/*z* 154, *m*/*z* 167, *m*/*z* 168, *m*/*z* 194, shared with ferulic acid, and *m*/*z* 210) (Fig. 1). PyMBMS spectra of the B73 and Mo17 parental lines were similar, but a difference spectrum indicated a small enrichment in B73 over Mo17 in mass fragments corresponding to total lignin, S-lignin, and G-lignin, and a decrease in 4-vinyl phenol (*m*/*z* 120), putatively diagnostic for *p*-coumaric acid, in *m*/*z* 114 diagnostic for pentose and in *m*/*z* 126 diagnostic for hexose (Fig. 1a and b).

We found the abundances of *m*/*z* 124, 137, 138, and 151, diagnostic for G lignin, and *m*/*z* 154, 167, 168, and 194, diagnostic for S lignin, to be normally distributed over 3 years and reasonably correlated with other lignin-related ion fragments (Fig. 2). Although abundant, fragments *m*/*z* 178 and *m*/*z* 210, typically related to G- and S-lignin, respectively, were not well correlated in our samples with the abundance of their related fragments. Pearson's correlations of the individual G- and S-lignin fragments to each other and to their totals were high (Fig. 2). Having substantiated that a set of four diagnostic fragments each were correlated with G- and S-lignin units, these values were summed to derive relative proportions of G-, S-, and total lignin when used in validation with chemical assays.

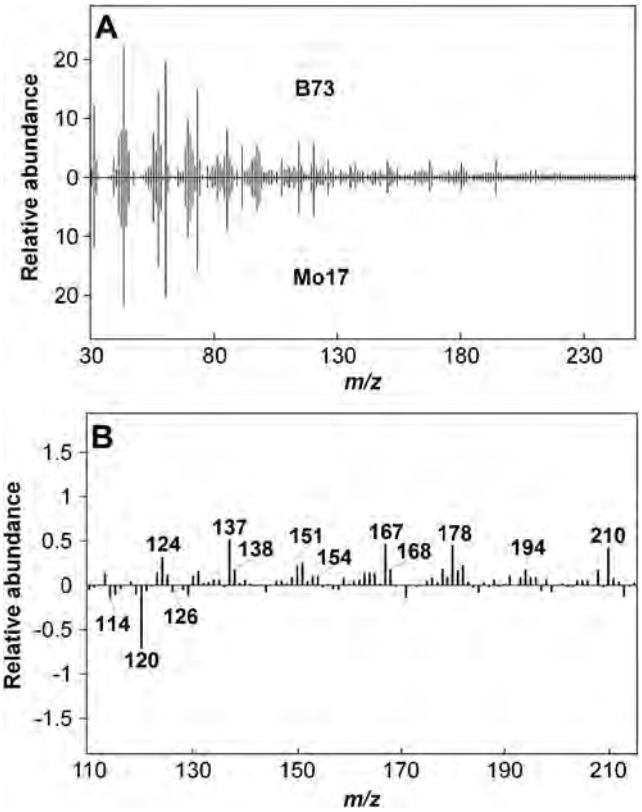

Fig. 1 PyMBMS provides relative abundances of phenylpropanoid and carbohydrates in maize cell-wall biomass. **a** PyMBMS spectra of stem cell-wall biomass from B73 (upper spectrum), Mo17 (lower spectrum), from *m*/*z* 30 to *m*/*z* 250. **b** Digital subtraction of Mo17 from B73 shows increases in fragments assigned to G-lignin (*m*/*z* 124, *m*/*z* 137, *m*/*z* 138, *m*/*z* 151, and *m*/*z* 178), and S-lignin (*m*/*z* 154, *m*/*z* 167, *m*/*z* 168, *m*/*z* 194, and *m*/*z* 210) decreases in carbohydrate (*m*/*z* 114, 126) and 4-vinylphenol (*m*/*z* 120) in B73

We also examined the correlation of fragments considered diagnostic for G- and S-lignin, *p*-coumaric acid, cellulose, and pentose (xylan) with each other. Weak positive correlations were found between syringyl lignin (*m*/*z* 167) and both guaiacyl lignin (*m*/*z* 137) and *p*-coumaric acid (*m*/*z* 120), but not between guaiacyl lignin and *p*-coumaric acid (Fig. 3a, b, and e). Weak negative correlations were found between hexose (*m*/*z* 126) and guaiacyl (*m*/*z* 137) and syringyl (*m*/*z* 167) lignin and *p*-coumaric acid (*m*/*z* 120; 4-vinyl phenol) (Fig. 3c, f, and h); weak negative correlations were also found between pentose (*m*/*z* 114) and syringyl lignin and *p*-coumaric acid, but a very weak positive correlation was seen between pentose

