

# Determination of Lignin Monomer Contents in Rice Straw Using Visible and Near-infrared Reflectance Spectroscopy

Zhen Hu,<sup>a,b,c</sup> Guifen Zhang,<sup>a,b,c</sup> Yuanyuan Chen,<sup>a,b,c</sup> Youmei Wang,<sup>a,b,c</sup> Yuqing He,<sup>b,d</sup> Liangcai Peng,<sup>a,b,c</sup> and Lingqiang Wang<sup>a,b,c,\*</sup>

Genetic modification of plant lignin composition is an important strategy for improving biomass enzymatic digestibility without sacrificing the normal growth of the plant. However, the absence of fast and convenient methods for rapid determination of lignin composition has impeded corresponding research. Near-infrared reflectance spectroscopy (NIRS) analysis has potential as a solution for this dilemma, while the NIRS measurement for expediently assaying lignin composition in rice straws is still lacking. In this study, visible and near-infrared reflectance spectroscopy (VIS/NIRS) and modified partial least squares (MPLS) method were combined to develop calibration models for predicting the lignin monomer contents in a diverse rice population. Four optimal equations for predicting the content of *p*-hydroxyphenyl (H), guaiacyl (G), and syringyl (S) lignin units and their total amount (H + S + G) were generated with acceptable determination coefficients for calibration (0.85 to 0.93), cross-validation (0.75 to 0.88), and external validation (0.82 to 0.88), and the ratio performance deviation (*RPD*, 2 to 3.01). This study was the first to demonstrate that VIS/NIRS could give a sufficiently accurate prediction of lignin monomer contents in rice and could be applied for rapid assessments of large-scale rice straw samples.

*Keywords:* Visible and near-infrared reflectance spectroscopy (VIS/NIRS); Calibration models; Lignin monomer; Rice

*Contact information:* a: Biomass and Bioenergy Research Centre, Huazhong Agricultural University, Wuhan, China; b: National Key Laboratory of Crop Genetic Improvement, Huazhong Agricultural University, Wuhan, China; c: College of Plant Science and Technology, Huazhong Agricultural University, Wuhan, China; d: College of Life Science and Technology, Huazhong Agricultural University, Wuhan, China; \*Corresponding author: lqwang@mail.hzau.edu.cn

## INTRODUCTION

Lignin is a crucial material in the evolution of plants from an aquatic environment to a land environment and provides land plants water transport, mechanical support, stress response, and pathogen resistance (Boerjan *et al.* 2003). Unfortunately, it has been identified as a major contributor to biomass recalcitrance, which is the blockage that leads to the high cost of lignocellulosic sugar production (Ding *et al.* 2012; Zeng *et al.* 2014). This economic effect has promoted lignin modification to become a popular research topic. Numerous studies have established that increased biomass digestibility could be gained by reduced lignin content and altered lignin composition (Fu *et al.* 2011; Bonawitz *et al.* 2014). However, a reduction in lignin often disturbs plant development and causes undesirable effects (Chen and Dixon 2007; Bonawitz and Chapple 2013).

A feasible method for lignin engineering to decrease biomass recalcitrance involves changing the proportion of three major types of lignin units, *p*-hydroxyphenyl (H), guaiacyl (G), and syringyl (S) units (Wang *et al.* 2016).

The lignin monomer content is typically a quantitative trait, while the studies that attempted to identify monolignol synthesis genes through quantitative trait locus/loci (QTL) mapping or genome-wide association study (GWAS) have been very limited (Barrière *et al.* 2008). A main reason that limits the use of classic quantitatively genetic methods in identification of the monolignol synthesis gene might be the complex process for assaying lignin monomer content on a large scale.

Traditional wet chemical methods used to analyze lignin composition include thioacidolysis (TA), nitrobenzene oxidation (NBO), and derivatization followed by reductive cleavage (DFRC) (Wen *et al.* 2013). Unfortunately, all of these methods are labor-intensive and time-consuming. Even though high-throughput systems with small vials have been developed to improve the efficiency of the process (Foster *et al.* 2010; Penning *et al.* 2014), additional costs of the equipment might limit their applications to some extent. Nonetheless, reliable, low-cost, and time-effective methods for lignin composition assessments are urgently needed.

The NIRS analysis has been shown to be a fast, environmentally friendly analytical method and has gained a wide variety of applications in the prediction of biomass property and component composition (Xu *et al.* 2013). The absorption of the near-infrared spectroscopy radiation is primarily induced by overtone and combination bands of fundamental stretching vibrations of O-H, C-H, and N-H, which represent the major chemical bonds in biological compounds (Bailleres *et al.* 2002). Through mathematical approaches, near-infrared spectral data, and reference parameters are combined to build the calibration models (Payne and Wolfrum 2015). Calibration models for accurate prediction of the biomass compositions of rice (Jin and Chen 2007), wheat (Owens *et al.* 2009), sorghum (Wu *et al.* 2015; Xu *et al.* 2015), *Miscanthus* (Jin *et al.* 2017), bamboo (Yang *et al.* 2016), sugarcane (Hoang *et al.* 2017), and wood (Bailleres *et al.* 2002), have been built using NIRS. However, the progress in quantitative prediction of lignin composition in plants is still very limited, and all of these studies were conducted on just wood (Bailleres *et al.* 2002; Alves *et al.* 2006; Yamada *et al.* 2006; Maranan and Laborie 2008; Robinson and Mansfield 2009).

Rice is an important food crop for human consumption. Its annual production generates about 800 million metric tons of straw, which provides an abundant biomass source for the production of renewable energy (Domínguez-Escribá and Porcar 2010). A reliable and fast method for assaying lignin composition of rice straw is necessary to efficiently design and screen rice varieties that are more amenable to enzymatic digestibility. Up until now, no research has been conducted to build NIRS calibration equations for the prediction of lignin composition in rice materials. The target of this work was to establish a rapid and accurate VIS/NIRS measurement for expediently assaying lignin composition in rice straw.

## EXPERIMENTAL

### Materials

The 100 rice straw samples were selected from a rice recombinant inbred line (RIL) population described by Yan *et al.* (2014). The rice lines were planted in the

experimental field of Huazhong Agricultural University (Wuhan, China) during the natural growing season in 2012. The straws of three individual plants for each line were collected after grain harvest. Then the leaves were removed, and the stems were inactivated at 105 °C for 30 min and dried at 60 °C to a constant weight. The dried stems were ground to pass through a 40-mesh screen and stored in a dry container until their use.

## Methods

### *Cell wall extraction and lignin composition assay*

Cell wall isolation and lignin composition assay were conducted using a high-throughput platform previously described by Foster *et al.* (2010). The isolated cell wall material was prepared by washing the samples with 1.5 mL of 70% aqueous ethanol and 1.5 mL of chloroform/methanol (1:1, v/v), followed by treatment with 35 µL of 0.01% sodium azide (NaN<sub>3</sub>), 35 µL amylase (50 µg/mL in H<sub>2</sub>O, from *Bacillus* species, Sigma), and 17 µL pullulanase (17.8 units, from *Bacillus acidopullulyticus*, Sigma). For lignin composition assay, 2 mg of isolated cell wall material was transferred into a screw-capped glass tube. Dioxane (175 µL), ethanethiol (EtSH, 20 µL 10%), and boron trifluoride diethyl etherate (BF<sub>3</sub>, 5 µL 2.5%) were added for thioacidolysis. The reaction was heated at 100 °C for 4 h with gentle mixing every hour. For purification, 1 mL of water and 0.5 mL of ethyl acetate was added, then mixed thoroughly with a vortex mixer, and the phases were left to separate. Next, 150 µL of the ethyl acetate layer (top layer) was transferred to a 2 mL Sarstedt tube.

