



A near infrared spectroscopic assay for stalk soluble sugars, bagasse enzymatic saccharification and wall polymers in sweet sorghum



Leiming Wu^{a,b,c}, Meng Li^{b,d,e}, Jiangfeng Huang^{a,b,c}, Hui Zhang^{a,b,c}, Weihua Zou^{a,b,c}, Shiwei Hu^{a,b,c}, Ying Li^{a,b,c}, Chunfen Fan^{a,b,c}, Rui Zhang^{a,b}, Haichun Jing^f, Liangcai Peng^{a,b,c}, Shengqiu Feng^{a,b,c,*}

^aNational Key Laboratory of Crop Genetic Improvement, Huazhong Agricultural University, Wuhan 430070, China

^bBiomass and Bioenergy Research Centre, Huazhong Agricultural University, Wuhan 430070, China

^cCollege of Plant Science and Technology, Huazhong Agricultural University, Wuhan 430070, China

^dNational Energy R&D Center for Non-food Biomass, Beijing 100193, China

^eCollege of Agronomy and Biotechnology, China Agricultural University, Beijing 100193, China

^fInstitute of Botany, Chinese Academy of Sciences, Beijing 100093, China

HIGHLIGHTS

- Sweet sorghum is a leading bioenergy crop with rich soluble sugars and high biomass.
- Diverse sweet sorghum samples are suitable for a stable and consistent NIRS assay.
- Nine calibration optimal equations of NIRS with much high R^2 , R^2_{cv} and RPD values.
- NIRS simultaneous assay for stalk sugars, bagasse saccharification and wall polymers.
- NIRS applicable for high-throughput screening of bioenergy sweet sorghums.

ARTICLE INFO

Article history:

Received 10 October 2014

Received in revised form 11 November 2014

Accepted 15 November 2014

Available online 21 November 2014

Keywords:

Sweet sorghum

Near infrared spectroscopy (NIRS)

Fermentable sugars

Cell wall polymers

Bioenergy crops

ABSTRACT

In this study, 123 sweet sorghum (*Sorghum bicolor* L.) accessions and 50 mutants were examined with diverse stalk soluble sugars, bagasse enzymatic saccharification and wall polymers, indicating the potential near infrared spectroscopy (NIRS) assay for those three important parameters. Using the calibration and validation sets and modified squares method, nine calibration optimal equations were generated with high determination coefficient on the calibration (R^2) (0.81–0.99), cross-validation (R^2_{cv}) (0.77–0.98), and the ratio performance deviation (RPD) (2.07–7.45), which were at first time applied by single spectra for simultaneous assay of stalk soluble sugars, bagasse hydrolyzed sugars, and three major wall polymers in bioenergy sweet sorghum.

© 2014 Elsevier Ltd. All rights reserved.

1. Introduction

Lignocellulose represents an enormous biomass for biofuels in bioenergy crops (Chen and Peng, 2013; Cotton et al., 2013). Principally, biomass process for bioethanol production involves three major steps: physical and chemical pretreatments for cell wall dissociation, enzymatic hydrolysis for soluble sugar release, and yeast fermentation for ethanol production. However, plant cell wall recalcitrance fundamentally determines a costly biomass

process with numerous compounds that inhibit biofuel conversion. As a promising solution, genetic modification of plant cell walls has been proposed for reducing biomass recalcitrance in bioenergy crops (Xie and Peng, 2011).

Among the energy crops, sweet sorghum (*Sorghum bicolor* L.) exceptionally contains a rich fermentable-sugar at stalk and abundant degradable-lignocellulose at bagasse (Byrt et al., 2011; Zegada-Lizarazu and Monti, 2012; Li et al., 2014a). As a typical C4 species, sweet sorghum is regarded as a leading candidate of bioenergy crops because of its rapid growth rate, great water-use efficiency, strong soil tolerance and wide adaptability (Li and Chan-Halbrendt, 2009; Xie and Su, 2012).

It has been reported that both stalk soluble sugars and dry bagasse could not significantly affect lignocellulose saccharification

* Corresponding author at: National Key Laboratory of Crop Genetic Improvement, Huazhong Agricultural University, Wuhan 430070, China. Tel.: +86 27 87281765; fax: +86 27 87280016.

E-mail address: fengsq@mail.hzau.edu.cn (S. Feng).

URL: <http://bbrc.hzau.edu.cn> (S. Feng).

under various chemical pretreatments by determining large populations of sweet sorghum accessions, suggesting the potential for selecting the elite sweet sorghum accessions with high stalk soluble sugars and bagasse saccharification yields (Li et al., 2014a). Notably, three major wall polymers have been determined with distinctive effects on biomass enzymatic digestibility in sweet sorghum accessions (Li et al., 2014a). Hence, it becomes important to find out an approach for quickly detecting multiple parameters (stalk soluble sugars, biomass digestibility and wall polymers) in bioenergy sweet sorghum accessions.

However, the traditional laboratory methods are labor-intensive, time-consuming and expensive for determining the stalk soluble sugars, bagasse enzymatic hydrolysis and the cell wall polymers (Roberts et al., 2011; Li et al., 2014a). Hence, the classic methods could not meet the need for efficiently selecting the elite sweet sorghum cultivars from large populations of germplasm distributed over the world. Near infrared spectroscopic assay (NIRS), which combines laboratory data and the spectral information, has been considered as a powerful tool for high-throughput screening of large populations of plant samples. For example, NIRS has been used for predicting soluble sugars in sugarcane (Rodriguez-Saona et al., 2001) and biomass enzymatic digestibility upon various physical and chemical pretreatments in *Miscanthus* (Huang et al., 2012). It could also analyze sweet sorghum bagasse digestibility and crude protein (Roberts et al., 2011). Furthermore, the soluble sugars and cell wall compositions in sweet sorghum have been separately measured by Fourier transform infrared (FTIR) spectra (Martin et al., 2013). However, little is reported about NIRS simultaneous assay for multiple parameters of plant samples, in particular on sweet sorghum.

