Wageningen Academic Publishers

Application of NIR transmission spectroscopy with effective wavelength selection in non-destructive determination of essential amino acid content of foxtail millet

J. Chen[1,2], X. Ren[1,2] and Q. Shen[1,2*]

[1]College of Food Science and Nutritional Engineering, China Agricultural University, 17 Tsinghua East Road, Beijing 100083, China P.R.; [2]National Engineering Research Center for Fruits and Vegetables Processing, 17 Tsinghua East Road, Beijing 100083, China P.R.; shenqun@cau.edu.cn

Received: 14 February 2016 / Accepted: 19 May 2016

RESEARCH ARTICLE

Abstract

The near infrared (NIR) spectroscopy models developed for rapid and accurate determination of nine essential amino acids and protein in foxtail millet were investigated. The effects of the status of the samples (intact or ground), amino acid/protein correlation and variable selection methods on the predictive ability of the models were analysed. According to results, although the average spectral patterns of the intact and ground samples were similar, the absolute values of peaks and valleys for the intact samples were slightly higher. The modelling results of intact samples were similar to those obtained with ground samples. In the results of the partial least squares models with full spectra, 81-94% of the amino acid variance could be explained, except for lysine, tryptophan and methionine with the coefficients of determination of cross-validation less than 0.70. The correlation between various amino acids and protein might affect the predictive ability of NIR spectroscopy for essential amino acids. After variable selection, the models for isoleucine, leucine, phenylalanine, threonine, tryptophan and valine were remarkably improved by the successive projections algorithm method. The results indicated that short-wave NIR spectroscopy coupled with variable selection is a promising approach for determination of essential amino acids in foxtail millet.

Keywords: near infrared transmission spectroscopy, successive projections algorithm, interval partial least squares, jack-knife algorithm

1. Introduction

Foxtail millet (*Setaria italica* Beauv.), also known as Italian millet, originates from China. As an important crop species, it has a high yield and can adapt better to arid and barren lands than other crops (Li and Wu, 1996). The mean protein content of foxtail millet varieties in China is 11.73±1.40%, which is similar to maize, rice and some other grains (Gu and Liu, 1989). Nowadays, foxtail millet is not only a main source of energy and protein for millions of the low-income population in Asia and Africa, but also an alternative protein source for people who are sensitive to the common protein sources, such as soybean, cow's milk, egg and peanuts (Kasaoka *et al.*, 1999).

Amino acid contents and composition play important roles in the nutritional and functional values of foxtail millet. Foxtail millet contains nine essential amino acids including histidine (His), which is an essential amino acid for infants. In addition, foxtail millet is rich in branched-chain amino acids (valine (Val), leucine (Leu) and isoleucine (Ile); BCAAs), which helps to promote protein synthesis and inhibit protein breakdown (Matsumoto *et al.*, 2009). A 48-month study demonstrated that the combination treatment with BCAAs and angiotensin-I-converting enzyme-inhibitory activity markedly improved the progression of liver fibrosis (Yoshiji *et al.*, 2012). All these mentioned above indicate that the study of the essential amino acids in foxtail millet is of great value.

However, wet chemical analysis of protein and amino acids is not only quite complicated and labour intensive, but also destructive and time-consuming. The near infrared (NIR) spectroscopy analysis can overcome these shortcomings

in the traditional methods. It is applied on the basis of the absorption of molecular overtone and combination vibrations of hydrogenous groups X-H (X=C, N, O ...) (800-2,500 nm). Consequently, the effect of the functional group is the most dominant in the NIR spectra (Zou *et al.*, 2010). Combined with the chemometrics methods, the models for various compositions and properties are developed. Currently, NIR spectroscopy is applied widely, and the standards of NIR detection have been established for some nutritional components of cereals including moisture (ICC Recommendation no. 202 (ICC, 1986)), protein (ICC Standard No. 159 (ICC, 1995), AACC Method 39-25 (AACCI, 2011a)), fat (GB/T24902-2010), ash (AACC Method 08-21 (AACCI, 2011b)), starch, crude fibre and amino acids (GB/T 18868-2002).[1] Although the applications of NIR spectroscopy on rice, rapeseed, wheat, maize, soybeans and barley (Chen *et al.*, 2011; Fontaine *et al.*, 2002; Wu *et al.*, 2002; Zhou *et al.*, 2012) are abundant, only few studies of NIR spectroscopy on foxtail millet were found.

We have determined the protein, carbohydrates and fat contents in foxtail millet by NIR spectroscopy successfully (Chen *et al.*, 2013). In this study, we attempted to expand our work to the rapid and precise, non-destructive and economical determination of essential amino acid contents in foxtail millet. Although a recentstudy on this topic was published (Yang *et al.*, 2013), the samples were ground and dried before scanned, which was destructive and time-consuming and might affect the moisture and amino acid contents. Furthermore, a multi-purpose analyser Fourier transform near infrared reflectance (FT-NIR) spectrometer was adopted to obtain the spectra in that study. Among the nine essential amino acids, methionine (Met) and tryptophan (Trp) were not analysed. The ratio of standard error of prediction to standard deviation (RPD) of lysine (Lys) and Val were unusable for sample screening (RPD=1.61 and 1.86, respectively), and the RPD of threonine (Thr) and His were unusable for most applications (RPD=2.4 and 2.06, respectively). Therefore, these models for essential amino acids need to be improved. On the one hand, according to Kovalenko *et al.* (2006), the NIR analyser with short-wave range of 850-1,048 nm demonstrated a significant advantage. The short-wave infrared spectra (800-1,100 nm) are the vital region to feed and grain analysis. Although the Fourier technique has much better SNR (signal-to-noise ratio) than the grating technique in the mid-infrared application, the grating technique is better than the Fourier technique in the NIR domain. This is because the moving mirror of Fourier technique will produce the mechanical frame error, and as shorter wavelengths (800-1,100 nm) were adopted to build the prediction models, the mechanical frame error has a greater impact on SNR. On the other hand, the models for amino acids might be improved using the effective wavelengths (EWs) selected by the methods such as the competitive adaptive reweighted sampling, genetic algorithms, and successive projection algorithm (SPA) methods (Liu *et al.*, 2010; Wei *et al.*, 2014). In addition, the FT-NIR instrument of the monochromator mainly uses a Michelson interferometer, which has a complex structure and is very expensive. Instead, the short-wave NIR transmission spectroscopy is more economic for wide use.

Therefore, in this study, we aimed to evaluate the feasibility of using the short-wave NIR transmittance spectra for non-destructive determination of the contents of nine essential amino acids including His, Ile, Leu, Lys, Met, phenylalanine (Phe), Thr, Trp and Val in foxtail millet. The specific objectives of this research were (1) to establish the quantitative models for essential amino acid contents using the partial least squares (PLS) methods with full spectra; (2) to compare the models built with intact and ground samples; (3) to analyse the influence of amino acid/protein correlation on the predictive ability of the models; (4) to establish the models for essential amino acid contents using the EWs selected by SPA, the interval partial least squares (iPLS) and the jack-knife algorithm (JK) methods to compare their predictive abilities with the full spectra models.