The solvent was evaporated by a concentrator with air. For the trimethylsilyl (TMS) derivatization, 500 µL of ethyl acetate, 20 µL of pyridine, and 100 µL of N, O-bis(trimethylsilyl) acetamide were added together, and the mixture was incubated for 2 h at 25 °C. The reaction products (100 µL) were analyzed by gas chromatography (GC) with a quadrupole mass-spectrometer (Santa Clara, CA, USA) or flame ionization detector (Restek Corporation, Bellefonte, PA, USA). Peaks were identified by characteristic mass spectrum ions of 299 m/z, 269 m/z, and 239 m/z for S, G, and H monomers, respectively. The composition of the lignin components was quantified by setting the total peak area to 100%. The chemical agents were purchased from Sigma-Aldrich, St. Louis, MO, USA.

### *Near-infrared reflectance spectra data collection and analysis*

Spectra collection and variability analysis of NIRS were conducted as described by Huang *et al.* (2012). Near-infrared reflectance spectra data was collected using a XDS Rapid Content™ Analyzer (FOSS, Co., LLC., Hillerod, Denmark) equipped with a dual detector system: silicon (400 nm to 1100 nm), lead sulfide (1100 nm to 2500 nm), and the ISIScan™ software (Infrasoft International LLC., Port Mathilda, PA, USA). The dried samples were placed into a mini-sample cup (standard ring cup) for screening. The reflectance (*R*) of each sample was recorded in triplicate at wavelengths ranging from 400 nm to 2500 nm with 2-nm intervals at room temperature.

The spectral absorbance values were recorded as  $\log 1/R$ . The WinISI III software package (Version 1.50e, Infrasoft International LLC., Port Matilda, PA, USA) was used for the chemometric management of data. A principal component analysis (PCA) algorithm was carried out to identify the spectral outlier sample and determine the structure and variability of spectral population (Cowe and McNicol 1985). The global H (GH) value of each sample was determined using the measured Mahalanobis distance from mean. The GH outlier (GH > 3.0) was finally eliminated after the PCA. Full spectra

wavelengths (408 nm to 2492 nm) were selected and pretreated with “1, 4, 4, 1” treatments for principal component analysis, the four digits orderly represented the number of the derivative, the gap over which the derivative is calculated, the number of the first smoothing, and the number of the second smoothing.

#### *Calibration and validation*

The prediction models were calibrated and validated following the procedure of Huang *et al.* (2012) with minor modifications. Two mathematical treatments as “0, 0, 1, 1” and “1, 4, 4, 1”, and seven scatter correction methods were used to build prediction equations.

The seven methods were: no scatter correction standard (None), standard normal variant (SNV), detrend only (DET), combination of SNV and detrend (SNVD), standard multiple scatter correction (SMSC), weighted multiple scatter correction (WMSC), and inverse multi scatter correction (IMSC). In addition, three wavelength ranges (408 nm to 2492 nm, 780 nm to 2492 nm, and 1108 nm to 2492 nm), and the modified partial least squares (MPLS) method were also utilized.

Cross-validation was conducted during the development of equations to select the optimal number of factors and to avoid over-fitting. In addition, one of every four samples sorted based on the laboratory value were selected for the external validation. The coefficient of determination of the cross-validation ( $R^2_{cv}$ ) was defined as the reference index to select the best equations.

#### *QTL mapping using measured data and predicted data*

The genetic map of the recombinant inbred line (RIL) population was built using 179 markers and MAPMAKER/EXP version 3.0b software (Whitehead Institute for Biomedical Research, Cambridge, MA, USA). The following inclusive composite interval mapping was carried out using QTL IciMapping version 3.3 software (Institute of Crop Science, Chinese Academy of Agricultural Sciences, Beijing, China) based on stepwise regression with simultaneous consideration of all marker information. The walking speed chosen for all QTL mapping was 1.0 cM.

#### *Statistical analysis*

A statistical analysis for calculating the minimum, maximum, mean, and standard deviation and the Kolmogorov-Smirnov test for checking the normal distribution were conducted using the software of the SPSS 17.0 (SPSS Inc., Chicago, IL, USA). The best-fit curve in frequency distributions and descriptors of regression analysis were developed using Origin 8.0 software (Microcal Software, Northampton, MA).

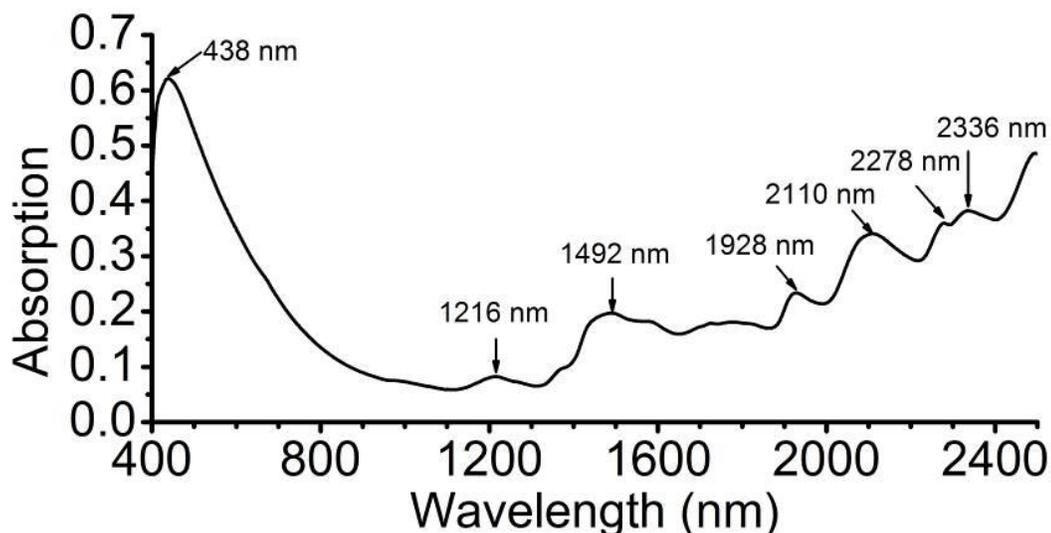
## **RESULTS AND DISCUSSION**

### **Characterization of Spectra and Identification of Outlier**

The averaged VIS/NIR spectrum of 100 rice samples with three replications was generated by the WinISI III software and is shown in Fig. 1. There was one absorption band peak that occurred in the visible region (400 nm to 770 nm) and six peaks that occurred in the near infrared region (780 nm to 2500 nm).