In our previous studies, we have determined diverse soluble sugar levels at stalk, varied biomass enzymatic digestibility and cell wall compositions at bagasse in large populations of sweet sorghum germplasm accessions (Li et al., 2014a) and genetic mutants. In this work, we used the previous data and generated nine NIRS models for simultaneously predicting those three parameters in sweet sorghum samples with high R^2 , R^2_{cv} and RPD values, and low $SECV$ value.

2. Methods

2.1. Materials

123 sweet sorghum germplasm accessions were collected worldwide (63 accessions were mentioned in Li et al. (2014a)) and 50 sweet sorghum mutants were selected by generating EMS mutagenesis pools of wild type (Nengsi 1). The sweet sorghum samples were grown in Hubei experimental field in 2011, and the mature stalks were harvested from 90 to 140 days after sowing.

2.2. Measurement of dry matter and soluble sugars at stalks

Measurements of dry matter and soluble sugar level at mature stalks were performed as described by Li et al. (2014a). The mature stalks were harvested after seeds and leaves were removed. The remaining stalk tissues were dried at 50 °C after inactivation at 105 °C for 20 min. The dried stalks were weighed as dry matter and then ground into powders through a 40 mesh screen for further analysis of stalk soluble sugars, bagasse enzymatic hydrolysis and cell wall polymers.

The biomass sample (0.3 g) was added with 6 mL distilled water, and shaken at 150 rpm for 2 h at 50 °C. After centrifugation at 3000g for 10 min, the biomass residues were washed five times with distilled water, and all supernatants were collected as stalk soluble sugars (pentoses and hexoses). Soluble sugars were

detected by colorimetric assay as described below. All experiments were performed in biological triplicate.

2.3. Determination of sugars yield released from enzymatic hydrolysis at bagasse

The biomass enzymatic digestibility (saccharification) was defined by determining the sugars yield released from enzymatic hydrolysis of bagasse after 1% NaOH (w/v) pretreatment as described by Li et al. (2014a). The bagasse sample was washed 5 times with 10 mL distilled water, and added with 6 mL 1% NaOH in the test tube. The sample tube was then shaken at 150 rpm for 2 h at 50 °C. After centrifugation at 3000g for 5 min, the pellet was washed three times with 10 mL distilled water and stored for enzyme hydrolysis. All experiments were conducted in biological triplicate.

The remaining samples from 1% NaOH pretreatment were washed 5 times with 6 mL distilled water, and once with 10 mL mixed-cellulases reaction buffer (0.2 M acetic acid-sodium acetate, pH 4.8). The washed residues were added with 0.16% (w/v) mixed-cellulases (containing β -glucanase $\geq 2.98 \times 10^4$ U, cellulase ≥ 298 U and xylanase $\geq 4.8 \times 10^4$ U from Imperial Jade Biotechnology Co. Ltd.) to 6 mL. During the enzymatic hydrolysis, the samples were shaken under 150 rpm at 50 °C for 48 h. After centrifugation at 3000g for 10 min, the supernatants were collected for bagasse sugars (pentoses and hexoses) yield assay. The samples without mixed-cellulases addition were shaken for 48 h at 50 °C as the control. All samples were carried out in biological triplicate.

2.4. Colorimetric assay of hexoses and pentoses

UV-VIS Spectrometer (V-1100D, Shanghai MAPADA Instruments Co. Ltd. Shanghai, China) was applied for the absorbance reading. Hexoses were detected using the anthrone/ H_2SO_4 method (Fry, 1988), whereas pentoses were detected using the orcinol/HCl method (Dische, 1962). Anthrone was purchased from Sigma-Aldrich Co. LLC., and ferric chloride and orcinol were obtained from Sinopharm Chemical Reagent Co. Ltd. The standard curves for hexoses and pentoses were plotted using D-glucose and D-xylose as standards (purchased from Sinopharm Chemical Reagent Co. Ltd.) respectively.

2.5. Plant cell wall fractionation and wall polymers measurement

Plant cell wall fractionation procedure was used for extraction of cellulose and hemicelluloses as described by Peng et al. (2000) and Li et al. (2014a). After the soluble sugars, lipids, starch and pectin were consecutively removed, the remaining pellet was extracted with 4 M KOH with 1.0 mg/mL sodium borohydride for 1 h at 25 °C, and washed with distilled water until soluble sugars were not detectable. The combined supernatant was then neutralized, dialyzed and lyophilized as KOH-extractable hemicelluloses. The remaining pellet was extracted with H_2SO_4 (67%, v/v) for 1 h at 25 °C and the supernatants were collected for determination of free hexoses and pentoses as total cellulose and non-KOH-extractable hemicelluloses (Li et al., 2014a). The total hemicelluloses were calculated by the KOH-extractable and non-KOH-extractable hemicelluloses as described above. Total lignin was assayed using a two-step acid hydrolysis method according to the Laboratory Analytical Procedure of the National Renewable Energy Laboratory (Sluiter et al., 2008) with minor modifications. The acid-insoluble lignin was calculated gravimetrically after correction for ash, and the acid-soluble lignin was measured using UV spectroscopy (Wu et al., 2013). All experiments were carried out in biological triplicate.

2.6. Near infrared spectra (NIRS) collection and variability analysis

Spectra collection and variability analysis of NIRS were described by Huang et al. (2012) with minor modification. Using a XDS Rapid Content™ Analyzer (FOSS, Co. LLC, Denmark) equipped with a dual detector system: silicon (400–1100 nm), lead sulfide (1100–2500 nm) and the ISIScan™ software, the spectral absorbance values were recorded as $\log(1/R)$, where R is the sample reflectance.