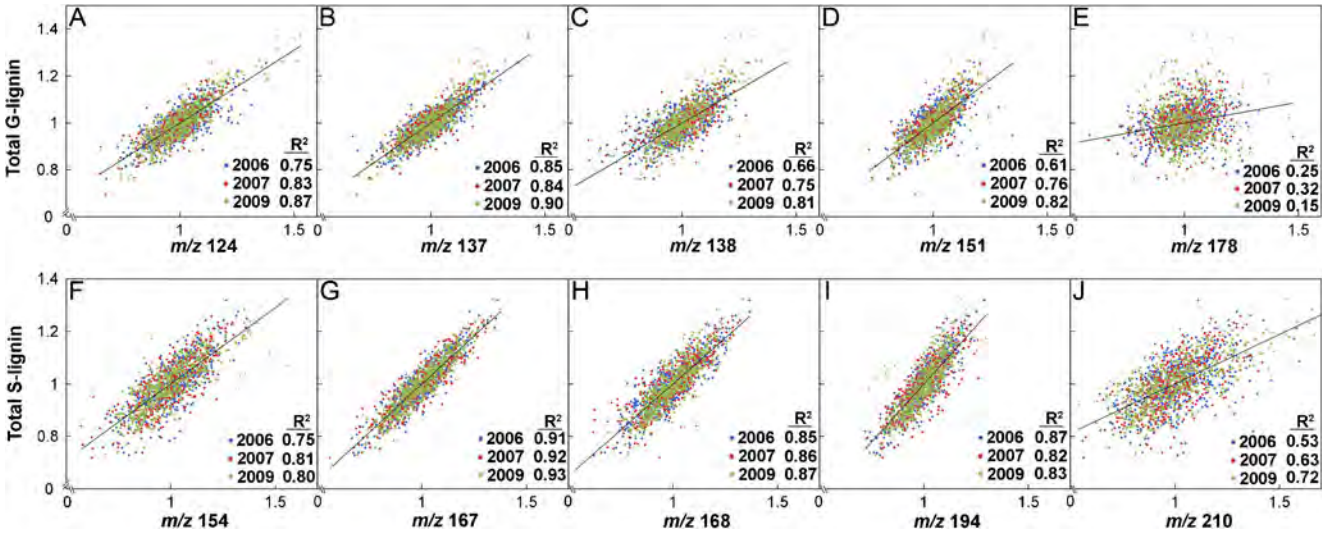

Fig. 2 Correlations of diagnostic fragments for G- and S-lignin for the IBM population. Comparative values are shown for diagnostic fragments normalized over 3 years of sampling. **a**–**e** Correlation of five diagnostic fragments for G-lignin (*m/z* 124 [0.205], *m/z* 137 [0.295], *m/z* 138 [0.159], *m/z* 151 [0.195], *m/z* 178 [0.146]) with the proportion of total G-lignin predicted from their sum. **f**–**j** Correlation of five diagnostic fragments for S-lignin (*m/z* 154 [0.164], *m/z* 167 [0.270], *m/z* 168 [0.150], *m/z* 194 [0.299], *m/z* 210[0.117]) with the proportion of total S-lignin predicted from their sum. Pearson's Correlation Coefficients (PCC) for each diagnostic fragment across the IBM population each year are inset. The average fractional proportion of each *m/z* fragment to the total G- and S-lignin is given in brackets

and hexose (Fig. 3j) and a very weak correlation in certain years was seen between pentose and guaiacyl lignin (Fig. 3g).

PyMBMS revealed up to twofold differences in relative abundance of diagnostic fragments of total lignin, *p*-coumaric acid (4-vinyl phenol), and hexose in the IBM population (Fig. 4a, c, and e). We also used the Shapiro–Wilk W test for normality appropriate for our data sample number to assess whether the RIL population was normally distributed for each diagnostic PyMBMS ion. Frequency distributions of three different *m/z* ions using the 2007 IBM data set with examples of Quantile–Quantile (Q–Q) plots illustrate that most samples fall on or near the normal distribution line, indicating that these ion abundances can be treated as quantitative traits (Fig. 4b, d, and f). We then employed several histochemical and biochemical assays to test the correlation of the PyMBMS diagnostic fragments with the respective abundances of G- and S-lignin, hydroxycinnamic acids, crystalline cellulose, and the non-cellulosic monosaccharides, glucose and xylose, across the range represented by the RIL population.

Lignin

Phloroglucinol reacts with aromatic aldehydes to yield either pink color with hydroxycinnamyl aldehydes or red-brown color with hydroxybenzaldehydes, indicative of the presence of both G- and S-lignin [31]. Stem internodes of IBM lines with significantly higher abundances of PyMBMS fragments predicted to be diagnostic for G-lignin and S-lignin gave more intense red staining than those predicted to be low lignin (Fig. 5).

The weight percentage of Klason lignin also correlated with the relative total lignin abundance derived from PyMBMS with sample representative of the range found in the IBM population (Fig. 6a). The abundance of G- and S-lignin derived from separation of copper-oxide oxidation products by GLC was also correlated with the total of the diagnostic fragments for each of these monolignol components determined by PyMBMS (Fig. 6b).