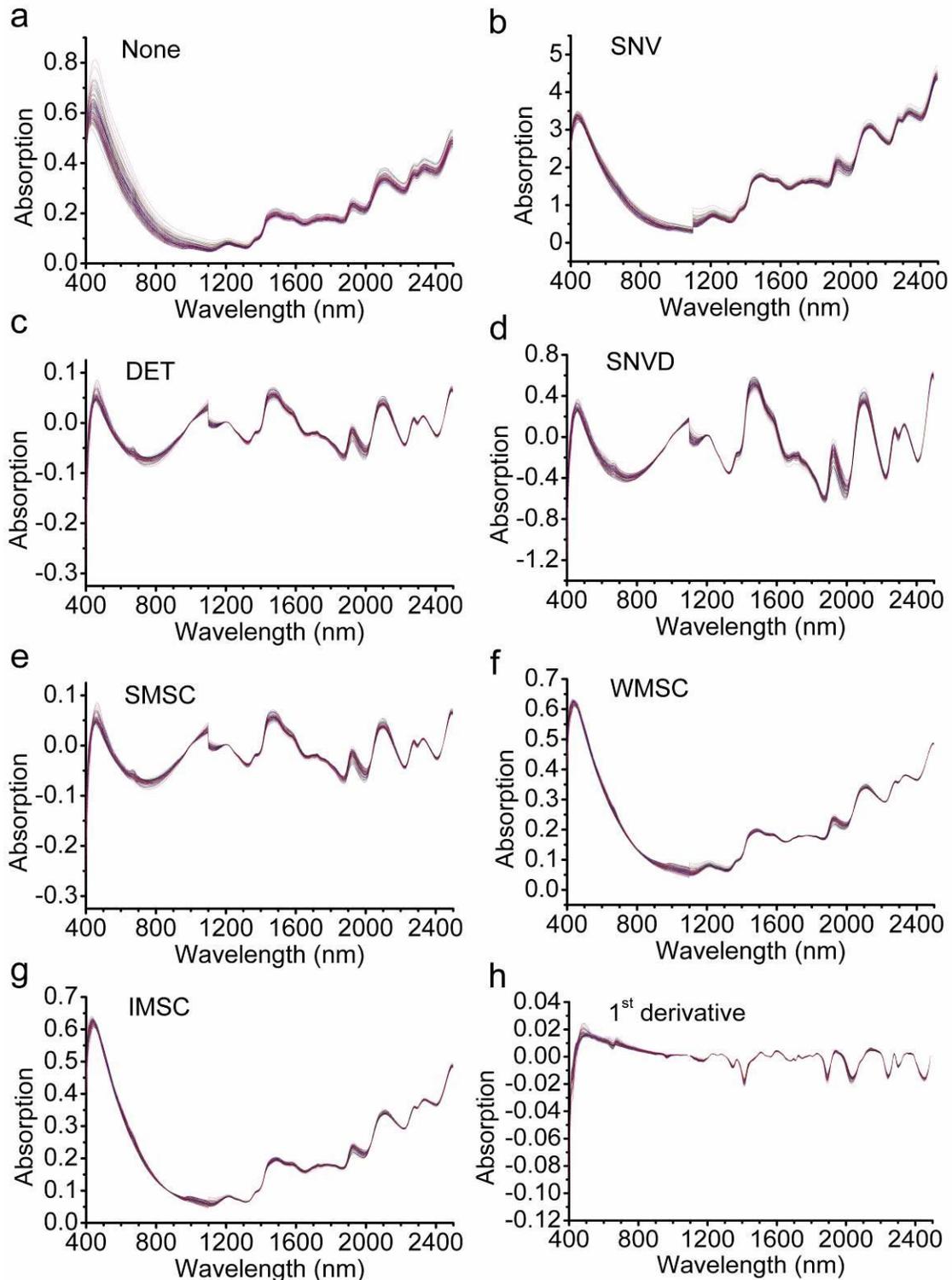
The strong peak in the visible region at 438 nm might have been induced by the yellow-orange color of the rice powders, as the spectral region from 425 nm to 492 nm represents the complementary color of yellow-orange (Shenk and Westerhaus 1994). The peak at 1216 nm and 1492 nm is primarily attributed to the 2<sup>nd</sup> and 1<sup>st</sup> overtone O–H stretching in cellulose, respectively (Schwanninger *et al.* 2011). The peak at 1928 nm is mainly attributed to the asymmetric stretching and deformation of O–H in water (Ali *et al.* 2001). The broad peak at 2110 nm is associated with the stretching and deformation of O–H in cellulose (Schwanninger *et al.* 2011). The peak at 2278 nm could be assigned to the O–H and C–C stretching and/or C–H stretching and deformation in cellulose (Ali *et al.* 2001). The last peak at 2336 nm is mainly attributed to the C–H stretching and C–H<sub>2</sub> deformation in lignin (Schwanninger *et al.* 2011). The original spectra and the pretreated spectra dealt with six scatter methods and the first-order derivative of all samples were also exhibited (Fig. 2).

The peaks in the pretreated spectra were clearer and steeper than those in original spectra and this might have been because the math pretreatments decreased the interferences induced by particle size, scatter coefficient, and path length variation.

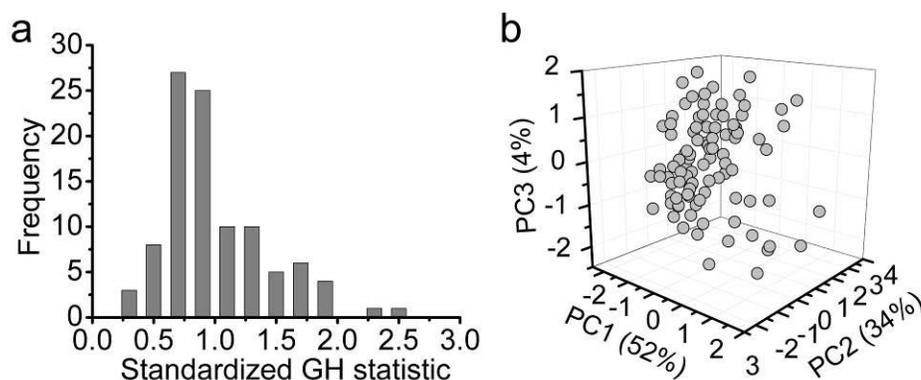


**Fig. 1.** Averaged visible-near infrared spectrum of original data of 100 rice samples. The main absorption band peaks are pointed out by arrows

In terms of the distribution and variability of the spectral population and identification of anomalous samples, linearly uncorrelated components were extracted to characterize the 100 averaged spectra. Consequently, the total of 16 components and the first three components explained 99.45% and 90% of the variance in the spectral data, respectively. Based on the calculated Global “H” values (GH), none of the samples was identified as an outlier (i.e.  $\text{GH} \geq 3$ ), while most of the GH values ranged between 0.25 and 2.0 (Fig. 3a). The distribution of PCA data was acceptably symmetrical in the 3D plot (Fig. 3b), which indicated that the spectra were appropriate for VIS/NIR spectra analysis.