The dried samples were placed into a mini-sample cup (standard ring cup). The reflectance each sample was recorded in triplicate at wavelengths ranging from 400 nm to 2500 nm with 2 nm intervals at room temperature. The reflectance values were then averaged for further analysis. The WinISI III software package (Version 1.50e, Infrasoft International LLC) 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). Full spectra wavelengths (408–2492 nm) were selected and pretreated with “1, 4, 4, 1” and “None” treatments for PCA. The global H (GH) of each sample was determined using the measured distance from mean. The sample, which is the GH outlier (GH > 3.0), was finally eliminated after the PCA.

2.7. Near infrared spectra data calibration and validation

Procedures for spectra data calibration and validation were described by Huang et al. (2012) with minor modifications. Derivative spectra and the gap and smoothing of derivative spectra were performed before calibration (Giese and French, 1955). Two mathematical treatments, “0, 0, 1, 1” and “1, 4, 4, 1”, five scatter correction methods, namely standard normal variate (SNV), detrend only (DET), standard multiple scatter correction (MSC), a combination of SNV and DET (NSVD), weighted multiple scatter correction (WMSC), and three wavelength ranges (408–2492 nm, 780–2492 nm, and 1108–2492 nm) were selected to obtain the best calibration equation.

The modified partial least squares (MPLS) method was performed to provide a prediction equation (Shenk and Westerhaus, 1991). Cross-validation is recommended when developing MPLS equations in order to select the optimal number of factors and to avoid over-fitting (Shenk and Westerhaus, 1995). What is more, one of every four samples sorted based on the laboratory value of predicted matters were selected into the external validation.

The coefficient determination of the calibration/cross-validation/external validation ($R^2/R^2_{cv}/R^2_{ev}$), the standard error of calibration (SEC), the standard error of prediction in external validation (SEP), and the standard error of cross-validation (SECV)

were used to evaluate the best equations. What is more, the ratio performance deviation (RPD), was used to evaluate the prediction capacity of the equation (Williams and Sobering, 1996).

3. Results and discussion

3.1. Diversity of stalk soluble sugars, bagasse saccharification and wall polymers

In the previous study, we have identified the 63 representative sweet sorghum germplasm accessions that displayed a varied soluble sugar level at stalk and diverse cell wall composition at bagasse (Li et al., 2014a). In this work, we characterized total 173 sweet sorghum samples including 123 sweet sorghum germplasm accessions and 50 genetic mutants, which represented two major genetic resources used in sweet sorghum breeding. In general, both sweet sorghum accessions and mutants were respectively examined with diverse dry matter and varied soluble sugars at stalk (Table 1 and Fig. A1). In particular, the 123 accessions showed the soluble hexoses level ranged from 0.71% to 40.82% (% dry matter). But, the mutants showed relatively higher levels of soluble

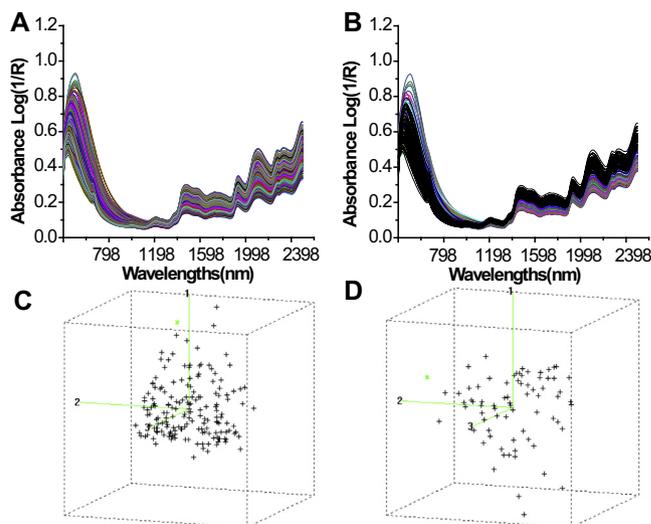


Fig. 1. Variation of NIRS absorbance spectra in sorghum samples. (A) 519 original scanned spectra; (B) averaged 173 spectra (colorful: the 123 accessions; black: the 50 mutants); (C) 3D plot of the PCA scores of 173 spectra (green: the GH outlier sample); (D) 3D plot of the PCA scores of 74 spectra (green: the GH outlier sample). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 1

Variations of dry matter, stalk sugars and bagasse hydrolyzed sugars in sweet sorghum accessions and mutants.

	Dry matter	Stalk hexoses	Stalk sugars	Bagasse hexoses yield	Bagasse sugars yield	Total hexoses in stalk and bagasse	Total sugars in stalk and bagasse
123 accessions							
Mean value	67.15 ± 42.85g* (5.58 g– 231.27 g#)	13.83 ± 10.04% (0.71–40.82%)	15.67 ± 10.54% (1.62–44.13%)	19.94 ± 3.22% (9.87–28.02%)	37.48 ± 5.96% (20.59–56.70%)	33.77 ± 8.29% (21.62–63.89%)	53.16 ± 7.50% (40.57–78.54%)
Coefficient variation	63.81%	72.57%	67.25%	16.15%	15.90%	24.55%	14.11%
50 mutants							
Mean value	191.75 ± 63.99g (106.14 g– 430.38 g)	40.32 ± 5.84% (26.79– 50.10%)	44.11 ± 6.29% (29.61– 54.66%)	16.63 ± 1.94% (13.00–21.69%)	27.71 ± 2.85% (22.79–34.23%)	57.05 ± 5.98% (42.69–67.03%)	71.82 ± 6.34% (57.61–82.22%)
Coefficient variation	33.37%	14.45%	14.25%	11.68%	10.30%	10.48%	8.83%

* Indicated mean ± SD.