2. Materials and methods

Sample preparation

In total, 195 varieties of foxtail millet from 16 provinces (Beijing, Gansu, Guangxi, Hainan, Hebei, Henan, Inner Mongolia, Jiangsu, Jilin, Liaoning, Ningxia, Qinghai, Shandong, Shanxi, Shaanxi and Sichuan) of China were used in this study provided by the Institute of Crop Science, Chinese Academy of Agricultural Sciences (Beijing, China P.R.). The foxtail millet samples were decorticated by a rice huller (SY88-TH, Ssangyong Ltd, Incheon, Korea 3 times and polished twice (15 s for each time) by a rice polisher (SY2001-NSART100, Ssangyong Ltd,). After decorticating and polishing, the samples were stored at 4 °C in sealed bags until analysis.

Near infrared spectra acquisition

Before spectra acquisition, the samples were kept at room temperature (20±2 °C) for 24 h to balance the moisture and the temperature. The samples were scanned in transmission mode in the NIR region of 850-1,048 nm with a 2 nm increment, using a NIR spectrometer (Infraneo Junior, Chopin, France). First, the intact foxtail millet samples were scanned, and then ground into whole foxtail millet flour by a grinder (BRC20K Mixer Grinder, Chopin, France). The

[1] GB/T24902-2010 and GB/T 18868-2002 are Chinese standard methods and can be found at: http://down.foodmate.net/standard/sort/3/22627.html and http://down.foodmate.net/standard/sort/3/16278.html, respectively (in Chinese).

process of grinding lasted 10 s and then paused 10 s in a total of 3 times in order to prevent overheating and moisture loss. After grinding, the samples were also scanned. Each sample was scanned in duplicate. The average spectrum of each sample was used in subsequent data analysis.

Reference methods

After spectra acquisition, the contents of moisture (dry matter), protein and nine essential amino acids including His, Ile, Leu, Lys, Met, Phe, Thr, Trp and Val were determined by reference methods. The compositions were measured using official methods of the American Association of Analytical Chemists (AOAC, 2016): moisture (AOAC-930.15, 1990), protein (AOAC-979.09, 1990), eight amino acids (AOAC-994.12, 2000) and Trp (AOAC 988.15, 1990). All the analyses were carried out in duplicate and the results were averaged and calculated on a dry basis.

Spectra pre-treatment and outlier elimination

In this study, multiplicative scatter correction (MSC) and Savitzkye-Golay second derivative with a 2nd-order polynomial and a 9-point window were applied (Rinnan *et al.*, 2009). The MSC method was used for scatter correction to reduce the physical variability of particle size during data collection (Isaksson and Næs, 1988). And Savitzky-Golay derivative was performed to adjust baseline shift, combined with a smoothing step. These spectra pre-treatment algorithms were performed in the Unscrambler software (version 9.7, Camo Software AS, Oslo, Norway).

Before sample selection, the PLS regression (leave-one-out cross-NIR validation) was first carried out on full spectra to identify outliers, through the leverage and the studentised residuals. On the one hand, the leverage could represent the influence of every sample on the model, and the outliers were detected with a threshold equal to 3 times the average leverage, where the average leverage was calculated as: H=(1+LVs)/n (n is the number of samples and LVs is the number of principal components) (Cao, 2013; Faber, 1999). On the other hand, samples with the student residual ≥2.5 were considered as chemical outliers with a confidence level of 95% (De Oliveira *et al.*, 2014).

Sample subsets partitioning

In this study, representative samples were obtained by sample partitioning. Accordingly, similar calibration samples and redundant calculations might be avoided and the models would be more robust. Samples were divided into calibration and validation subsets at the ratio of 3:1 for multivariate modelling using the joint x-y distance (SPXY) algorithm. The models were constructed by full cross-validation. According to the differences in both X (spectra) and Y (reference values) domains, a stepwise procedure to select samples was adopted. The SPXY technique is an advantageous alternative to Kennard-Stone and random sampling methods (Galvão *et al.*, 2005).

Variable selection

The multivariate calibration with three variable selection methods was evaluated: (1) the SPA method with the multiple linear regression (MLR) models was implemented; (2) the iPLS algorithm with the PLS regression was performed; (3) the JK method with the PLS regression was applied. The SPA is a forward selection method, and it can solve the collinearity problem by selecting the wavelengths whose information content is minimally redundant (Araújo *et al.*, 2001). The iPLS method divides the spectra data into several intervals and selects a subset of variables. It compares all the local models with the full spectra models on the basis of the lowest root mean square error of cross-validation (RMSECV), and the iPLS method could give superior performance (Nørgaard *et al.*, 2000). Through jack-knife estimation of the uncertainty of the model parameters, variables with significant regression coefficients are obtained. As a result, the useless or unreliable X- or Y-variables can be eliminated automatically, in order to simplify the final model and making it more reliable (Anderssen *et al.*, 2006). The variable selection procedure was performed in the Matlab environment (version R2013b, the Mathworks, Natick, MA, USA) with the SPA routine (revision 1.0) and PLS Toolbox (version 7.5.2, Eigenvector Research, Inc., Manson, WA, USA) and in the Unscrambler (version 9.7, CAMO Software AS), respectively.

Multivariate calibration and validation

The PLS and the MLR regressions were used to develop all the prediction models. The PLS algorithm compresses the original matrix into principal components (PCs), which maximises the covariance between the reference values and the spectra. The PLS models were established using full cross-validation. The MLR is a common algorithm for calibration with the advantages of being simple and easily interpreted, and the influence of collinearity on the MLR models can be eliminated effectively by the SPA method (Huang *et al.*, 2011). After the cross-validation, the external validation (R^2_{val}) was also carried out. Both methods were applied to avoid overfitting and to ensure the predictive ability of the calibration model. The models were built with PLS Toolbox.

Statistical analysis

The statistical analysis of chemical values for calibration and validation sets was carried out in SPSS (Chicago, IL, USA). The coefficients of determination of cross-validation (R^2_{cv}) and R^2_{val} are used to determine the amount of variation in the data which is modelled by the calibration equation

(Burns and Ciurczak, 2007). The RMSECV and root mean square error of prediction (RMSEP) are the most efficient and simplest indicators of the uncertainty of future predictions. Besides, the RPD can be used to evaluate the precision of the models (Williams and Sobering, 1996) and the bias is a good similarity index between validation samples and the calibration set which determines the accuracy.

3. Results and discussion

Variability of composition in foxtail millets

After some outliers were eliminated for each constituent according to the studentised residuals and the leverage with a 95% confidence limit, samples were divided into calibration and validation sets using the SPXY algorithm. The descriptive statistics for different datasets are presented in Tables 1 and 2. According to the coefficient of variation (CV), the results reflected a broad range of variability in each chemical content, especially in the calibration sets. Besides, the means, ranges and standard deviations between the calibration and validation sets were close and the validation sets were within the limits of the calibration sets. These observations suggested that the calibration and validation sets were selected properly and both sets can represent the variability in the foxtail millet samples (Zhang *et al.*, 2011).