**Fig. 2.** The original spectra and pretreated spectra of rice samples: (a) original spectra; (b) standard normal variant (SNV) pretreated spectra; (c) detrend (DET) pretreated spectra; (d) combination of SNV and detrend (SNVD) pretreated spectra; (e) standard multiple scatter correction (SMSC) pretreated spectra; (f) weighted multiple scatter correction (WMSC) pretreated spectra; (g) inverse multiple scatter correction (IMSC) pretreated spectra; and (h) 1<sup>st</sup> derivative pretreated spectra



**Fig. 3.** Histogram of global H (GH) value of 100 averaged spectra (a) and the three-dimensional plot of the PCA scores of spectra in principal components space (b)

### Reference Data and Determination of Calibration/ External Validation Sets

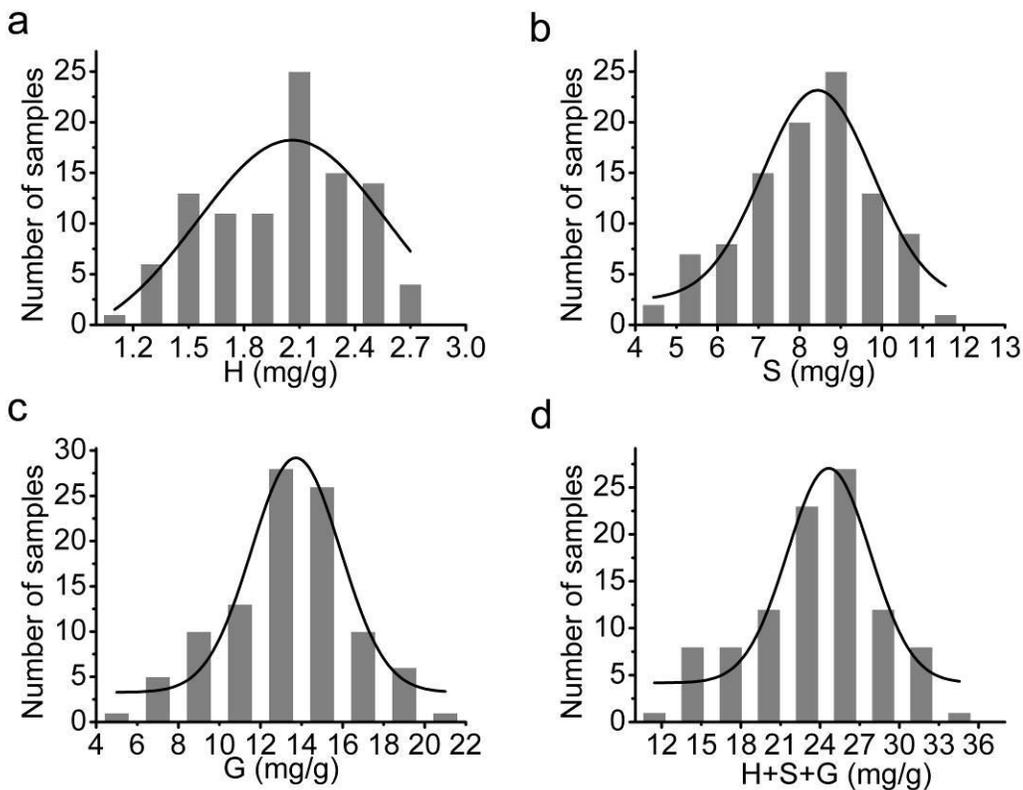
Minimum and maximum contents of the three major lignin monomers (H, S, and G) measured/observed in the samples are presented in Table 1. The results showed that G represented the most abundant lignin unit in rice straw, followed by S. The total lignin monomers represented 2.4% of the cell wall material, which was consistent to previous studies that the thioacidolysis method released approximately 20% by weight of the lignin and the lignin levels in rice were approximately 11% to 19% of the dry matter basis (Jourdes *et al.* 2007; Li *et al.* 2015). Both the lignin monomers and the total monomer displayed high levels of diversity with the coefficient of variation (CV) were about 20% and relatively normal distribution (Table 1, Fig. 4). The substantial variation of these traits showed that the rice straw samples were suitable for NIRS assay.

**Table 1.** Statistics of the Lignin Monomer Contents of Rice Samples

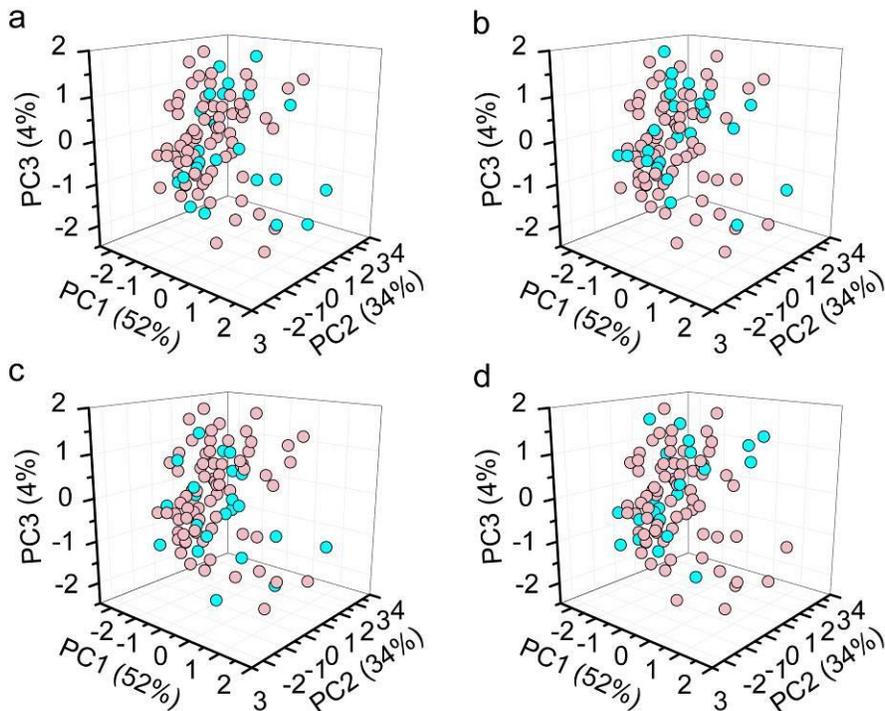
Traits	Number	Minimum	Maximum	Mean $\pm$ SD <sup>a</sup>	CV <sup>b</sup>	P <sup>c</sup>
H (mg/g)	100	1.08	2.68	2.02 $\pm$ 0.38	18.81	0.11
S (mg/g)	100	4.46	11.47	8.21 $\pm$ 1.57	19.12	0.20
G (mg/g)	100	5.73	20.15	13.3 $\pm$ 3.13	23.53	0.07
H + S + G (mg/g)	100	11.27	33.85	23.53 $\pm$ 4.84	20.57	0.06

Note: <sup>a</sup> Average  $\pm$  standard deviation; <sup>b</sup> Coefficient of variation (SD\*100/mean); <sup>c</sup> P value of Kolmogorov-Smirnov; P > 0.05 means the data are normally distributed

Based on the order of reference value, one of every four samples was selected and merged into the validation sets, and the remaining samples were included in the calibration sets. Thus, the external validation sets included 25 samples, and the calibration sets contained 75 samples. The trait values of the samples in the calibration and external validation sets displayed comparable range and similar standard deviation (Table 2). Moreover, samples in the external validation set were distributed evenly in the corresponding calibration samples as displayed in the three-dimensional plot (Fig. 5). Thus, the calibration and external validation sets were suitable for building and verifying equations.