Indicated the minimum and maximum values.

Table 2
Calibration and validation sets for sugar contents and cell wall polymers in 173 sorghum samples.

	Calibration					Validation				
	N ^a	Min ^b	Max ^c	Mean	SD ^d	N	Min	Max	Mean	SD
Stalk hexoses	129	0.71	49.55	21.41	15.24	44	0.72	50.10	21.84	14.75
Stalk sugars	129	1.62	53.93	23.76	16.20	44	1.62	54.66	24.28	15.73
Bagasse hexoses yield	129	9.87	28.02	18.96	3.38	44	12.90	25.24	19.05	2.96
Bagasse sugars yield	129	20.59	56.70	34.77	7.23	44	23.54	45.78	34.32	5.76
Total hexoses in stalk and bagasse	129	21.62	66.01	40.36	13.16	44	22.30	67.03	40.88	12.95
Total sugars in stalk and bagasse	129	40.93	81.60	58.53	11.02	44	40.57	82.22	58.59	11.49
Cellulose	55	18.88	34.19	25.72	4.39	19	18.03	34.06	25.59	4.66
Hemicelluloses	55	13.43	32.56	23.82	4.93	19	12.97	31.33	23.44	5.20
Lignin	55	11.37	19.45	15.52	1.96	19	12.07	19.09	15.12	2.05

^a N, sample number.

^b Min, minimum value.

^c Max, maximum value.

^d SD, standard deviation.

Table 3
Calibration and validation statistics for equations generated for prediction of stalk soluble sugars, bagasse sugars yield and total hexoses and sugars both in stalk and bagasse (% dry matter).

	Calibration							Cross validation				External validation			
	N ^a	Mean	SD ^b	SEC ^c	R ^{2d}	Spectrum range (nm)	DT ^e	SCM ^f	Terms ^g	SECV ^h	R ^{2cv} ⁱ	RPD ^j	N	SEP ^k	R ^{2ev} ^l
Stalk hexoses	115	21.32	15.08	1.53	0.99	780–2492	1,4,4,1	MSC ^m	7	1.76	0.99	8.56	43	5.75	0.86
Stalk sugars	115	24.49	16.24	1.82	0.99	408–2492	1,4,4,1	MSC	8	2.21	0.98	7.35	41	2.34	0.98
Bagasse hexoses yield	123	18.98	3.11	1.39	0.80	780–2492	0,0,1,1	DET ⁿ	10	1.65	0.73	1.89	44	2.17	0.63
Bagasse sugars yield	121	34.37	6.58	2.36	0.87	1180–2492	0,0,1,1	NSVD ^o	8	2.67	0.84	2.47	44	4.07	0.71
Total hexoses in stalk and bagasse	119	40.47	12.94	3.04	0.94	1180–2492	1,4,4,1	DET	3	3.16	0.94	4.09	42	4.63	0.88
Total sugars in stalk and bagasse	123	58.69	11.08	3.93	0.87	780–2492	0,0,1,1	None ^p	9	4.33	0.85	2.56	44	4.89	0.82

^a N, sample number.

^b SD, standard deviation of reference value.

^c SEC, standard error of calibration.

^d R², determination coefficient.

^e DT, derivative treatment.

^f SCM, scatter correction methods.

^g Terms, number of principal component used for calibration.

^h SECV, standard error of cross validation.

ⁱ R^{2cv}, determination coefficient of cross validation.

^j RPD, ratio performance deviation (SD/SECV).

^k SEP, standard error of prediction in external validation.

^l R^{2ev}, determination coefficient of external validation.

^m MSC, standard multiple scatter.

ⁿ DET, detrend only.

^o NSVD, a combination of SNV and DET.

^p None, none scatter correction.

hexoses ranging from 26.79% to 50.1%, compared with the germplasm accessions. As genetic mutants could be directly used in bioenergy crop breeding (Xie and Peng, 2011), the test samples combining germplasm accessions with genetic mutants should be advantageous for selecting the optimal genetic bioenergy sweet sorghum lines by NIRS assay in this work. In addition, the sweet sorghum samples also displayed a diverse total sugars (hexoses and pentoses) level at stalks.

Furthermore, we determined either the hexoses or total sugars (hexoses and pentoses) yield released from enzymatic hydrolysis of bagasse after 1% NaOH pretreatment in the sweet sorghum accessions and mutants (Table 1 and Fig. A1). As a result, the 123 sweet sorghum accessions exhibited the hexoses yield ranged from 9.87% to 28.02%, and the mutants varied from 13% to 21.69%. Compared with the hexoses yield, both the sweet sorghum samples could even show high variations of total sugars (hexoses and pentoses) yield at bagasse from 20.59% to 56.7% in accessions and from 22.79% to 34.23% in mutants. Hence, both sweet sorghum accessions and mutants also exhibited high variations of total hexoses

and sugars (hexoses and pentoses) released from stalk and bagasse (Table 1 and Fig. A1).

As biomass enzymatic digestibility (hexoses and sugars yield) is affected by plant cell wall compositions (Xu et al., 2012; Li et al., 2013; Wu et al., 2013; Zhang et al., 2013; Li et al., 2014a,b,c; Jia et al., 2014), we determined three major wall polymers (cellulose, hemicelluloses and lignin) level in the 59 accessions and 15 mutants (Table A1). Notably, the sweet sorghum samples also showed a diverse cell wall composition (Fig. A1). Taken all together, therefore, the diversity of stalk soluble sugars, bagasse enzymatic digestibility and cell wall compositions suggested that the sweet sorghum samples could be applicable for NIRS modeling (Huang et al., 2012).