Hydroxycinnamic Acids

PyMBMS fragment *m/z* 164 is considered diagnostic for *p*-coumaric acid and an isoeugenol derivative of G-lignin [9, 10]. Secondary fragments of *m/z* 147 [M–17] and *m/z* 120 [M–44] are considered equally diagnostic without coincident fragments from other aromatic substances, although H-lignin of grasses could potentially contribute to these fragment abundances. Only weak correlation was found between the abundances of *m/z* 120 and the amounts of saponifiable *p*-coumaric acid from the maize IBM lines (Fig. 7). Fragment *m/z* 194 is assigned as diagnostic for ferulic acid, with contributions of dimethoxy-propenylphenol from S-lignin. Amounts of saponifiable *p*-coumaric acid average about 4.5-fold higher than those of ferulic acid across the range found in the IBM lines. No correlation between the relative abundance of *m/z* 194 and saponifiable ferulic acid was observed (not shown).

Polysaccharides

Fragment *m/z* 126 is assigned to hexose, and *m/z* 114 is diagnostic mainly of pentose with some secondary

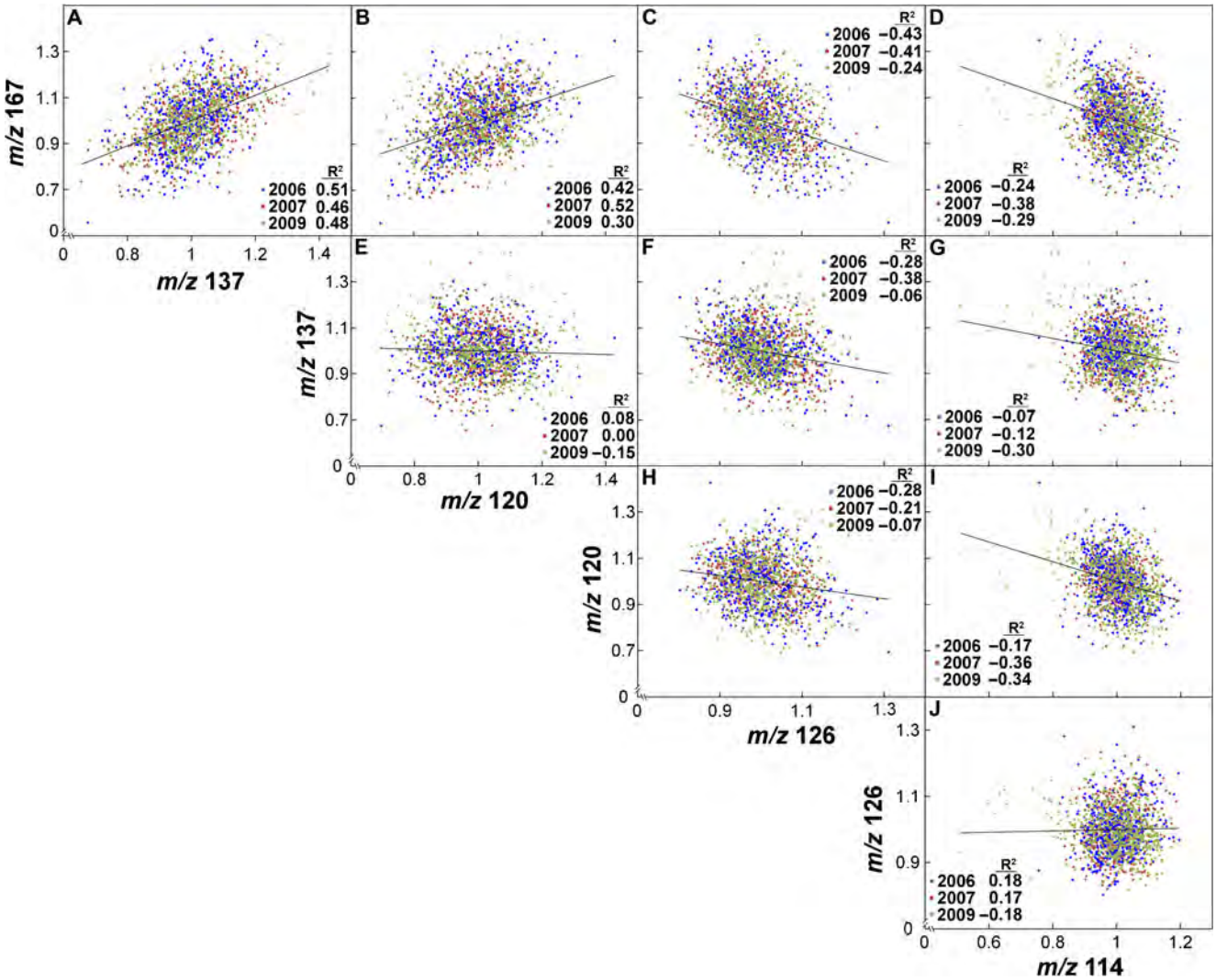

Fig. 3 Comparison of diagnostic fragments for G- and S-lignin, *p*-coumaric acid (4-vinyl phenol), crystalline cellulose, non-cellulosic glucose, and xylose. Comparative values are shown for diagnostic fragments for all 3 years of sampling. Relative abundances from PyMBMS are defined for G-lignin (*m/z* 137), S-lignin (*m/z* 167), *p*-coumaric acid (*m/z* 120), and cellulosic and non-cellulosic glucose (hexose; *m/z* 126) and xylose (pentose *m/z* 114). Pearson's Correlation Coefficients (PCC) for each diagnostic fragment across the IBM population each year are inset

fragmentation of hexose derivatives [9]. We assayed cellulose both as acetic-nitric acid insoluble, crystalline cellulose [29] and as the 2 M TFA-insoluble glucose equivalents in assays of non-cellulosic monosaccharides by GC–MS of their alditol acetates. No correlation was observed between *m/z* 126 abundance and cellulosic or non-cellulosic glucose content as determined biochemically (Fig. 8a and b). No correlation was observed using the sum of *m/z* 126 and *m/z* 114 (data not shown).