**Fig. 4.** Frequency distributions of lignin monomer content in rice samples; frequency distribution was fitted by a Gaussian curve



**Fig. 5.** Distribution of the calibration set and external validation set for H (a), S (b), G (c), and H + S + G (d) in principal components space; the purple and blue circle represents the calibration and external validation set, respectively

**Table 2.** Calibration and External Validation Sets for Lignin Monomer Contents

Traits	Calibration					External Validation				
	<i>N</i> <sup>a</sup>	Min <sup>b</sup>	Max <sup>c</sup>	Mean	<i>SD</i> <sup>d</sup>	<i>N</i>	Min	Max	Mean	<i>SD</i>
H (mg/g)	75	1.08	2.68	2.02	0.38	25	1.23	2.64	2.01	0.39
S (mg/g)	75	4.46	11.47	8.22	1.58	25	4.77	11.01	8.18	1.59
G (mg/g)	75	5.73	20.15	13.32	3.13	25	6.34	19.49	13.24	3.18
H + S + G (mg/g)	75	11.27	33.85	23.55	4.88	25	12.97	32.10	23.46	4.82

<sup>a</sup> *N*- sample number; <sup>b</sup> Min- minimum value; <sup>c</sup> Max- maximum value; <sup>d</sup> *SD*- standard deviation

**Table 3.** Coefficient of Determination for Cross-validation ( $R^2_{cv}$ ) of Models Developed Using Different Spectrum Ranges and Mathematical Treatments for Lignin Monomers in Rice Materials

Spectrum Range (nm)	Derivative Treatment <sup>a</sup>	Scatter Methods <sup>b</sup>	Determination Coefficient for Cross-validation			
			H	S	G	H + S + G
408 to 2492	0,0,1,1	None	0.459	0.765	0.731	0.775
	0,0,1,1	SNV	0.477	0.739	0.743	0.797
	0,0,1,1	DET	0.466	0.773	0.731	0.827
	0,0,1,1	SNVD	0.457	0.763	0.787	0.796
	0,0,1,1	SMSC	0.471	0.766	0.764	0.789
	0,0,1,1	WMSC	0.514	0.776	0.814	0.788
	0,0,1,1	IMSC	0.453	0.756	0.758	0.798
	1,4,4,1	None	0.564	0.755	0.814	0.880
	1,4,4,1	SNV	0.687	0.781	0.767	0.833
	1,4,4,1	DET	0.647	0.745	0.842	0.879
	1,4,4,1	SNVD	0.714	0.768	0.828	0.825
	1,4,4,1	SMSC	0.687	0.781	0.768	0.834
	1,4,4,1	WMSC	0.734	0.781	0.827	0.826
	1,4,4,1	IMSC	0.687	0.781	0.767	0.832
780 to 2492	0,0,1,1	None	0.489	0.752	0.833	0.807
	0,0,1,1	SNV	0.397	0.739	0.810	0.803
	0,0,1,1	DET	0.459	0.764	0.852	0.812
	0,0,1,1	SNVD	0.512	0.764	0.764	0.823
	0,0,1,1	SMSC	0.483	0.810	0.816	0.807
	0,0,1,1	WMSC	0.600	0.766	0.831	0.811
	0,0,1,1	IMSC	0.517	0.800	0.813	0.815
	1,4,4,1	None	0.716	0.805	0.856	0.867
	1,4,4,1	SNV	0.669	0.790	0.821	0.842
	1,4,4,1	DET	0.655	0.790	0.858	0.872
	1,4,4,1	SNVD	0.712	0.767	0.845	0.841
	1,4,4,1	SMSC	0.669	0.789	0.821	0.843
	1,4,4,1	WMSC	0.711	0.789	0.828	0.828
	1,4,4,1	IMSC	0.669	0.790	0.821	0.842
1108 to 2492	0,0,1,1	None	0.487	0.783	0.827	0.823
	0,0,1,1	SNV	0.473	0.781	0.814	0.833
	0,0,1,1	DET	0.545	0.782	0.844	0.841
	0,0,1,1	SNVD	0.570	0.768	0.850	0.857
	0,0,1,1	SMSC	0.606	0.775	0.854	0.833
	0,0,1,1	WMSC	0.562	0.794	0.854	0.812
	0,0,1,1	IMSC	0.566	0.808	0.816	0.843
	1,4,4,1	None	0.729	0.765	0.844	0.847
	1,4,4,1	SNV	0.745	0.778	0.845	0.847
	1,4,4,1	DET	0.677	0.781	0.873	0.873
	1,4,4,1	SNVD	0.718	0.800	0.845	0.850
	1,4,4,1	SMSC	0.745	0.778	0.844	0.847
	1,4,4,1	WMSC	0.727	0.782	0.865	0.845
	1,4,4,1	IMSC	0.745	0.778	0.845	0.846

<sup>a</sup> The four digits orderly represent the number of the derivative, the gap over which the derivative was calculated, the number of the first smoothing, and the number of the second smoothing

<sup>b</sup> None, no scatter correction standard; SNV, standard normal variant; DET, detrend only; SNVD, combination of SNV and detrend; SMSC, standard multiple scatter correction; WMSC, weighted multiple scatter correction; IMSC, inverse multiple scatter correction

### Calibration and Validation of Lignin Monomeric Models

For the correlation between the spectral data and the laboratory values, seven scatter correction methods, two derivative treatments, and three spectrum regions combining the MPLS regression technique were applied to produce a total of 42 calibration equations for each trait (Table 3). The best equations according to the  $R^2_{cv}$  for the four traits were selected (Table 4). Generally speaking, no single math treatment or spectrum regions gave the best prediction for all traits. For H, G, and H + S + G, the best equation came from the 1, 4, 4, 1 derivative treatment using 6 factors, while the best equation for S came from the 0, 0, 1, 1 derivative treatment using 8 factors. Unlike single lignin monomers, the total lignin monomer was predicted best using the whole wavelength range including the visible spectrum, which indicated that the color of the rice powder might contain the information of the total lignin monomer content. The best equations for all four traits had high  $R^2$  values that ranged from 0.85 for H and 0.93 for H + S + G in calibration. In cross-validation, the equations also displayed acceptable  $R^2_{cv}$  values from 0.75 to 0.88 and ratio performance deviation (RPD) values from 2 to 3.01 (Arana *et al.* 2005). The external validation for the equations was conducted using the “Monitor results” module of the WinISI III software. The  $R^2_{ev}$  value of equations for these four traits ranged from 0.82 to 0.88 (Table 4). In addition, the predicted and reference values of these four traits showed good correlations, as indicated in the plots (Fig. 6) and the corresponding table of regression descriptors (Table 5).

**Table 4.** Calibration and Validation for the Best Equations Generated for Prediction of Lignin Monomer Contents in Rice Samples

Traits	Calibration				Cross-validation					External Validation	
	SEC <sup>a</sup>	R <sup>2b</sup>	Spectrum (nm)	DT <sup>c</sup>	SCM <sup>d</sup>	Terms <sup>e</sup>	SECV <sup>f</sup>	R <sup>2cv</sup> <sup>g</sup>	RPD <sup>h</sup>	SEP <sup>i</sup>	R <sup>2ev</sup> <sup>j</sup>
H	0.14	0.85	1108-2498	1,4,4,1	SMSC <sup>k</sup>	6	0.19	0.75	2	0.24	0.83
S	0.56	0.88	780-2498	0,0,1,1	SMSC	8	0.69	0.81	2.29	0.68	0.82
G	0.9	0.92	1108-2498	1,4,4,1	DET <sup>l</sup>	6	1.14	0.87	2.75	1.2	0.86
H + S + G	1.28	0.93	400-2498	1,4,4,1	None <sup>m</sup>	6	1.62	0.88	3.01	1.91	0.88

<sup>a</sup> SEC, standard error of calibration; <sup>b</sup> R<sup>2</sup>, determination coefficient of calibration

<sup>c</sup> DT, derivative treatment; the four digits orderly represent the number of the derivative, the gap over which the derivative was calculated, the number of the first smoothing, and the number of the second smoothing