3.2. NIRS modeling for predicting stalk soluble sugars and bagasse biomass saccharification

Using 123 sweet sorghum accessions and 50 mutants, we performed NIRS spectra collections with three replicates (Fig. 1A).

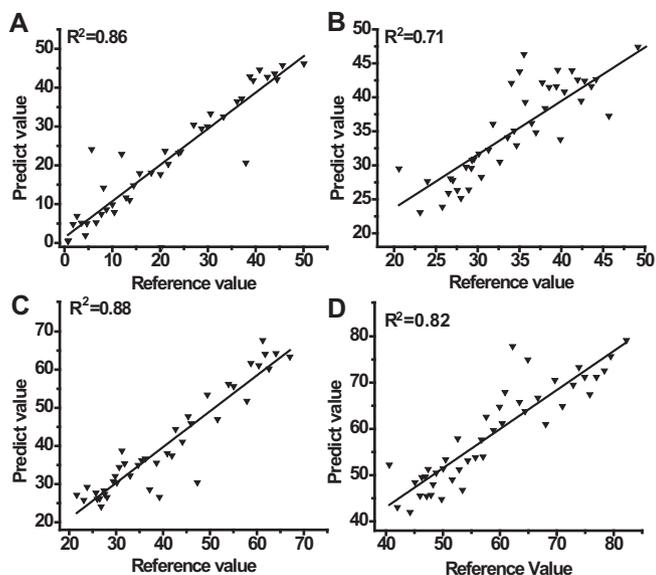


Fig. 2. Correlation between the predicted and measured values for stalk soluble sugars, bagasse sugars yield and total sugars. (A) Stalk hexoses; (B) bagasse sugars yield; (C) total hexoses in stalk and bagasse; (D) total sugars in stalk and bagasse.

For near infrared assay, the averaged spectra were generated by means of the WinISI software. The different distribution of the averaged spectra between the 123 accessions and 50 mutants was shown in Fig. 1B, consistent with the values of various soluble sugars and wall polymers (Tables 1 and A1). With regard to the distribution of the total 173 samples, a 3D plot was given to present the data at three dimensions. Using the PCA algorithm of 173 spectra population, 17 components were kept to interpret the 99.72% of the variance, and just one sample was eliminated as a GH outlier ($GH \geq 3.0$) (Fig. 1C). Hence, the symmetrical distribution indicated the spectra behaved well in calibration and prediction.

One of every four samples based on the value of predicted objects were included into the validation sets, and the remaining was used for the calibration sets. As a result, the validation sets included 44 samples, and the calibration sets contained 129 samples. The calibration and validation sets for stalk soluble sugars, bagasse hydrolyzed sugars and total sugars in stalk and bagasse were compared and the variations between the calibration and validation sets were calculated (Table 2 and Fig. A2).

The calibration was processed with five cross-validation groups, a maximum of 11 terms and two outlier elimination passes, and the GH and T outlier samples were then eliminated. And 36 calibration equations were produced by six scatter correction methods, two derivative treatments, and three spectrum regions selections, in which the best calibration equation subjective to high R^2 , R^2cv and low SEC and $SECV$ (Huang et al., 2012). Hence, the calibration optimal equations for stalk soluble sugars, bagasse hydrolyzed sugars and total sugars had R^2 values ranging from 0.80 to 0.99 during calibration (Table 3). By comparison, the prediction for stalk hexoses was almost entirely accurate, consistent with the previous reports (Rodriguez-Saona et al., 2001; Duarte et al., 2002; Martin et al., 2013). With respect to the bagasse hydrolyzed sugars, the R^2 value for the hexoses in bagasse was 0.80, and the R^2 value for the total hexoses in stalk and bagasse was also high up to 0.94 (Table 3).

According to the cross-validation data, the calibration optimal equations for stalk soluble sugars, bagasse hydrolyzed sugars and total sugars showed the R^2cv values ranging from 0.73 to 0.99 (Table 3), in agreement with the calibration data. Furthermore, we conducted an external validation using the “Monitor results” module of the WinISI III software in order to compare the correlations between the predicted and reference values. All of the R^2ev values were higher than 0.71 except for the R^2ev value of the bagasse hydrolyzed hexoses (Fig. 2 and Table 3).

In sum, the calibration optimal equations for stalk soluble sugars, bagasse enzymatic sugars and total sugars in stalk and bagasse displayed high R^2 , R^2cv and low SEC and $SECV$ values (Table 3), which suggested a high applicability for near infrared modeling.

3.3. NIRS modeling for cell wall polymers assay

It has been reported that plant cell wall compositions could determine biomass enzymatic digestibility in plants (Chen and Dixon, 2007; Fu et al., 2011; Xu et al., 2012; Li et al., 2013; Wu et al., 2013; Zhang et al., 2013; Li et al., 2014a,b,c; Jia et al., 2014). In particular, cellulose and lignin exhibit significantly negative effects on biomass saccharification in sweet sorghum (Li et al., 2014a). Hence, it becomes important to perform high-throughput NIRS assay for wall polymers. Based on their stalk soluble sugars, bagasse dry matter and hydrolyzed sugars, total 74 representative sweet sorghum samples were selected in this study including 59 sweet sorghum accessions and 15 mutants. As described previously, the 74 sweet samples exhibited a diverse cell wall composi-

Table 4
Calibration and validation statistics for equations generated for prediction of cell wall polymers.