Xylose from glucurono(arabino)xylan comprises 67 to 82 mol% of the total non-cellulosic monosaccharide content, with the remainder, in descending order of abundance, in glucose, arabinose, and galactose. As with cellulosic and non-cellulosic glucose content with *m/z* 126, no correlation was observed between relative abundance of fragment *m/z* 114 and xylose or pentose content (Fig. 8c).

Discussion

PyMBMS is a valuable tool as a relatively high-throughput means to obtain a snapshot of the composition of lignocellulosic biomass. In their hallmark paper, Evans and Milne [9] provided several guiding principles for interpreting over a hundred *m/z* fragments derived from biomass and understanding their limitations as diagnostic for polysaccharide and lignin constituents. The ion profile is dependent on both the temperature of pyrolysis and the residence time [9]. We found that generation of all fragments to occur within the first 30 s of a 90-s acquisition time.

The utility of PyMBMS is increasing as our understanding of pyrolytic mechanisms is evolving to explain chemically why simple substrates, such as cellulose and xylan, give rise to an array of fragment ions [32]. Standard materials, such as

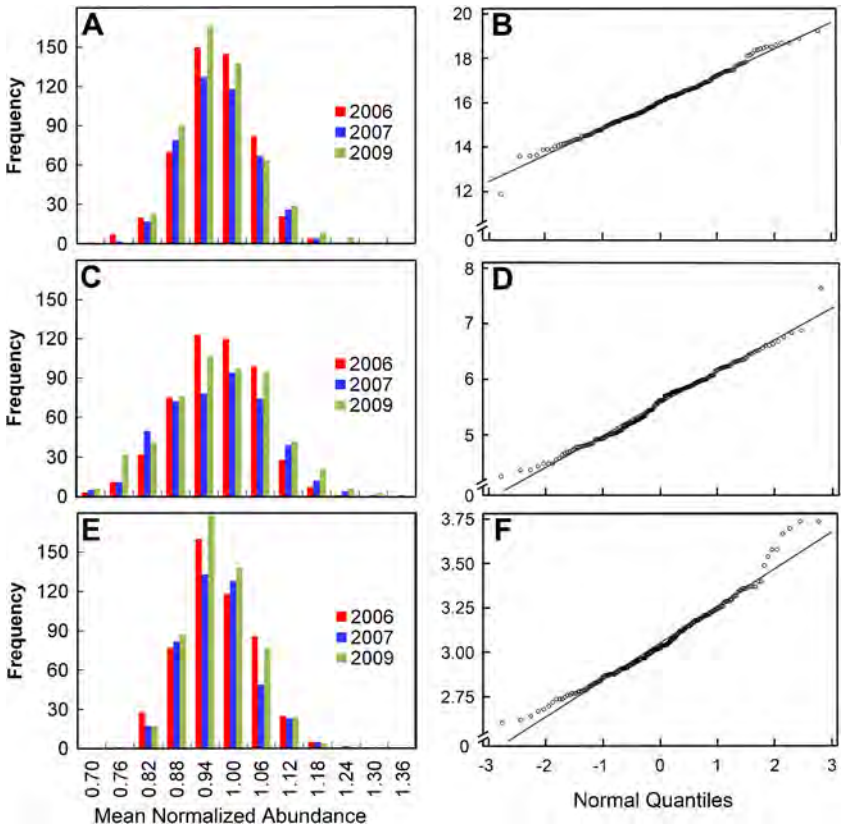

Fig. 4 Frequency distributions and Quantile–Quantile plots of the abundance of diagnostic fragments for lignin, *p*-coumaric acid, and hexose in the IBM population. Abundance of assigned fragments in the IBM population is shown for **a** total lignin (sum of G-lignin: *m/z* 124, *m/z* 137, *m/z* 138, *m/z* 151, and S-lignin: *m/z* 154, *m/z* 167, *m/z* 168, *m/z* 198), **b** *p*-coumaric acid (*m/z* 120; 4-vinyl phenol), and **c** hexose (*m/z* 126). Frequency distributions for normalized relative abundance are shown for all 3 years of sampling. Shown are QQ plots from the SAS Univariate function test of normality, including Shapiro–Wilk tests with the 2007 population, for **b** total lignin, **d** *p*-coumaric acid, and **e** hexose