<sup>d</sup> SCM, scatter correction methods

<sup>e</sup> Terms, number of principal component used for calibration

<sup>f</sup> SECV, standard error of cross-validation

<sup>g</sup> R<sup>2cv</sup>, determination coefficient of cross-validation

<sup>h</sup> RPD, ratio performance deviation ( $SD/SECV$ )

<sup>i</sup> SEP, standard error of prediction in external validation

<sup>j</sup> R<sup>2ev</sup>, determination coefficient of external validation

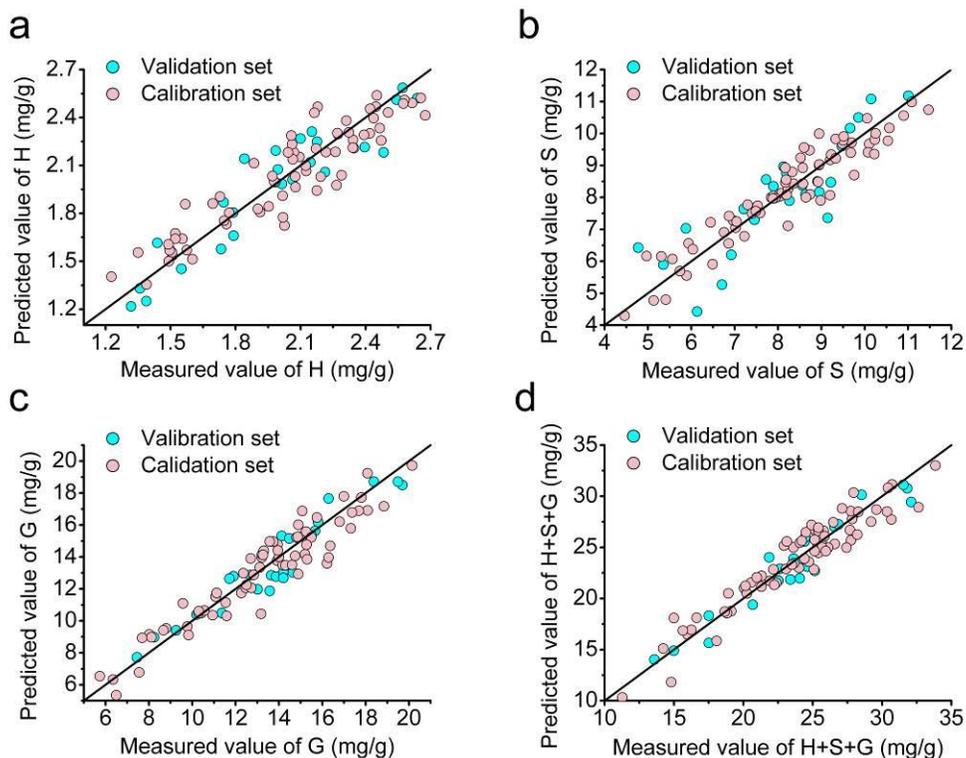
<sup>k</sup> SMSC, standard multiple scatter correction; <sup>l</sup> DET, detrend only

<sup>m</sup> None, no scatter correction

**Table 5.** Descriptors of Regression Line between Measured and Predicted Values of Calibration Set and External Validation Set

	Calibration Set				External Validation Set			
	H	S	G	H + S + G	H	S	G	H + S + G
Slope	0.84	0.89	0.91	0.91	0.94	0.52	0.93	0.96
Intercept	0.33	0.89	1.16	2.14	0.15	0.93	0.76	0.67
$R^2$ <sup>a</sup>	0.85	0.89	0.90	0.90	0.86	0.72	0.91	0.92

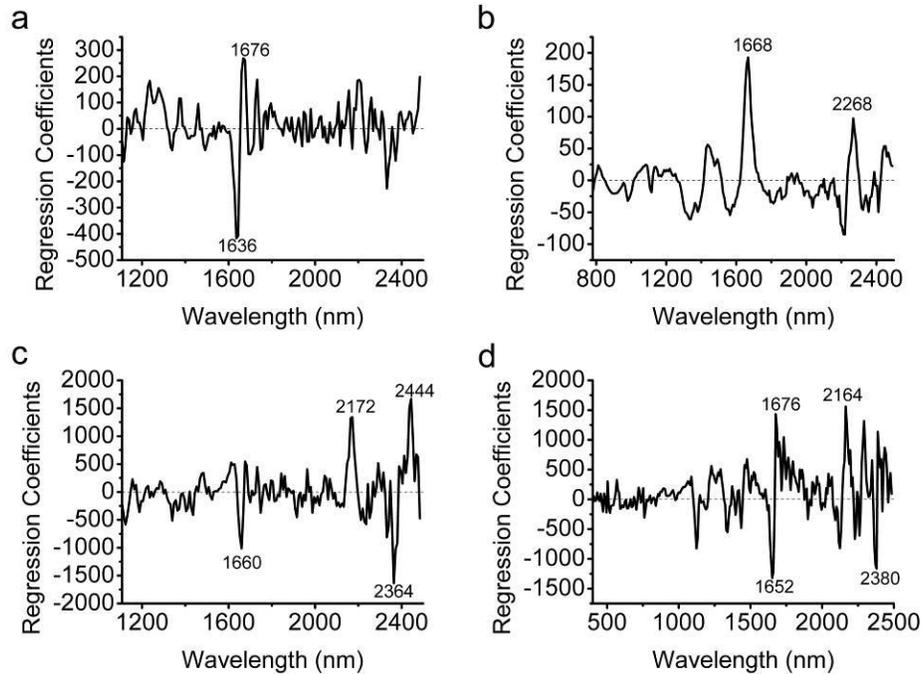
<sup>a</sup> Coefficient of determination, represents the degree of fit of the regression



**Fig. 6.** Correlation between the predicted and measured values of external validation set and calibration set for H (a), S (b), G (c), and H + S + G (d); the solid line indicates the best linear relationship (1:1)

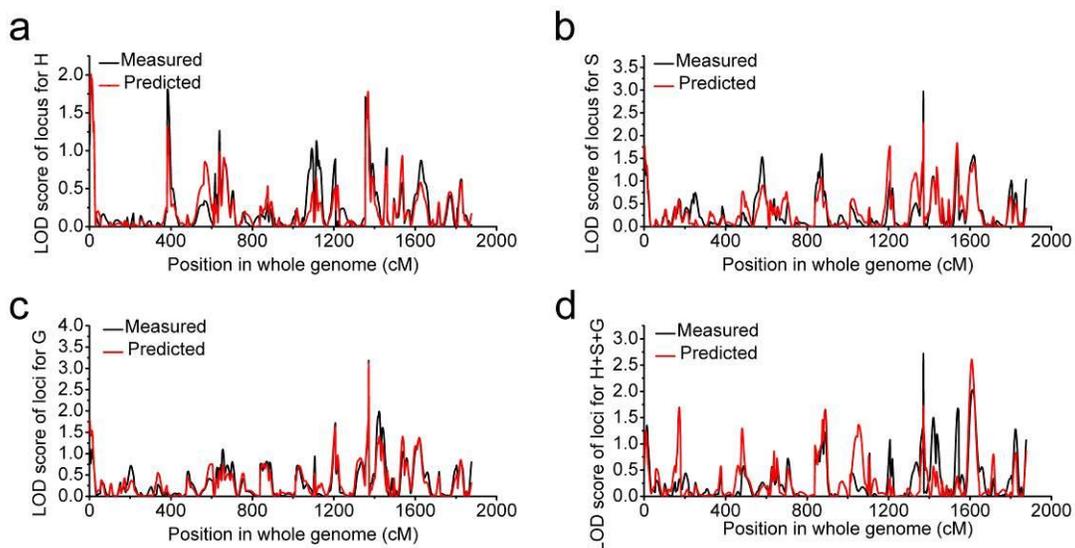
The regression coefficients that could be used to gain insight on the molecular features underlying the best calibration models are displayed in Fig. 7. The strong correlation coefficients were found at wavelengths centered at 1636 nm and 1676 nm for H, 1668 nm and 2268 nm for S, 1660 nm, 2172 nm, 2364 nm, and 2444 nm for G, and 1652 nm, 1676 nm, 2164 nm, and 2380 nm for H + S + G.