	Calibration								Cross validation				External validation		
	N^a	Mean	SD^b	SEC^c	R^2d	Spectrum range(nm)	DT^e	SCM^f	Terms ^g	$SECV^h$	R^2cv^i	RPD^j	N	SEP^k	R^2ev^l
Cellulose	51	25.72	4.43	1.09	0.94	408–2492	0,0,1,1	SNV ^m	5	1.23	0.92	3.59	19	1.15	0.94
Hemicelluloses	54	23.93	4.82	1.47	0.91	780–2492	1,4,4,1	DET ⁿ	3	1.74	0.87	2.78	19	1.73	0.90
Lignin	50	15.44	1.84	0.49	0.93	408–2492	1,4,4,1	None ^o	5	0.77	0.83	2.40	19	0.86	0.87

^a N , sample number.

^b SD , standard deviation of reference value.

^c SEC , standard error of calibration.

^d R^2 , determination coefficient.

^e DT , derivative treatment.

^f SCM , scatter correction methods.

^g Terms, number of principal component used for calibration.

^h $SECV$, standard error of cross validation.

ⁱ R^2cv , determination coefficient of cross validation.

^j RPD , ratio performance deviation (SD/SEC).

^k SEP , standard error of prediction in external validation.

^l R^2ev , determination coefficient of external validation.

^m SNV, standard normal variate.

ⁿ DET, detrend only.

^o None, none scatter correction.

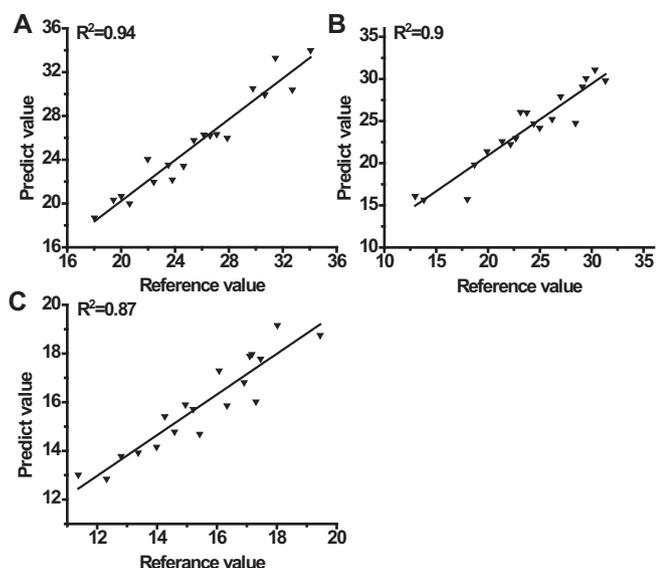


Fig. 3. Correlation between the predicted and measured values for cell wall polymers. (A) Cellulose; (B) hemicelluloses; (C) lignin.

Table 5

Statistics for calibrated optimal equations on stalk soluble sugars, bagasse sugars yield and total hexoses and sugars in stalk and bagasse.

	Calibration									Cross validation		
	N ^a	Mean	SD ^b	SEC ^c	R ^{2d}	Spectrum range(nm)	DT ^e	SCM ^f	Terms ^g	SECV ^h	R ^{2cv} ⁱ	RPD ^j
Stalk hexoses	156	21.69	15.32	1.87	0.99	1180–2492	1,4,4,1	None ^k	6	2.06	0.98	7.45
Stalk sugars	156	24.09	16.32	1.99	0.99	1180–2492	1,4,4,1	DET ^l	6	2.22	0.98	7.37
Bagasse hexoses yield	163	18.81	3.16	1.36	0.81	780–2492	1,4,4,1	DET	9	1.53	0.77	2.07
Bagasse sugars yield	158	34.16	6.37	2.17	0.88	1180–2492	1,4,4,1	NSVD ^m	8	2.43	0.85	2.62
Total hexoses in stalk and bagasse	159	40.68	13.17	2.39	0.97	780–2492	1,4,4,1	SNV ⁿ	9	2.83	0.95	4.65
Total sugars in stalk and bagasse	164	58.78	11.15	4.04	0.87	1180–2492	1,4,4,1	None	2	4.22	0.86	2.64

^a N, sample number.

^b SD, standard deviation of reference value.

^c SEC, standard error of calibration.

^d R², determination coefficient.

^e DT, derivative treatment.

^f SCM, scatter correction methods.

^g Terms, number of principal component used for calibration.

^h SECV, standard error of cross validation.

ⁱ R^{2cv}, determination coefficient of cross validation.

^j RPD, ratio performance deviation (SD/SECV).

^k None, none scatter correction.

^l DET, detrend only.

^m NSVD, a combination of SNV and DET.

ⁿ SNV, standard normal variate.

Table 6

Statistics for calibrated optimal equations on cell wall polymers.

	Calibration									Cross validation		
	N ^a	Mean	SD ^b	SEC ^c	R ^{2d}	Spectrum range (nm)	DT ^e	SCM ^f	Terms ^g	SECV ^h	R ^{2cv} ⁱ	RPD ^j
Cellulose	68	25.70	4.46	0.85	0.96	1180–2492	1,4,4,1	NSVD ^k	4	0.98	0.95	4.57
Hemicelluloses	71	23.88	4.94	1.38	0.92	408–2492	1,4,4,1	DET ^l	3	1.57	0.90	3.15
Lignin	70	15.49	1.95	0.51	0.93	1180–2492	1,4,4,1	None ^m	7	0.64	0.89	3.03

^a N, sample number.

^b SD, standard deviation of reference value.

^c SEC, standard error of calibration.

^d R², determination coefficient.

^e DT, derivative treatment.

^f SCM, scatter correction methods.

^g Terms, number of principal component used for calibration.

^h SECV, standard error of cross validation.

ⁱ R^{2cv}, determination coefficient of cross validation.

^j RPD, ratio performance deviation (SD/SECV).

^k NSVD, a combination of SNV and DET.

^l DET, detrend only.