cellulose, larch xylan, and lignin, give complex, sometimes similar spectra as a result of secondary and tertiary reactions that occur in the gas phase, and apparent abundance of the few diagnostic fragments of these spectra can be impacted by the physical state of the material. Pyrolysis of cellulose results in three putatively diagnostic fragments of, in decreasing order of abundance, *m/z* 144, *m/z* 162, and *m/z* 126, but a vast majority of the ions generated are a complex set of values below *m/z* 100. However, if the cellulose is pre-treated with mono- and dibasic potassium carbonate, the profile changes markedly to reveal large amounts of *m/z* 126, followed by *m/z* 144 and *m/z* 114. Larch wood xylan clearly gives a peak at *m/z* 114 but no *m/z* 126 or *m/z* 144, and about the same amounts of the various fragments below *m/z* 100 that are seen for cellulose. In contrast to polysaccharide fragmentation, sweet gum lignin yields a vast majority of fragments between *m/z* 137 and *m/z* 210, representing primarily G- and S-lignin derivatives.

Relative abundances of diagnostic fragments varied widely among biomass sources but were well correlated in samples with greatly differing composition. For example, the xylan- and S-lignin-rich birch wood gave abundant *m/z* 114 for xylose and *m/z* 167, *m/z* 180, and *m/z* 210 for syringyl derivatives, whereas mannan- and G-lignin-rich Douglas fir gave abundant *m/z* 126 for hexose and *m/z* 137 and *m/z* 180 from guaiacyl units [9]. Pyrolysis of grass bagasse gave a distinctive fragmentation pattern that gave high *m/z* 114 consistent

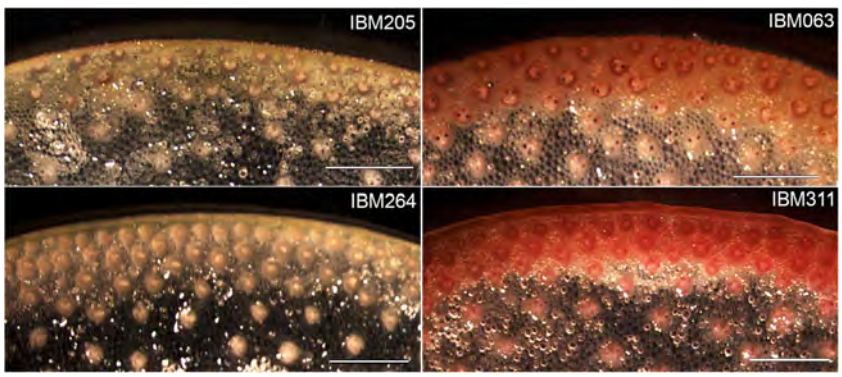

Fig. 5 Phloroglucinol comparative staining of low- and high-lignin with selected high- and low-lignin IBM lines. *Red color* is a result of staining coniferyl and sinapyl aldehydes and gives relative content of lignin. Low lignin lines are IBM205 and IBM264; high-lignin lines are IBM063 and IBM311. Samples are from internode 4 at the time of tassel appearance. Scale bars=1 mm

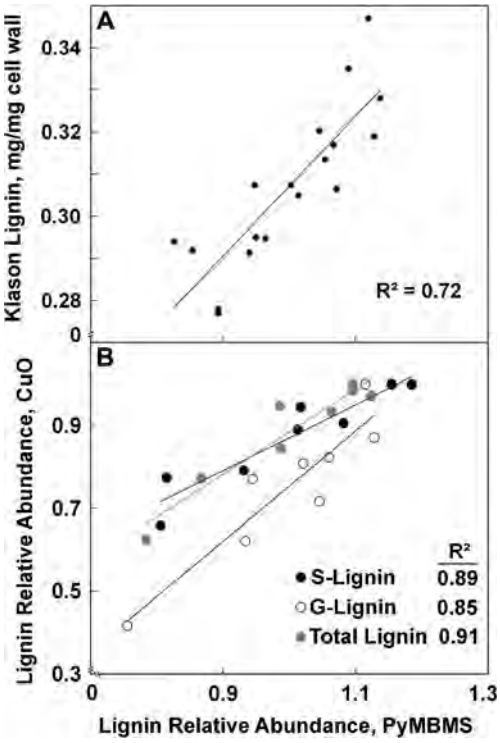

Fig. 6 Correlation of lignin determinations with total G- and S-lignin relative abundance of their diagnostic fragments derived from PyMBMS. Relative abundance of G-lignin is the sum of *m/z* 124, *m/z* 137, *m/z* 138, and *m/z* 151 and that of S-lignin is the sum of *m/z* 154, *m/z* 167, *m/z* 168, and *m/z* 198. **a** Comparison of PyMBMS values of the samples of the IBM population with their Klason lignin. **b** Comparison of PyMBMS values of the samples of the IBM population with their G- and S-lignin content from CuO oxidation analysis. R^2 values for the correlations of each diagnostic fragment with their respective lignin determinations each year are inset