Through the PCs plot, one can find out the relationships between different variables, and interpret sample groupings, similarities or differences. The first factors explain more variation of the data than other components. The closer the samples are in the score plot, the more similar they are with respect to the two components concerned (Jolliffe, 2002). According to the PC1-PC2 score plot from the raw spectra and the pre-processed spectra (Figure 1), the first factor explained 61.39 and 46.52% of X-variation, while the second factor explained additional 21.32 and 28.70% of X-variation. Besides, no obvious clustering tendency was found, although it was more compact between ground samples. Consequently, all foxtail millet samples could be used to build a unified regression model (Zhang *et al.*, 2004).

Spectra overview

The average raw and pre-processed spectra for intact and ground foxtail millet are shown in Figure 2A and 2B, respectively. The average spectral patterns of the intact samples resembled those shown in the ground samples. However, in both raw and pre-processed spectra, the absolute values of peaks and valleys for intact samples at 914, 950 and 992 nm were slightly higher than those for ground samples. Similar observation was reported by Guy *et al.* (2011). The MSC and second-order derivative pre-processed spectra in Figure 2B showed that the overlapping peaks and baseline effects were removed. The main absorption features such as peaks and valleys were sharper. Similar patterns were found in the average standard normal variate second derivative spectra of wheat as well (Pojic *et al.*, 2012). The largest spectra variations were observed in the region around 914 nm assigned to the third overtone of C-H stretching for methyl (Lu *et al.*, 2000), around 950 nm corresponding to O-H 2nd overtone for water, and around 994 nm corresponding to N-H 2nd overtone for peptides and proteins (Pojic *et al.*, 2012).

Table 1. Laboratory reference value statistics of intact foxtail millet for the calibration set and the validation set.[1]

Comp[2]	Calibration set						Validation set					
	n	mean	SD	min	max	CV	n	mean	SD	min	max	CV
His	143	0.27	0.03	0.20	0.36	11.68	47	0.26	0.02	0.21	0.31	8.53
Ile	145	0.48	0.06	0.34	0.65	12.95	48	0.46	0.05	0.37	0.61	11.36
Leu	143	1.79	0.27	1.14	2.53	15.05	48	1.72	0.26	1.22	2.28	14.84
Lys	145	0.25	0.02	0.20	0.30	8.30	47	0.24	0.02	0.22	0.28	6.23
Met	145	0.39	0.05	0.28	0.49	11.84	48	0.38	0.03	0.31	0.44	9.05
Phe	142	0.71	0.09	0.51	0.97	12.70	47	0.70	0.08	0.55	0.86	11.29
Thr	141	0.48	0.06	0.34	0.63	11.93	47	0.47	0.05	0.37	0.56	10.85
Trp	142	0.20	0.03	0.13	0.28	14.16	47	0.19	0.03	0.13	0.26	13.85
Val	144	0.62	0.08	0.43	0.84	12.85	48	0.61	0.07	0.49	0.79	12.25
Prot	143	12.19	1.46	8.90	15.84	12.01	47	12.16	1.38	8.98	15.01	11.36

[1] SD = standard deviation (%); CV = coefficient of variation (%).

[2] The composition (%) modelled includes histidine (His), isoleucine (Ile), leucine (Leu), lysine (Lys), methionine (Met), phenylalanine (Phe), threonine (Thr), tryptophan (Trp) and valine (Val) and protein (Prot).

Table 2. Laboratory reference value statistics of ground foxtail millet for the calibration set and the validation set.[1,2]

Comp	Calibration set						Validation set					
	n	mean	SD	min	max	CV	n	mean	SD	min	max	CV
His	145	0.27	0.03	0.20	0.36	11.85	48	0.26	0.03	0.21	0.31	10.34
Ile	146	0.48	0.06	0.34	0.65	12.89	48	0.46	0.05	0.36	0.58	11.40
Leu	145	1.79	0.28	1.14	2.53	15.39	48	1.72	0.23	1.22	2.14	13.57
Lys	145	0.25	0.02	0.20	0.30	8.49	48	0.24	0.02	0.22	0.28	6.91
Met	143	0.39	0.05	0.29	0.49	11.52	48	0.38	0.04	0.31	0.44	9.27
Phe	143	0.72	0.09	0.52	0.97	12.25	48	0.69	0.09	0.51	0.87	12.40
Thr	141	0.49	0.06	0.34	0.63	11.84	47	0.46	0.05	0.36	0.56	10.36
Trp	143	0.20	0.03	0.13	0.26	14.07	47	0.20	0.02	0.14	0.23	10.93
Val	146	0.63	0.08	0.43	0.84	12.58	49	0.59	0.07	0.44	0.73	12.55
Prot	145	12.31	1.46	8.90	15.84	11.84	48	11.86	1.37	8.98	14.37	11.58

[1] SD = standard deviation (%); CV = coefficient of variation (%).

[2] The composition (%) modelled includes histidine (His), isoleucine (Ile), leucine (Leu), lysine (Lys), methionine (Met), phenylalanine (Phe), threonine (Thr), tryptophan (Trp) and valine (Val) and protein (Prot).

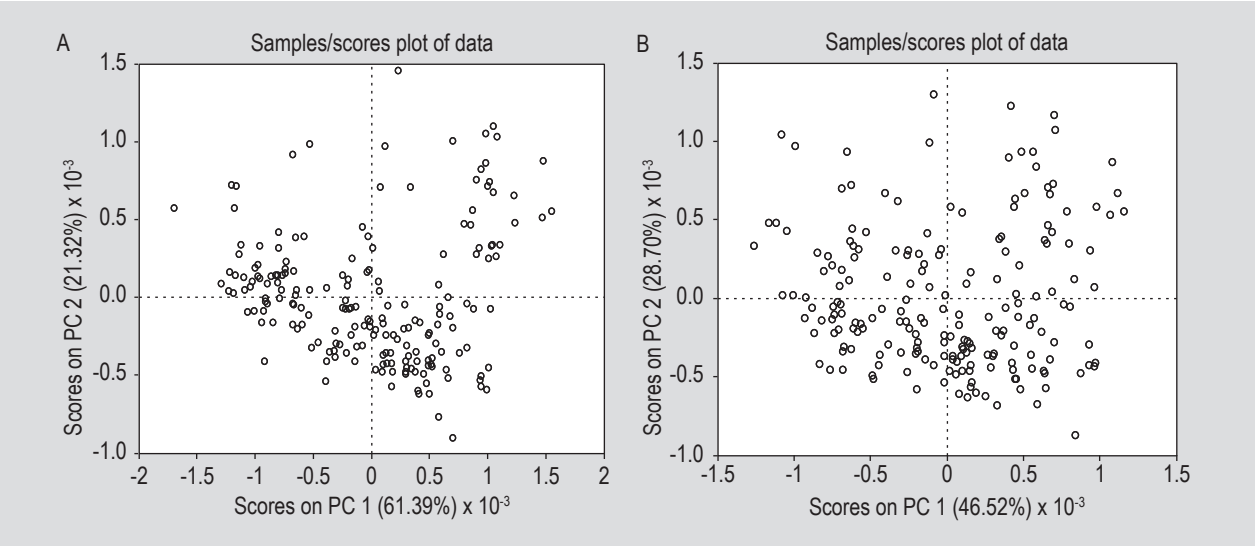

Figure 1. Score plots of the first two principal components (PC1 and PC2) of (A) intact and (B) ground foxtail millet samples.