^m None, none scatter correction.

tion, indicating those samples applicable for the NIRS assay (Fig. A1 and Table A1). As a result, the NIRS models for the three cell wall polymers all had high R² values over 0.9 during calibration (Table 4). Based on the cross-validation data, the R^{2cv} values of the calibration optimal equations for cellulose, hemicelluloses and lignin were 0.92, 0.87 and 0.83, respectively. Furthermore, the 1/4 samples for the external validation also showed the high correlations between NIRS predicted and reference values (Fig. 3). Therefore, the developed NIRS assay was well enough for fast determining cell wall polymers, which could be used to screen out large populations of sweet sorghum samples with high biomass enzymatic digestibility.

3.4. Simultaneous assay of stalk soluble sugars, bagasse hydrolyzed sugars, total sugars and cell wall polymers

Based on the obtained calibration and validation data as described above, all sweet sorghum samples were recalibrated, and six new calibration optimal equations were finally generated for simultaneously determining the stalk soluble sugars, bagasse hydrolyzed sugars, and total sugars (Table 5). Meanwhile, other three new calibration optimal equations were established for measuring cell wall compositions in sweet sorghum (Table 6). Notably,

all the nine calibration optimal equations exhibited much high R^2 (0.81–0.99), R^2_{cv} (0.77–0.98), and RPD (2.07–7.45) values, with relatively low $SECV$ values. Hence, the NIRS models could be powerful for sufficient and accurate predictions of stalk soluble sugars, bagasse hydrolyzed sugars, total sugars and cell wall polymers in sweet sorghum samples. To our knowledge, it was at the first time to provide a platform for fast and simultaneous assay of multiple traits/factors in sweet sorghum samples using NIRS, which should be applicable for bioenergy sweet sorghum breeding and biofuel process.

4. Conclusion

123 sweet sorghum accessions and 50 genetic mutants have been examined with diverse stalk soluble sugars, bagasse enzymatic digestibility and cell wall compositions. Using the NIRS of 173 sweet sorghum samples, nine calibration optimal equations have been generated with much high R^2 , R^2_{cv} , and RPD values and relatively low $SECV$. Those equations could be applied for simultaneously and rapidly predicting stalk soluble sugars, bagasse hydrolyzed sugars, total sugars and cell wall polymers. Therefore, the high-throughput NIRS models could offer the fast approach for bioenergy sweet sorghum selection.

Acknowledgements

This work was supported in part by grants from the Fundamental Research Funds for the Central Universities Project (2013QC042) and the 111 Project of Ministry of Education of China (B08032).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.biortech.2014.11.073>.