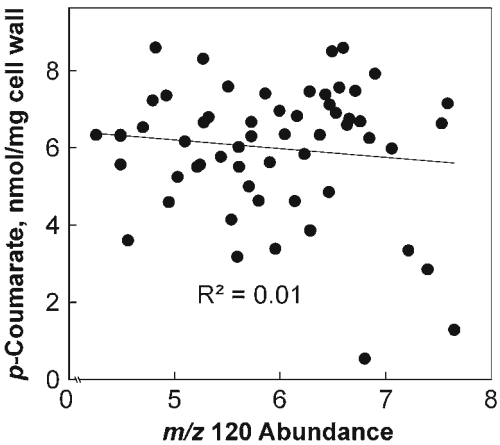

Fig. 7 Correlation of saponifiable *p*-coumaric acid content with relative abundance of its diagnostic fragment derived from PyMBMS. Relative abundance of *p*-coumaric acid by PyMBMS is based on fragment *m/z* 120 (4-vinyl phenol). Saponifiable *p*-coumaric acid was determined without derivatization by HPLC using 2,3,4-trimethyl-*trans*-cinnamic acid as a quantitative standard. R^2 values for the correlations of fragment *m/z* 120 abundance with *p*-coumaric acid determination each year are inset

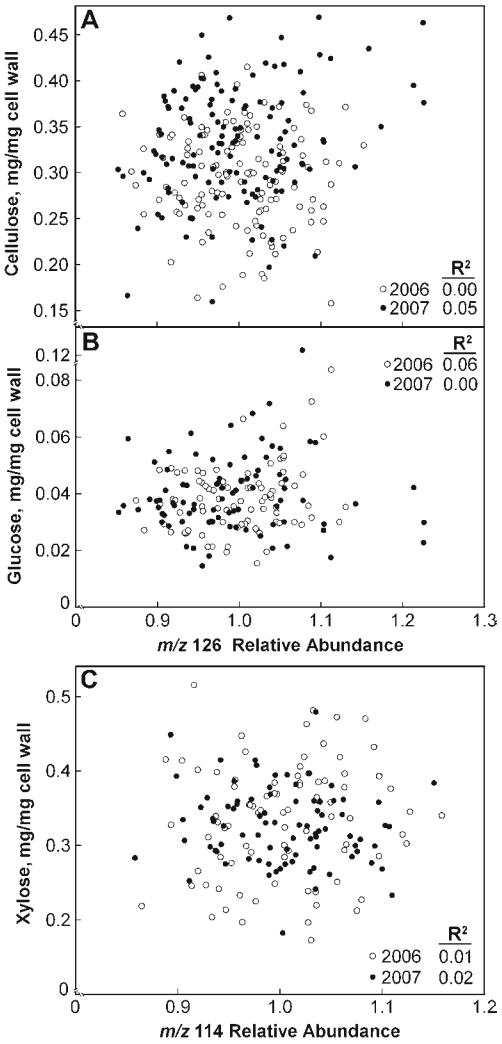

Fig. 8 Correlation of cellulosic and non-cellulosic polysaccharide content with the relative abundance of their diagnostic fragments from PyMBMS. **a** Cellulosic glucose content vs. relative abundance of *m/z* 126. **b** Non-cellulosic glucose content with relative abundance of *m/z* 126. **c** Xylose content with relative abundance of *m/z* 114. R^2 values for the correlations of each diagnostic fragment with their respective cellulosic or non-cellulosic monosaccharide determinations each year are inset

with expected xylan enrichment, and fragments of *m/z* 150 and *m/z* 120 indicated coumaryl derivatives, presumably from the hydroxycinnamic acid groups, *p*-coumaric and ferulic acid, in addition to guaiacyl and syringyl lignin derivatives.

Consistent with the observations of others using woody tissues, [11, 18] we find reasonable correlation of the abundance of fragments attributed to lignin with both Klason lignin determinations and CuO products of G-, S-, and even H-lignin (Figs. 5 and 6), and little to no correlation between polysaccharide fragment assignments from PyMBMS and measurements of cellulose, xylan, non-cellulosic glucose, or saponifiable hydroxycinnamic acids (Figs. 7 and 8). In contrast to these results, Agblevor et al. [33] found good correlation between carbohydrate constituents but only weak correlation

with lignin components. Different molecules can give rise to fragments of the same size and a lot of this variability depends upon conditions of pyrolysis and the architecture of the biomass, and we suggest that these ambiguities may account for inconsistency in the ability to correlate all fragment abundances with their respective constituent.

That said, our lack of correlation between the PyMBMS fragments and hydroxycinnamic acids is puzzling. Numerous reports show the enrichment in grasses of *p*-coumaric acid associated with an abundant *m*/*z* 164 fragment and secondary fragmentation to *m*/*z* 120 (4-vinyl phenol) [9, 34]. While H-lignin is present as a minor component of lignin and is responsible for the *m*/*z* 150 peak (4-coumaryl alcohol), much of the *m*/*z* 120 fragment should indeed be evolving from *p*-coumaric acid. It is well established that close to 100 % of the *p*-coumaric acid, if not all, is esterified and saponifiable [35–37]. Our lack of correlation might stem from architectural features or other factors, such as salt concentration that result in secondary and tertiary fragmentations that preclude a simple quantitative correlation.