Models using full spectra

The prediction results of essential amino acids and protein contents in intact and ground samples using full spectra are illustrated in Tables 3 and 4. In the results of full PLS (the PLS method with full spectra), the ranges of R^2_{cv} and R^2_{val} values with full spectra were 0.15-0.94 and 0.11-0.95 for intact samples; for ground samples the ranges were 0.12-0.95 and 0.16-0.94. The R^2_{cv} and R^2_{val} for Lys, Trp and Met were less than 0.70, and the R^2_{cv} for the models of other amino acids (His, Ile, Leu, Phe, Thr, Val) and protein were more than 0.80. It indicated that more than 80% of the variance in these amino acids concentration could be explained by the spectra variables. The prediction models obtained with intact samples were similar to those obtained with ground samples. However, Wu *et al.* (2002) and Fontaine *et al.* (2001) reported that the calibration models based on ground rice and soybean samples were more accurate than those based on intact samples. Following reasons may contribute to the results mentioned above. Firstly, the grain size of foxtail millet is relatively small; secondly, the protein and amino acids uniformly distribute in foxtail millet; lastly, the mass of the samples varies in different studies (Pazdernik *et al.*, 1997). According to Zhang *et al.* (2011), the models with large samples (3 g) outperformed those with small samples (500 mg), which might increase the diversity of the surface area. In this research, large mass of the intact samples (40-45 g) improved the models. Above

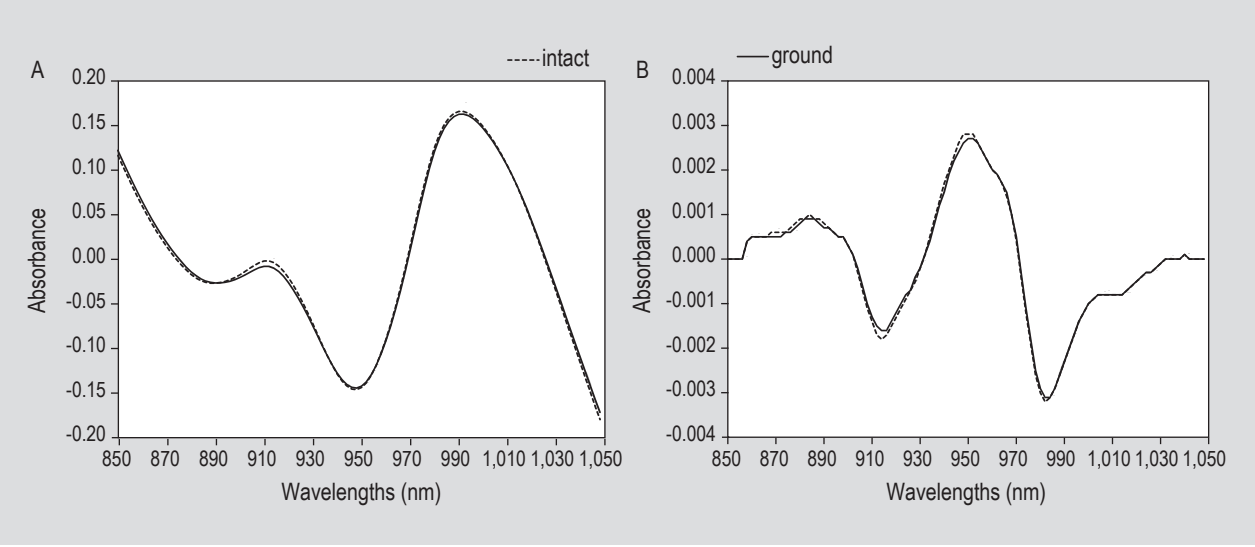

Figure 2. (A) Average raw spectra, and (B) average pre-processed spectra for intact and ground foxtail millet samples.

Table 3. Statistics of the models for essential amino acids and protein contents prediction in intact foxtail millet using full spectra.

Comp[1]	PCs[2]	Calibration[3]		Validation[4]				Correlation coefficients[5]	
		R^2_{cv}	RMSECV	R^2_{val}	RMSEP	Bias	RPD	r_{spc}	r_{prot}
His	4	0.88	0.011	0.83	0.009	0.000	2.4	0.94	0.94
Ile	5	0.84	0.025	0.87	0.019	-0.001	2.8	0.91	0.92
Leu	4	0.94	0.067	0.95	0.056	0.002	4.6	0.97	0.97
Lys	6	0.15	0.019	0.11	0.015	0.004	1.0	0.36	0.22
Met	2	0.68	0.026	0.7	0.019	0.001	1.8	0.82	0.84
Phe	4	0.91	0.027	0.91	0.024	0.002	3.3	0.95	0.96
Thr	3	0.89	0.019	0.90	0.016	0.000	3.2	0.94	0.95
Trp	2	0.44	0.021	0.68	0.016	-0.002	1.7	0.69	0.72
Val	5	0.81	0.035	0.81	0.032	-0.005	2.3	0.90	0.91
Prot	5	0.97	0.254	0.96	0.279	-0.026	5.0	0.98	–

[1] The composition modelled include histidine (His), isoleucine (Ile), leucine (Leu), lysine (Lys), methionine (Met), phenylalanine (Phe), threonine (Thr), tryptophan (Trp) and valine (Val) and protein (Prot).

[2] PCs = number of principal components.

[3] The calibration set includes two parameters: R^2_{cv}, the coefficients of determination of cross-validation; RMSECV, the root of mean square error of cross-validation.

[4] The validation set includes four parameters: R^2_{val}, the coefficients of determination of external validation; RMSEP, the root of mean square error of prediction; RPD, the ratio of standard error of prediction to standard deviation.

[5] Correlation coefficients: the r_{spc} means the correlation coefficients of the cross-validation model using the full spectra; the r_{prot} means the correlation coefficients between the protein contents and each variety of amino acids.

all, satisfactory results could be obtained for determination of essential amino acids and protein contents of the intact foxtail millet, so that the pre-treatment was simplified and the water loss was minimised.