In identified wavelength, the regression correlations at 1660 nm and 1668 nm were consistent with the chemical structure of lignin, as the bands at these wavelengths correspond to the first overtone of aromatic C-H stretching in lignin (Barton *et al.* 1992; Michell and Schimleck 1996).



**Fig. 7.** Corresponding plots of regression coefficients for the prediction model for H (a), S (b), G (c), and H + S + G (d); the main peaks are delineated

To investigate whether the accuracy of these equations were satisfied for QTL analysis, QTL mapping was conducted using the measured data and predicted data of 87 rice samples whose genotype data had been investigated. The logarithm of odds (LOD) scores of genetic loci calculated by two sets of data for all traits were similar in full genetic map (Fig. 8), indicating the equations were sufficiently accurate to generate phenotype data for mapping genetic loci controlling lignin composition in rice straw.



**Fig. 8.** LOD scores of genetic locus for H (a), S (b), G (c), and H + S + G (d) calculated by predicted and measured values

## CONCLUSIONS

1. Visible and near-infrared spectroscopy (VIS/NIRS) and lignin monomer content data of rice materials were correlated by the modified partial least squares method for the calibration of a VIS/NIR model.
2. Four calibration equations for rapid predicting of the content of H, S, G, and H + S + G were built.
3. The equations showed an acceptable determination coefficient for calibration, cross-validation, and external validation that ranged from 0.85 to 0.93, 0.75 to 0.88, and 0.82 to 0.88, respectively.
4. These equations make it feasible to determine the lignin composition of a large number of rice samples in a time-effective and low-cost manner.

## ACKNOWLEDGMENTS

This work was supported by grants from the National Natural Science Foundation of China (Nos. 30900890, 31171524, and 31371550), the Fundamental Research Funds for the Central Universities of China (Nos. 2011PY047, 2662015PY207, and 2662015PY009), and the China Scholarship Council (File No. 201208420115).

## REFERENCES CITED

- Ali, M., Emsley, A. M., Herman, H., and Heywood, R. J. (2001). "Spectroscopic studies of the ageing of cellulosic paper," *Polymer* 42(7), 2893-2900. DOI: 10.1016/S0032-3861(00)00691-1
- Alves, A., Schwanninger, M., Pereira, H., and Rodrigues, J. (2006). "Calibration of NIR to assess lignin composition (H/G ratio) in maritime pine wood using analytical pyrolysis as the reference method," *Holzforschung* 60(1), 29-31. DOI: 10.1515/HF.2006.006
- Arana, I., Jarén, C., and Arazuri, S. (2005). "Maturity, variety and origin determination in white grapes (*Vitis Vinifera* L.) using near infrared reflectance technology," *J. Near Infrared Spec.* 13(6), 349-357. DOI: 10.1255/jnirs.566
- Bailleres, H., Davrieux, F., and Ham-Pichavant, F. (2002). "Near infrared analysis as a tool for rapid screening of some major wood characteristics in a eucalyptus breeding program," *Ann. For. Sci.* 59(5-6), 479-490. DOI: 10.1051/forest:2002032
- Barrière, Y., Thomas, J., and Denoue, D. (2008). "QTL mapping for lignin content, lignin monomeric composition, *p*-hydroxycinnamate content, and cell wall digestibility in the maize recombinant inbred line progeny F838 × F286," *Plant Sci.* 175(4), 585-595. DOI:10.1016/j.plantsci.2008.06.009
- Barton, F. E., Himmelsbach, D. S., Duckworth, J. H., and Smith, M. J. (1992). "Two-dimensional vibration spectroscopy: Correlation of mid- and near-infrared regions," *Appl. Spectrosc.* 46(3), 420-429. DOI: 10.1366/0003702924125375
- Boerjan, W., Ralph, J., and Baucher, M. (2003). "Lignin biosynthesis," *Annu. Rev. Plant*

- Biol.* 54(1), 519-546. DOI: 10.1146/annurev.arplant.54.031902.134938
- Bonawitz, N. D., and Chapple, C. (2013). "Can genetic engineering of lignin deposition be accomplished without an unacceptable yield penalty?," *Curr. Opin. Biotech.* 24(2), 336-343. DOI: 10.1016/j.copbio.2012.11.004
- Bonawitz, N. D., Kim, J. I., Tobimatsu, Y., Ciesielski, P. N., Anderson, N. A., Ximenes, E., Maeda, J., Ralph, J., Donohoe, B. S., Ladisch, M., *et al.* (2014). "Disruption of mediator rescues the stunted growth of a lignin-deficient *Arabidopsis* mutant," *Nature* 509(7500), 376-380. DOI: 10.1038/nature13084
- Chen, F., and Dixon, R. A. (2007). "Lignin modification improves fermentable sugar yields for biofuel production," *Nat. Biotechnol.* 25(7), 759-761. DOI: 10.1038/nbt1316
- Cowe, I. A., and McNicol, J. W. (1985). "The use of principal components in the analysis of near-infrared spectra," *Appl. Spectrosc.* 39(2), 257-265. DOI: 10.1366/0003702854248944
- Ding, S. Y., Liu, Y. S., Zeng, Y. N., Himmel, M. E., Baker, J. O., and Bayer, E. A. (2012). "How does plant cell wall nanoscale architecture correlate with enzymatic digestibility?," *Science* 338(6110), 1055-1060. DOI: 10.1126/science.1227491
- Domínguez-Escribá, L., and Porcar, M. (2010). "Rice straw management: The big waste," *Biofuel. Bioprod. Bior.* 4(2), 154-159. DOI: 10.1002/bbb.196
- Foster, C. E., Martin, T. M., and Pauly, M. (2010). "Comprehensive compositional analysis of plant cell walls (lignocellulosic biomass) part I: Lignin," *Jove-J. Vis. Exp.* (37), e1745. DOI: 10.3791/1745
- Fu, C. X., Mielenz, J. R., Xiao, X. R., Ge, Y. X., Hamilton, C. Y., Jr., Rodriguez, M., Chen, F., Foston, M., Ragauskas, A., Bouton, J., *et al.* (2011). "Genetic manipulation of lignin reduces recalcitrance and improves ethanol production from switchgrass," *P. Natl. Acad. Sci.-USA* 108(9), 3803-3808. DOI: 10.1073/pnas.1100310108
- Hoang, N. V., Furtado, A., Donnan, L., Keffe, E. C., Botha, F. C., and Henry, R. J. (2017). "High-throughput profiling of the fiber and sugar composition of sugarcane biomass," *Bioenerg. Res.* 10(2), 400-416. DOI: 10.1007/s12155-016-9801-8
- Huang, J., Xia, T., Li, A., Yu, B., Li, Q., Tu, Y., Zhang, W., Yi, Z., and Peng, L. (2012). "A rapid and consistent near infrared spectroscopic assay for biomass enzymatic digestibility upon various physical and chemical pretreatments in *Miscanthus*," *Bioresource Technol.* 121(10), 274-281. DOI: 10.1016/j.biortech.2012.06.015
- Jin, S., and Chen, H. (2007). "Near-infrared analysis of the chemical composition of rice straw," *Ind. Crop. Prod.* 26(2), 207-211. DOI: 10.1016/j.indcrop.2007.03.004
- Jin, X., Chen, X., Shi, C., Li, M., Guan, Y., Yu, C. Y., Yamada, T., Sacks, E. J., and Peng, J. (2017). "Determination of hemicellulose, cellulose and lignin content using visible and near infrared spectroscopy in *Miscanthus sinensis*," *Bioresource Technol.* 241(2017), 603-609. DOI: 10.1016/j.biortech.2017.05.047
- Jourdes, M., Cardenas, C. L., Laskar, D. D., Moinuddin, S. G., Davin, L. B., and Lewis, N. G. (2007). "Plant cell walls are enfeebled when attempting to preserve native lignin configuration with poly-*p*-hydroxycinnamaldehydes: Evolutionary implications," *Phytochemistry* 68(14), 1932-1956. DOI: 10.1016/j.phytochem.2007.03.044
- Li, F., Zhang, M., Guo, K., Hu, Z., Zhang, R., Feng, Y., Yi, X., Zou, W., Wang, L., Wu, C., *et al.* (2015). "High-level hemicellulosic arabinose predominately affects lignocellulose crystallinity for genetically enhancing both plant lodging resistance and biomass enzymatic digestibility in rice mutants," *Plant Biotechnol. J.* 13(4), 514-