PyMBMS fragment abundance and quantitation has allowed discrimination between control and Bt toxin transgenic *Populus* species [38], lignin variability in different portions of *Populus* stem [18] and lignin content in switchgrass grown under varied environmental conditions [13], where Principal Components Analyses and similar statistical packages can resolve populations of samples using multivariate analyses. Lignin abundance [11] or specific fragment ions used as sole traits [12] can also be used for QTL studies. In addition, strong QTL for *m*/*z* 120 (4-vinyl phenol) have also been found, despite the lack of correlation between this fragment and *p*-coumaric acid abundance [39].

Conclusions

PyMBMS spectra provide a relative measure, rather than an absolute value, for components of the pyrolyzed material, depending on the pyrolysis temperature, the ramp time to that temperature (residence time), and the physical state of the material pyrolyzed. However, a relative value is preferred for evaluation of lignocellulosic biomass in large populations. Despite this dependency of spectral profiles on conditions, different species still show markedly different profiles consistent with our biochemical knowledge of their wall composition. Our data indicate that lignin ions correlate well with chemical determinations of abundance, and that reproducible values for relative abundance of any assigned fragment can be used as a quantitative trait. Thus, PyMBMS is validated as a high-throughput screen to utilize the wealth of genetic resources of maize and other model grass species.

Acknowledgments We thank Steve Moose, University of Illinois, for providing access to the maize stover and helpful discussions. Multiyear sampling of the IBM population was supported by the U.S. Department of Energy Feedstock Genomics Program, Office of Biological and Environmental Research, Office of Science; screening of the IBM population by PyMBMS was supported by the BioEnergy Science Center a U.S. Department of Energy Bioenergy Research Center supported by the Office of Biological and Environmental Research in the DOE Office of Science Contract No. DE-AC36-08-GO28308, and by the National Science Foundation "Hy-Bi", an Emerging Frontiers in Research and Innovation (EFRI) program, Award No. 0938033; histochemical and analytical validations of PyMBMS for lignin and carbohydrate abundance were supported by the Center for Direct Catalytic Conversion of Biomass to Biofuels (C3Bio), an Energy Frontier Research Center funded by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, Award Number DE-SC0000997.

References

1. Dweikat I, Weil C, Moose S, Kochian L, Mosier N, Ileleji K et al (2012) Envisioning the transition to a next-generation biofuels industry in the US Midwest. BioFPR 6:376–386
2. Romay MC, Millard MJ, Glaubitz JC, Peiffer JA, Swarts KL, Casstevens TM et al (2013) Comprehensive genotyping of the USA national maize inbred seed bank. Genome Biol 14:R55
3. Liu K, Goodman M, Muse S, Smith JS, Buckler ES, Doebley J (2003) Genetic structure and diversity among maize inbred lines as inferred from DNA microsatellites. Genetics 165:2117–2128
4. Flint-Garcia S, Thullet A-C, Yu J, Pressoir G, Romero SM, Mitchell SE, Doebley J, Kresovich S, Goodman MM, Buckler ES (2005) Maize association population: a high-resolution platform for quantitative trait locus dissection. Plant J 44:1054–1064
5. Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES (2007) TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics 23:2633–2635
6. Lee M, Sharopova N, Beavis WD, Grant D, Katt M et al (2002) Expanding the genetic map of maize with the intermated B73×Mo17 (IBM) population. Plant Mol Biol 48:453–461
7. Sharopova N, McMullen M, Schultz L, Schroeder S, Sanchez-Villeda H, Gardiner J et al (2002) Development and mapping of SSR markers for maize. Plant Mol Biol 48:463–481
8. Yu J, Holland JB, McMullen MD, Buckler ES (2008) Genetic design and statistical power of nested association mapping in maize. Genetics 178:539–551
9. Evans RJ, Milne TA (1987) Molecular characterization of the pyrolysis of biomass. 1. Fundamentals. Energ Fuels 1:123–137
10. Boon JJ (1989) An introduction to pyrolysis mass spectrometry of lignocellulosic material: case studies of barley straw, corn stem and *Agropyron*. In: Chesson A, Ørskov ER (eds) Physicochemical characterization of plant residues for industrial feed use. Elsevier Applied Science, Amsterdam, pp 25–49
11. Tuskan G, West D, Bradshaw HD, Neale D, Sewell M, Wheeler N et al (1999) Two high-throughput techniques for determining wood properties as part of a molecular genetics analysis of hybrid poplar and loblolly pine. Appl Biochem Biotechnol 77–79:55–65
12. Yin T, Zhang X, Gunter L, Priya R, Sykes R, Davis M, et al. (2010) Differential detection of genetic loci underlying stem and root lignin content in Populus. PLoS One 5: Art. No. e14021.
13. Mann DGJ, Labbé N, Sykes RW, Gracom K, Kline L, Swamidoss IM et al (2009) Rapid assessment of lignin content and structure in switchgrass (*Panicum virgatum* L.) grown under different environmental conditions. Bioenerg Res 2:246–256