Considering the results of both intact and ground foxtail millet samples were consistent, it is unnecessary to discuss the results for intact and ground samples separately. On the basis of guidelines for interpretation of RPD outlined by Williams and Sobering (1996), the models of Val, His and Ile (RPD=2.3, 2.4 and 2.8) were usable for sample screening; the models of Leu, Phe, Thr and protein (RPD=4.6, 3.3, 3.2 and 5.0) were usable with caution for most applications. Compared with another research about prediction of amino acids in foxtail millet (Yang *et al.*, 2013), the models for Leu, Thr, Phe and Val were improved in our study, with the

Table 4. Statistics of the models for essential amino acids and protein contents prediction in ground foxtail millet using full spectra.

Comp[1]	PCs[2]	Calibration[3]		Validation[4]			
		R^2_{cv}	RMSECV	R^2_{val}	RMSEP	Bias	RPD
His	4	0.86	0.012	0.88	0.009	0.000	3.0
Ile	5	0.83	0.026	0.84	0.021	0.002	2.5
Leu	6	0.95	0.059	0.94	0.060	-0.001	3.9
Lys	3	0.12	0.020	0.19	0.016	0.006	1.0
Met	2	0.73	0.023	0.70	0.020	0.002	1.8
Phe	4	0.89	0.029	0.92	0.025	-0.006	3.4
Thr	5	0.93	0.016	0.91	0.015	-0.004	3.2
Trp	3	0.47	0.020	0.65	0.013	-0.003	1.6
Val	5	0.76	0.039	0.83	0.031	-0.003	2.4
Prot	6	0.97	0.254	0.96	0.273	0.014	5.0

[1] The composition modelled include histidine (His), isoleucine (Ile), leucine (Leu), lysine (Lys), methionine (Met), phenylalanine (Phe), threonine (Thr), tryptophan (Trp) and valine (Val) and protein (Prot).
[2] PCs = number of principal components.
[3] The calibration set includes two parameters: R^2_{cv}, the coefficients of determination of cross-validation; RMSECV, the root of mean square error of cross-validation.
[4] The validation set includes four parameters: R^2_{val}, the coefficients of determination of external validation; RMSEP, the root of mean square error of prediction; RPD, the ratio of standard error of prediction to standard deviation.

RPD value rising by 73.7, 54.4, 15.5 and 24.7%, respectively. Although the RPD of Lys (RPD=1.6) was higher than the value (RPD=1.0) in our study, it was still unusable for sample screening. The RPD of His and Ile were similar to the results in this study. However, the RMSEP of His was twice as large as it in this study, which means that the future prediction of the model in our study is more accurate. Another literature on the amino acid contents in soybeans reported a better result of Lys (RPD=2.89) and inferior results of Ile, Thr and Phe (RPD=2.32, 2.01 and 2.82), which obtained the models by the PLS algorithm with the spectra range of 850-1,048 nm.

We attempted to explain the variation of predictive ability in various amino acids models, through correlating the RPD to the concentrations and the CVs of these amino acids. And there were no significant relationships. However, when the correlation coefficients between the NIR spectra and amino acid contents (r_{spc}) were regressed against the correlation coefficients between protein and amino acid contents (r_{prot}) (Figure 3), the variation in the predictive ability of NIR spectroscopy for essential amino acids was determined by how a certain amino acid was correlated to protein. This implied that NIR spectroscopy predicted amino acid contents in foxtail millet indirectly by deriving it from the total nitrogen contents. If the correlation was poor, as in the cases of Lys and Trp, the prediction by NIR spectroscopy would be correspondingly inaccurate. A similar observation for soybean meal and full-fat soy was made by Kovalenko *et al.* (2006) and Fontaine *et al.* (2001). In their studies, the correlation between Lys and protein in soybean was far closer than that in foxtail millet. The r between Leu and protein in peanuts (r=0.89) (Wang *et al.*, 2012) was lower than that in foxtail millet (r=0.97). Therefore, for prediction of amino acids using NIR spectroscopy in some legumes and cereals, the introduction of calibration samples (possibly different species) with various amino acid profiles might be one way to break the correlation.

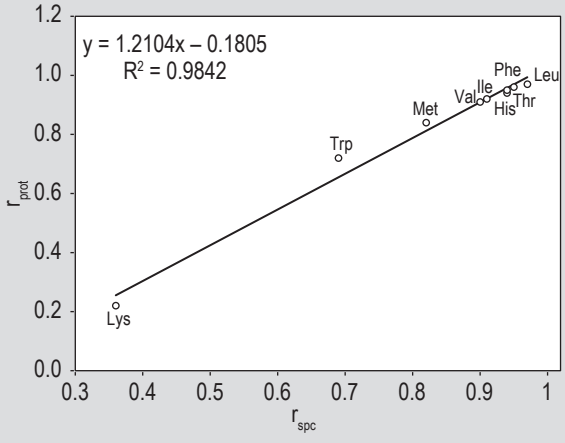

Figure 3. Correlation coefficients between near infrared spectra and amino acids contents (r_{spc}) versus correlation coefficients between protein and amino acids contents (r_{prot}).

Models using selected effective wavelengths

The prediction results of essential amino acids and protein contents using selected wavelengths are illustrated in Table 5. In addition, most of the specific EWs selected by the SPA, iPLS and JK methods for a certain amino acid were not identical, due to the different mechanism of the three methods. The R_{cv}^2 of the model using full spectra for Lys was 0.15. After variable selection, the R_{cv}^2 wasn't increased for the models built with the selected variables by SPA, iPLS and JK algorithms, respectively. That means no significant improvement of the models for Lys can be observed and

Table 5. Statistics of the models for essential amino acids and protein contents prediction in intact foxtail millet using selected wavelengths.

Comp[1]	Methods[2]	n[3]	PCs[4]	Calibration[5]		Validation[6]		
				R^2_{cv}	RMSECV	R^2_{val}	RMSEP	RPD
His	SPA	7	–	0.89	0.011	0.84	0.009	2.4
	iPLS	30	4	0.90	0.010	0.84	0.009	2.4
	JK	55	4	0.89	0.010	0.83	0.009	2.4
Ile	SPA	7	–	0.83	0.025	0.88	0.018	2.9
	iPLS	36	3	0.83	0.030	0.85	0.020	2.6
	JK	51	5	0.83	0.025	0.87	0.019	2.8
Leu	SPA	11	–	0.96	0.054	0.97	0.043	5.9
	iPLS	20	3	0.95	0.060	0.96	0.050	5.1
	JK	71	4	0.94	0.065	0.96	0.051	5.0
Lys	SPA	1	–	0.01	0.021	0.01	0.016	1.0
	iPLS	100	6	0.15	0.019	0.11	0.015	1.0
	JK	16	6	0.09	0.020	0.09	0.015	1.0
Met	SPA	9	–	0.75	0.023	0.73	0.019	1.8
	iPLS	20	2	0.72	0.025	0.68	0.020	1.7
	JK	30	2	0.71	0.025	0.70	0.020	1.7
Phe	SPA	14	–	0.92	0.025	0.93	0.021	3.8
	iPLS	42	4	0.93	0.025	0.93	0.021	3.8
	JK	69	3	0.91	0.027	0.91	0.024	3.3
Thr	SPA	9	–	0.91	0.017	0.92	0.014	3.6
	iPLS	45	3	0.92	0.016	0.92	0.015	3.4
	JK	77	3	0.91	0.018	0.91	0.016	3.2
Trp	SPA	3	–	0.48	0.020	0.74	0.014	1.9
	iPLS	40	2	0.50	0.020	0.73	0.014	1.9
	JK	26	1	0.48	0.020	0.71	0.015	1.8
Val	SPA	10	–	0.83	0.033	0.85	0.029	2.6
	iPLS	60	3	0.80	0.040	0.83	0.030	2.5
	JK	40	5	0.80	0.036	0.78	0.035	2.1
Prot	SPA	7	–	0.97	0.272	0.98	0.220	6.3
	iPLS	46	3	0.97	0.237	0.96	0.257	5.4
	JK	77	5	0.97	0.236	0.96	0.269	5.1