References

- Byrt, C., Grof, C., Furbank, R., 2011. C4 plants as biofuel feedstocks: optimising biomass production and feedstock quality from a lignocellulosic perspective. *J. Integr. Plant Biol.* 53, 120–135.
- Chen, F., Dixon, R., 2007. Lignin modification improves fermentable sugar yields for biofuel production. *Nat. Biotechnol.* 25, 759–761.
- Chen, P., Peng, L., 2013. The diversity of lignocellulosic biomass resources and their evaluation for use as biofuels and chemicals. In: Sun, J.Z., Ding, S.Y., Peterson, J.D. (Eds.), *Biological Conversion of Biomass for Fuels and Chemicals: Exploration from Natural Biomass Utilization Systems*. Royal Society of Chemistry, Oxfordshire, pp. 83–113.
- Cotton, J., Burow, G., Acosta-Martinez, V., Moore-Kucera, J., 2013. Biomass and cellulosic ethanol production of forage sorghum under limited water conditions. *Bioenergy Res.* 6, 711–718.
- Cowe, I., McNicol, J., 1985. The use of principle components in the analysis of near-infrared spectra. *Appl. Spectrosc.* 39, 257–266.
- Dische, Z., 1962. Color reactions of carbohydrates. In: Whistler, R.L., Wolfrom, M.L. (Eds.), *Methods in Carbohydrate Chemistry*, vol. 1. Academic Press, New York, pp. 477–512.
- Duarte, I., Barros, A., Delgado, I., Almeida, C., Gil, A., 2002. Application of FTIR spectroscopy for the quantification of sugars in mango juice as a function of ripening. *J. Agric. Food Chem.* 50, 3104–3111.
- Fry, S.C., 1988. *The Growing Plant Cell Wall: Chemical and Metabolic Analysis*. Longman, London, pp. 95–97.
- Fu, C., Mielenz, J.R., Xiao, X., Ge, Y., Hamilton, C.Y., Rodriguez Jr., M., Chen, F., Foston, M., Ragauskas, A., Bouton, J., Dixon, R.A., Wang, Z., 2011. Genetic manipulation of lignin reduces recalcitrance and improves ethanol production from switchgrass. *Proc. Natl. Acad. Sci.* 108, 3803–3808.
- Giese, A., French, C., 1955. The analysis of overlapping spectral absorption bands by derivative spectrophotometry. *Appl. Spectrosc.* 9, 78–96.
- Huang, J., Xia, T., Li, A., Yu, B., Li, Q., Tu, Y., Zhang, W., Yi, Z., Peng, L., 2012. A rapid and consistent near infrared spectroscopic assay for biomass enzymatic digestibility upon various physical and chemical pretreatments in *Miscanthus*. *Bioresour. Technol.* 121, 274–281.
- Jia, J., Yu, B., Wu, L., Wang, H., Wu, Z., Li, M., Huang, P., Feng, S., Chen, P., Zheng, Y., Peng, L., 2014. Biomass enzymatic saccharification is determined by the non-KOH-extractable wall polymer features that predominately affect cellulose crystallinity in corn. *PLoS ONE* 9, e108449.
- Li, S., Chan-Halbrendt, C., 2009. Ethanol production in (the) People's Republic of China: potential and technologies. *Appl. Energy* 86, 162–169.
- Li, F., Ren, S., Zhang, W., Xu, Z., Xie, G., Chen, Y., Tu, Y., Li, Q., Zhou, S., Li, Y., Tu, F., Liu, L., Wang, Y., Jiang, J., Qin, J., Li, S., Li, Q., Jing, H., Zhou, F., Guttererson, N., Peng, L., 2013. Arabinose substitution degree in xylan positively affects lignocellulose enzymatic digestibility after various NaOH/H₂SO₄ pretreatments in *Miscanthus*. *Bioresour. Technol.* 130, 629–637.
- Li, M., Feng, S., Wu, L., Li, Y., Fan, C., Zhang, R., Zou, W., Tu, Y., Jing, H., Li, S., Peng, L., 2014a. Sugar-rich sweet sorghum is distinctively affected by wall polymer features for biomass digestibility and ethanol fermentation in bagasse. *Bioresour. Technol.* 167, 14–23.
- Li, M., Si, S., Hao, B., Zha, Y., Wan, C., Hong, S., Kang, Y., Jia, J., Zhang, J., Li, M., Zhao, C., Tu, Y., Zhou, S., Peng, L., 2014b. Mild alkali-pretreatment effectively extracts guaiacyl-rich lignin for high lignocellulose digestibility coupled with largely diminishing yeast fermentation inhibitors in *Miscanthus*. *Bioresour. Technol.* 169, 447–454.
- Li, Z., Zhao, C., Zha, Y., Wan, C., Si, S., Liu, F., Zhang, R., Li, F., Yu, B., Yi, Z., Xu, N., Peng, L., Li, Q., 2014c. The minor wall-networks between monolignols and interlinked-phenolics predominantly affect biomass enzymatic digestibility in *Miscanthus*. *PLoS ONE* 9, e105115.
- Martin, A., Palmer, W., Byrt, C., Furbank, R., Grof, C., 2013. A holistic high-throughput screening framework for biofuel feedstock assessment that characterises variations in soluble sugars and cell wall composition in *Sorghum bicolor*. *Biotechnol. Biofuels* 6, 186.
- Peng, L., Hocart, C., Redmond, J., Williamson, R., 2000. Fractionation of carbohydrates in *Arabidopsis* root cell walls shows that three radial swelling loci are specifically involved in cellulose production. *Planta* 211, 406–414.
- Roberts, C., Houx, J., Fritschi, F., 2011. Near-infrared analysis of sweet sorghum bagasse. *Crop Sci.* 51, 2284.
- Rodriguez-Saona, L., Fry, F., McLaughlin, M., Calvey, E., 2001. Rapid analysis of sugars in fruit juices by FT-NIR spectroscopy. *Carbohydr. Res.* 336, 63–74.
- Shenk, J.S., Westerhaus, M.O., 1991. Population structuring of near infrared spectra and modified partial least squares regression. *Crop Sci.* 31, 1548–1555.
- Shenk, J.S., Westerhaus, M.O., 1995. The application of near infrared reflectance spectroscopy (NIRS) to forage analysis. In: Fahey, G.C., Collins, M., Mertens, D.R., Moser, L.E. (Eds.), *Forage Quality, Evaluation, and Utilization*. American Society of Agronomy, Madison, WI, USA, pp. 406–449.
- Sluiter, A., Hames, B., Ruiz, R., Scarlata, C., Sluiter, J., Templeton, D., Crocker, D., 2008. Determination of Structural Carbohydrates and Lignin in Biomass. *Tech. Rep. NREL/TP-510-42618*, NREL, Golden Co.
- Williams, P.C., Sobering, D.C., 1996. How do we do it: a brief summary of the methods we use in developing near infrared calibration. In: Davies, A.M.C., Williams, P. (Eds.), *Near Infrared Spectroscopy: The Future Waves*. NIR Publications, Chichester, UK, pp. 185–188.
- Wu, Z., Zhang, M., Wang, L., Tu, Y., Zhang, J., Xie, G., Zou, W., Li, F., Guo, K., Li, Q., Gao, C., Peng, L., 2013. Biomass digestibility is predominantly affected by three factors of wall polymer features distinctive in wheat accessions and rice mutants. *Biotechnol. Biofuels* 6, 183.
- Xie, G., Peng, L., 2011. Genetic engineering of energy crops: a strategy for biofuel production in China. *J. Integr. Plant Biol.* 53, 143–150.
- Xie, T., Su, P., 2012. Canopy and leaf photosynthetic characteristics and water use efficiency of sweet sorghum under drought stress. *Russ. J. Plant Physiol.* 59, 224–234.
- Xu, N., Zhang, W., Ren, S., Liu, F., Zhao, C., Liao, H., Xu, Z., Huang, J., Li, Q., Tu, Y., Yu, B., Wang, Y., Jiang, J., Qin, J., Peng, L., 2012. Hemicelluloses negatively affect lignocellulose crystallinity for high biomass digestibility under NaOH and H₂SO₄ pretreatments in *Miscanthus*. *Biotechnol. Biofuels* 5, 58.
- Zegada-Lizarazu, W., Monti, A., 2012. Are we ready to cultivate sweet sorghum as a bioenergy feedstock? A review on field management practices. *Biomass Bioenergy* 40, 1–12.
- Zhang, W., Yi, Z., Huang, J., Li, F., Hao, B., Li, M., Hong, S., Lv, Y., Sun, W., Ragauskas, A., Hu, F., Peng, J., Peng, L., 2013. Three lignocellulose features that distinctively affect biomass enzymatic digestibility under NaOH and H₂SO₄ pretreatments in *Miscanthus*. *Bioresour. Technol.* 130, 30–37.