14. Carpita NC (1996) Structure and biogenesis of the cell walls of grasses. Annu Rev Plant Physiol Plant Mol Biol 47:445–476
15. Carpita NC, Gibeaut DM (1993) Structural models of primary cell walls in flowering plants: consistency of molecular structure with the physical properties of the walls during growth. Plant J 3:1–30
16. Barrière Y, Riboulet C, Méchin V, Maltese S, Pichon M, Cardinal A et al (2007) Genetics and genomics of lignification in grass cell walls based on maize as model species. Genes, genomics, and genetics. Global Science Books, In, pp 133–156
17. Vanholme R, Demedts B, Morreel K, Ralph J, Boerjan W (2010) Lignin biosynthesis and structure. Plant Physiol 153:895–905
18. Sykes R, Kodrzycki B, Tuskan G, Foutz K, Davis M (2008) Within tree variability of lignin composition in *Populus*. Wood Sci Technol 42:649–661
19. Edwards AL (1976) The correlation coefficient. An introduction to linear regression and correlation. WH Freeman, San Francisco, In, pp 33–46
20. Mecklin C (2007) Shapiro–Wilk test for normality. In: Salkind N (ed) Encyclopedia of measurement and statistics. SAGE Publications, Thousand Oaks, CA, pp 884–887
21. Sindhu A, Langewisch T, Olek A, Multani DS, McCann MC, Vermerris W et al (2007) Maize *Brittle stalk2* encodes a COBRA-like protein expressed in early organ development but required for tissue flexibility at maturity. Plant Physiol 145:1444–1459
22. Theander O, Westerlund EA (1986) Studies on dietary fiber. 3. Improved procedures for analysis of dietary fiber. J Agric Food Chem 34:330–336
23. Hatfield RD, Jung HJG, Ralph J, Buxton DR, Weimer PJ (1994) A comparison of the insoluble residues produced by the Klason lignin and acid detergent lignin procedures. J Sci Food Agric 65:51–58
24. Hemm MR, Ruegger MO, Chapple C (2003) The Arabidopsis ref2 mutant is defective in the gene encoding CYP83A1 and shows both phenylpropanoid and glucosinolate phenotypes. Plant Cell 15:179–194
25. Hedges JI, Mann DC (1979) Characterization of plant tissues by their lignin oxidation products. Geochim Cosmochim Acta 43:1803–1807
26. Hedges JI, Ertel JR (1982) Characterization of lignin by gas capillary chromatography of cupric oxide oxidation-products. Anal Chem 54: 174–178
27. Filley TR, Boutton TW, Liao JD, Jastrow D (2008) Chemical changes to non-aggregated particulate soil organic matter following grassland-to-woodland transition in a subtropical savanna. J Geophys Res 113, G03009
28. Dubois M, Gilles KA, Hamilton JK, Rebers PA, Smith F (1956) Colorimetric method for determination of sugars and related substances. Anal Chem 28:350–356
29. Updegraff D (1969) Semimicro determination of cellulose in biological materials. Anal Biochem 32:420–424
30. Gibeaut DM, Carpita NC (1991) Tracing cell-wall biogenesis in intact cells and plants. Selective turnover and alteration of soluble and cell-wall polysaccharides in grasses. Plant Physiol 97:551–561
31. Pomar F, Merino F, Barceló AR (2002) *O*-4-Linked coniferyl and sinapyl aldehydes in lignifying cell walls are the main targets of the Wiesner (phloroglucinol-HCl) reaction. Protoplasma 220:17–28
32. Venkatakrishnan VK, Degenstein JC, Smeltz AD, Delgass WN, Agrawal R, Ribeiro FH (2013) High pressure fast-pyrolysis, fast-hydropyrolysis and catalytic hydrodeoxygenation of cellulose: Production of liquid fuel from biomass. Green Chemistry, in press.
33. Agblevor FR, Evans RJ, Johnson KD (1994) Molecular-beam mass-spectrometric analysis of lignocellulosic materials. 1. Herbaceous biomass. J Anal Appl Pyrolysis 30:125–144
34. Ralph J, Hatfield RD (1991) Pyrolysis-GC–MS characterization of forage materials. J Agric Food Chem 39:1426–1437
35. Mueller-Harvey I, Hartley RD, Harris PJ, Curzon EH (1986) Linkage of *p*-coumaroyl and feruloyl groups to cell-wall polysaccharides of barley straw. Carbohydr Res 148:71–85
36. Ralph J, Hatfield RD, Quideau S, Helm RF, Grabber JH, Jung HJG (1994) Pathway of *p*-coumaric acid incorporation into maize lignin as revealed by NMR. J Am Chem Soc 116:9448–9456
37. Grabber JH, Hatfield RD, Ralph J, Zon J, Amrhein N (1995) Ferulate cross-linking in cell walls isolated from maize cell suspensions. Phytochemistry 40:1077–1082
38. Davis MF, Tuskan GA, Payne P, Tschaplinski TJ, Meilan R (2006) Assessment of *Populus* wood chemistry following the introduction of a Bt toxin gene. Tree Physiol 26:557–564
39. Penning BW, Sykes WR, Babcock NC, Dugard CK, Held MA, Klimek JF, et al. (2013) Quantitative trait loci contributing to phenylpropanoid abundance are different from those for saccharification yield in a maize recombinant inbred population. In review.