[1] The composition modelled include histidine (His), isoleucine (Ile), leucine (Leu), lysine (Lys), methionine (Met), phenylalanine (Phe), threonine (Thr), tryptophan (Trp) and valine (Val) and protein (Prot).

[2] The variable selection includes three methods: SPA = the successive projections algorithm; iPLS = interval partial least squares; JK = the jack-knife algorithm.

[3] n = number of effective wavelengths.

[4] PCs = number of principal components.

[5] The calibration set includes two parameters, R^2_{cv} = the coefficients of determination of cross-validation; RMSECV = the root of mean square error of cross-validation.

[6] The validation set includes three parameters; R^2_{val} = the coefficients of determination of external validation; RMSEP, the root of mean square error of prediction; RPD = the ratio of standard error of prediction to standard deviation.

these models cannot be used to predict the Lys content accurately. The SPA method improved the models of almost all the essential amino acids, compared with the full PLS models, except for His and Ile. In the iPLS models, the outcomes of Leu, Phe, Thr and Trp were improved in both the cross-validation and the R^2_{val}, compared with the full PLS models. In addition, the effect of the JK method used for variable selection on the models was not significant. In the models for His, Phe, Thr and Trp, the performance of the models with the spectra selected by the SPA and iPLS methods was equally good. In the models for Ile, Leu, Met and Val, the SPA method yielded superior performance, compared to the predictive results of the iPLS models, since the R^2_{cv} and RPD increased and the RMSECV went down. It indicated that the MLR models could be improved based on a small subset of wavelengths selected by the SPA method, minimising the collinearity problem of the MLR. For protein, the RPD and R^2_{val} in the SPA model exceeded them in the iPLS model. The RMSECV was 0.272, which was much higher than it (RMSECV=0.237) in the iPLS model. The model with the EWs selected by the iPLS method was more precise among the protein models. Wei *et al.* (2014) determined the BCAAs (Ile, Leu, Val) contents in fermented *Cordyceps sinensis* mycelium, the EWs for the three amino acids were mainly concentrated around the wavenumbers of 12,010-10,050, 7,500-6,000 and 5,000-4,000 cm^{-1} selected by competitive adaptive reweighted sampling and genetic algorithms. Although the wavelength range was wider than that in this study of foxtail millet, the performance of the models was inferior to our results.

The EWs selected by the SPA, iPLS and JK methods were compared with the wavelengths with large absolute values in the regression coefficient plots of the models with full spectra (Figure 4). In the full PLS models, the regression coefficients were used to calculate the response values from the X variables, and X variables with large absolute values play an important role in the regression models. Most of the curves for amino acids followed a similar pattern with protein, except for Trp and Met with lower RPD. It confirmed that NIR spectroscopy predicted amino acid contents in foxtail millet indirectly by deriving it from

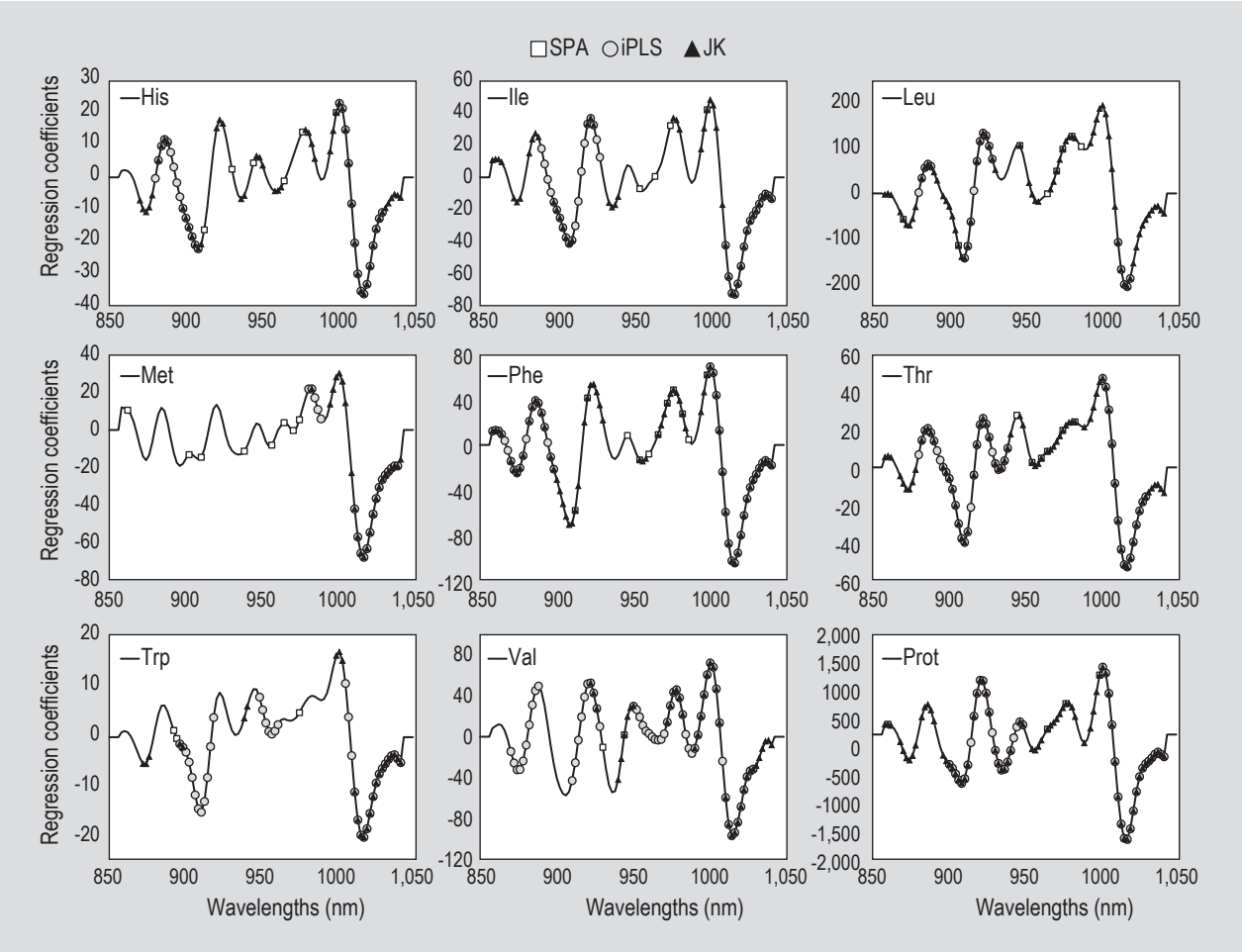

Figure 4. Comparison of effective wavelengths selected by the successive projection algorithm (interval partial least squares), interval partial least squares (the successive projections algorithm), the jack-knife algorithm method and regression coefficients in full partial least squares for eight essential amino acids and protein contents.