525. DOI: 10.1111/pbi.12276
- Maranan, M. C., and Laborie, M. P. G. (2008). "Rapid prediction of the chemical traits of hybrid poplar with near infrared spectroscopy," *J. Biobased Mater. Bio.* 2(1), 57-63. DOI: 10.1166/jbmb.2008.202
- Michell, A. J., and Schimleck, L. R. (1996). "NIR spectroscopy of woods from *Eucalyptus globulus*," *Appita J.* 49(1), 23-26. DOI: 10.1007/s00226-010-0368-9
- Owens, B., McCann, M. E., McCracken, K. J., and Park, R. S. (2009). "Prediction of wheat chemical and physical characteristics and nutritive value by near-infrared reflectance spectroscopy," *Brit. Poultry Sci.* 50(1), 103-122. DOI: 10.1080/00071660802635347
- Payne, C. E., and Wolfrum, E. J. (2015). "Rapid analysis of composition and reactivity in cellulosic biomass feedstocks with near-infrared spectroscopy," *Biotechnol. Biofuels* 8(1), 43. DOI: 10.1186/s13068-015-0222-2
- Penning, B. W., Sykes, R. W., Babcock, N. C., Dugard, C. K., Klimek, J. F., Gamblin, D., Davis, M., Filley, T. R., Mosier, N. S., Weil, C. F., *et al.* (2014). "Validation of PyMBMS as a high-throughput screen for lignin abundance in lignocellulosic biomass of grasses," *Bioenerg. Res.* 7(3), 899-908. DOI: 10.1007/s12155-014-9410-3
- Robinson, A. R., and Mansfield, S. D. (2009). "Rapid analysis of poplar lignin monomer composition by a streamlined thioacidolysis procedure and near-infrared reflectance-based prediction modeling," *Plant J.* 58(4), 706-714. DOI: 10.1111/j.1365-3113X.2009.03808.x
- Schwanninger, M., Rodrigues, J. C., and Fackler, K. (2011). "A review of band assignments in near infrared spectra of wood and wood components," *J. Near Infrared Spec.* 19(5), 287-308. DOI: 10.1255/jnirs.955
- Shenk, J. S., and Westerhaus, M. O. (1994). "The application of near infrared reflectance spectroscopy (NIRS) to forage analysis," *Forage Quality, Evaluation, and Utilization* 10, 406-449. DOI: 10.2134/1994.foragequality.c10
- Wang, Y., Fan, C., Hu, H., Li, Y., Sun, D., and Peng, L. (2016). "Genetic modification of plant cell walls to enhance biomass yield and biofuel production in bioenergy crops," *Biotechnol. Adv.* 34(5), 997-1017. DOI: 10.1016/j.biotechadv.2016.06.001
- Wen, J. L., Sun, S. L., Xue, B. L., and Sun, R. C. (2013). "Recent advances in characterization of lignin polymer by solution-state nuclear magnetic resonance (NMR) methodology," *Materials* 6(1), 359-391. DOI: 10.3390/ma6010359
- Wu, L., Li, M., Huang, J., Zhang, H., Zou, W., Hu, S., Li, Y., Fan, C., Zhang, R., Jing, H., *et al.* (2015). "A near infrared spectroscopic assay for stalk soluble sugars, bagasse enzymatic saccharification and wall polymers in sweet sorghum," *Bioresource Technol.* 177, 118-124. DOI: 10.1016/j.biortech.2014.11.073
- Xu, F., Yu, J. M., Tesso, T., Dowell, F., and Wang, D. H. (2013). "Qualitative and quantitative analysis of lignocellulosic biomass using infrared techniques: A mini-review," *Appl. Energ.* 104(2), 801-809. DOI: 10.1016/j.apenergy.2012.12.019
- Xu, F., Zhou, L., Zhang, K., Yu, J. M., and Wang, D. H. (2015). "Rapid determination of both structural polysaccharides and soluble sugars in sorghum biomass using near-infrared spectroscopy," *Bioenerg. Res.* 8(1), 130-136. DOI: 10.1007/s12155-014-9511-z
- Yamada, T., Yeh, T. F., Chang, H. M., Li, L., Kadla, J. F., and Chiang, V. L. (2006). "Rapid analysis of transgenic trees using transmittance near-infrared spectroscopy (NIR)," *Holzforschung* 50(1), 1040-1028. DOI: 10.1515/HF.2006.005

- Yan, B., Liu, R., Li, Y., Wang, Y., Gao, G., Zhang, Q., Liu, X., Jiang, G., and He, Y. (2014). "QTL analysis on rice grain appearance quality, as exemplifying the typical events of transgenic or backcrossing breeding," *Breeding Sci.* 64(3), 231-239. DOI: 10.1270/jsbbs.64.231
- Yang, Z., Li, K., Zhang, M., Xin, D., and Zhang, J. (2016). "Rapid determination of chemical composition and classification of bamboo fractions using visible-near infrared spectroscopy coupled with multivariate data analysis," *Biotechnol. Biofuels* 9(1), 35. DOI: 10.1186/s13068-016-0443-z
- Zeng, Y., Zhao, S., Yang, S., and Ding, S. Y. (2014). "Lignin plays a negative role in the biochemical process for producing lignocellulosic biofuels," *Curr. Opin. Biotech.* 27(6), 38-45. DOI: 10.1016/j.copbio.2013.09.008

Article submitted: September 30, 2017; Peer review completed: January 28, 2018;  
Revised version received and accepted: March 9, 2018; Published: March 14, 2018.  
DOI: 10.15376/biores.13.2.3284-3299