the spectral information of the total nitrogen content. Compared to the curve of protein, the valley at 990 nm vanished in Trp and an extra small peak at 965 nm could be observed in Met. Remarkably, the number of selected variables for protein and various amino acids was different. The reason might be that protein and amino acids molecules had specific absorption bands in the spectra regions of NIR. It indicated that variable selection was efficacious and specific for diverse components (Huang *et al.*, 2013).

The EWs selected by the JK method for protein were distributed around each peaks and valleys, and involved 77% wavelengths of the entire spectrum. Moreover, they comprised all the ranges of EWs selected for the essential amino acids. Besides, the range of 990-1,042 nm corresponding to N-H 2nd overtone for proteins was included for all the amino acids. The spectral regions selected by the JK method for Thr and His (except for 860-864 nm), for Leu, Ile and Phe (except for 930-960 nm), and for Met, Trp and Val (except for 850-960 nm) were consistent with the regions for protein prediction. Just like the JK method, the region of 1,002-1,042 nm selected by the iPLS algorithm for all amino acids were also involved, which was assigned to the N-H 2nd overtone and the combination of C-H in methyl and methylene. The range of 912-930 nm assigned to the third overtone of C-H in methyl was selected for Ile, Leu and Val, and the possible reason was that all of them contained two methyl (Lu *et al.*, 2000). The number of EWs was the least in the SPA models, only 3-11% of the wavelengths were included. Besides, the wavelengths for amino acids fell within the regions selected by the iPLS algorithm, except in the model for Met where there was only one wavelength selected in the range of 990-1,048 nm. However, the SPA model for Met was no better than the iPLS and JK models, which indicated that Met had little relationship with the band of N-H and with the spectra in the range of 850-1,048 nm.

5. Conclusions

NIR spectroscopy was successfully utilised for non-destructive determination of essential amino acids and protein contents in foxtail millet, except for Lys, Trp and Met. The correlation coefficients between various amino acids and protein contents were different and the variation of the correlation coefficients affects the predictive ability of the models for these amino acids. In addition, compared with the models built with the full spectra and with the EWs selected by the JK and iPLS methods, the models built with SPA algorithm were simpler and had better predictive ability, because less subsets of spectra regions were used and the RPD values were higher. In summary, NIR spectroscopy combined with the SPA methods could effectively promote the prediction results for amino acids in foxtail millet.

Acknowledgements

This work has been financially supported by the China Agricultural Research System (No. CARS-07-12.5). We are also grateful to Xin Qiao for providing language help and invaluable technical support. The authors declared that they have no conflicts of interest to this work.

References

American Association of Cereal Chemists International (AACI), 2011a. AACCI Method 39-25.01. Near-infrared reflectance method for protein content in whole-grain wheat. Approved Methods of the American Association of Cereal Chemists, 11th edition. AACC, St Paul, MN, USA. Available at: http://methods.aaccnet.org/summaries/39-25-01.aspx.

American Association of Cereal Chemists International (AACCI), 2011b. AACCI Method 08-21.01. Prediction of ash content in wheat flour – near-infrared method. Approved Methods of the American Association of Cereal Chemists, 11th edition. AACC, St Paul, MN, USA. Available at: http://methods.aaccnet.org/summaries/08-21-01.aspx.

Anderssen, E., Dyrstad, K., Westad, F. and Martens, H., 2006. Reducing over-optimism in variable selection by cross-model validation. Chemometrics and Intelligent Laboratory Systems 84: 69-74.

Araújo, M.C.U., Saldanha, T.C.B., Galvão, R.K.H., Yoneyama, T., Chame, H.C. and Visani, V., 2001. The successive projections algorithm for variable selection in spectroscopic multicomponent analysis. Chemometrics and Intelligent Laboratory Systems 57: 65-73.

Association of Analytical Communities International (AOAC), 2016. Official methods of analysis, 20th edition. AOAC International, St Paul, MN, USA.

Burns, D.A. and Ciurczak, E.W., 2007. Handbook of near-infrared analysis, 35. CRC press, Boca Raton, FL, USA.

Cao, N., 2013. Calibration optimization and efficiency in near infrared spectroscopy. Graduate Theses and Dissertations Thesis, Iowa State University, Ames, IA, USA, 184 pp.

Chen, G.L., Zhang, B., Wu, J.G. and Shi, C.H., 2011. Nondestructive assessment of amino acid composition in rapeseed meal based on intact seeds by near-infrared reflectance spectroscopy. Animal Feed Science and Technology 165: 111-119.

Chen, J., Ren, X., Zhang, Q., Diao, X. and Shen, Q., 2013. Determination of protein, total carbohydrates and crude fat contents of foxtail millet using effective wavelengths in NIR spectroscopy. Journal of Cereal Science 58: 241-247.

De Oliveira, G.A., Bureau, S., Renard, C., Pereira-Netto, A.B. and De Castilhos, F., 2014. Comparison of NIRS approach for prediction of internal quality traits in three fruit species. Food Chemistry 143: 223-230.

Faber, N.K.M., 1999. Estimating the uncertainty in estimates of root mean square error of prediction: application to determining the size of an adequate test set in multivariate calibration. Chemometrics and Intelligent Laboratory Systems 49: 79-89.

Fontaine, J., Horr, J. and Schirmer, B., 2001. Near-infrared reflectance spectroscopy enables the fast and accurate prediction of the essential amino acid contents in soy, rapeseed meal, sunflower meal, peas, fishmeal, meat meal products, and poultry meal. Journal of Agricultural and Food Chemistry 49: 57-66.

Fontaine, J., Schirmer, B. and Horr, J., 2002. Near-infrared reflectance spectroscopy (NIRS) enables the fast and accurate prediction of essential amino acid contents. 2. Results for wheat, barley, corn, triticale, wheat bran/middlings, rice bran, and sorghum. Journal of Agricultural and Food Chemistry 50: 3902-3911.

Galvão, R.K.H., Araujo, M.C.U., José, G.E., Pontes, M.J.C., Silva, E.C. and Saldanha, T.C.B., 2005. A method for calibration and validation subset partitioning. Talanta 67: 736-740.

Gu, S. and Liu, S., 1989. Study on protein content and amino acid composition of Chinese foxtail millet. Acta Agriculturae Boreali-Sinica 4: 8-15.

Guy, F., Prache, S., Thomas, A., Bauchart, D. and Andueza, D., 2011. Prediction of lamb meat fatty acid composition using near-infrared reflectance spectroscopy (NIRS). Food Chemistry 127: 1280-1286.

Huang, L., Wu, D., Jin, H., Zhang, J., He, Y. and Lou, C., 2011. Internal quality determination of fruit with bumpy surface using visible and near infrared spectroscopy and chemometrics: a case study with mulberry fruit. Biosystems Engineering 109: 377-384.

Huang, Z., Sha, S., Rong, Z., Chen, J., He, Q., Khan, D.M. and Zhu, S., 2013. Feasibility study of near infrared spectroscopy with variable selection for non-destructive determination of quality parameters in shell-intact cottonseed. Industrial Crops and Products 43: 654-660.

International Association for Cereal Science and Technology (ICC), 1986. ICC Standard Methods 202. Procedure for near infrared (NIR) reflectance analysis of ground wheat and milled wheat products. Available at: https://www.icc.or.at/standard_methods.

International Association for Cereal Science and Technology (ICC), 1995. ICC Standard Methods 159. Determination of protein by near infrared reflectance (NIR) Spectroscopy. Available at: https://www.icc.or.at/standard_methods.

Isaksson, T. and Næs, T., 1988. The effect of multiplicative scatter correction (MSC) and linearity improvement in NIR spectroscopy. Applied Spectroscopy 42: 1273-1284.

Jolliffe, I.T., 2002. Graphical representation of data using principal components, principal component analysis. Springer, New York, NY, USA, pp. 78-110.

Kasaoka, S., Oh-hashi, A., Morita, T. and Kiriyama, S., 1999. Nutritional characterization of millet protein concentrates produced by a heat-stable α-amylase digestion. Nutrition Research 19: 899-910.

Kovalenko, I.V., Rippke, G.R. and Hurburgh, C.R., 2006. Determination of amino acid composition of soybeans (*Glycine max*) by near-infrared spectroscopy. Journal of Agricultural and Food Chemistry 54: 3485-3491.

Li, Y. and Wu, S., 1996. Traditional maintenance and multiplication of foxtail millet (*Setaria italica* (L.) P. Beauv.) landraces in China. Euphytica 87: 33-38.

Liu, F., Jin, Z.L., Naeem, M.S., Tian, T., Zhang, F., He, Y., Fang, H., Ye, Q.F. and Zhou, W.J., 2010. Applying near-infrared spectroscopy and chemometrics to determine total amino acids in herbicide-stressed oilseed rape leaves. Food and Bioprocess Technology 4: 1314-1321.

Lu, W.Z., Yuan, H.F., Xu, G.T. and Qiang, D.M., 2000. The technology of modern near infrared spectral analysis. China Petrochemical Press, Beijing, China P.R.

Matsumoto, K., Koba, T., Hamada, K., Sakurai, M., Higuchi, T. and Miyata, H., 2009. Branched-chain amino acid supplementation attenuates muscle soreness, muscle damage and inflammation during an intensive training program. Journal of Sports Medicine and Physical Fitness 49: 424-431.

Nørgaard, L., Saudland, A., Wagner, J., Nielsen, J.P., Munck, L. and Engelsen, S., 2000. Interval partial least-squares regression (iPLS): a comparative chemometric study with an example from near-infrared spectroscopy. Applied Spectroscopy 54: 413-419.

Pazdernik, D.L., Killam, A.S. and Orf, J.H., 1997. Analysis of amino and fatty acid composition in soybean seed, using near infrared reflectance spectroscopy. Agronomy Journal 89: 679-685.

Pojic, M., Mastilovic, J. and Majcen, N., 2012. Robustness of the near infrared spectroscopy method determined using univariate and multivariate approach. Food Chemistry 134: 1699-1705.

Rinnan, Å., Berg, F.v.d. and Engelsen, S.B., 2009. Review of the most common pre-processing techniques for near-infrared spectra. Trends in Analytical Chemistry 28: 1201-1222.

Wang, L., Wang, Q., Liu, H., Liu, L. and Du, Y., 2012. Determining the contents of protein and amino acids in peanuts using near-infrared reflectance spectroscopy. Journal of the Science of Food and Agriculture 93: 118-124.

Wei, X., Xu, N., Wu, D. and He, Y., 2014. Determination of branched-amino acid content in fermented *Cordyceps sinensis* mycelium by using FT-NIR spectroscopy technique. Food and Bioprocess Technology 7: 184-190.

Williams, P.C. and Sobering, D.C., 1996. How do we do it: a brief summary of the methods we use in developing near infrared calibrations. In: Davies, A.M.C., Williams, P.C. (eds.) Infrared Spectroscopy: The future waves. NIR Publications, Chichester, UK, pp. 185-188.

Wu, J.G.G., Shi, C.H. and Zhang, X.M., 2002. Estimating the amino acid composition in milled rice by near-infrared reflectance spectroscopy. Field Crops Research 75: 1-7.

Yang, X.S., Wang, L.L., Zhou, X.R., Shuang, S.M., Zhu, Z.H., Li, N., Li, Y., Liu, F., Liu, S.C., Lu, P., Ren, G.X. and Dong, C., 2013. Determination of protein, fat, starch, and amino acids in foxtail millet *Setaria italica* (L.) Beauv. by Fourier transform near-infrared reflectance spectroscopy. Food Science and Biotechnology 22: 1495-1500.

Yoshiji, H., Noguchi, R., Ikenaka, Y., Kaji, K., Aihara, Y., Douhara, A., Yamao, J., Toyohara, M., Mitoro, A. and Sawai, M., 2012. Combination of branched-chain amino acid and angiotensin-converting enzyme inhibitor improves liver fibrosis progression in patients with cirrhosis. Molecular Medicine Reports 5: 539.

Zhang, B., Rong, Z.Q., Shi, Y., Wu, J.G. and Shi, C.H., 2011. Prediction of the amino acid composition in brown rice using different sample status by near-infrared reflectance spectroscopy. Food Chemistry 127: 275-281.

Zhang, M.H., Luypaert, J., Fernández Pierna, J.A., Xu, Q.S. and Massart, D.L., 2004. Determination of total antioxidant capacity in green tea by near-infrared spectroscopy and multivariate calibration. Talanta 62: 25-35.

Zhou, L.J., Zhang, L.Y., Zhang, E.X., Li, J.T., Yang, W.J. and Wang, Z.Y., 2012. Rapid determination of swine available energy and amino acids in corn distillers dried grains with solubles by near-infrared reflectance spectroscopy. Animal Feed Science and Technology 175: 198-202.

Zou, X.B., Zhao, J.W., Povey, M.J.W., Holmes, M. and Mao, H., 2010. Variables selection methods in near-infrared spectroscopy. Analytica Chimica Acta 667: 14-32.