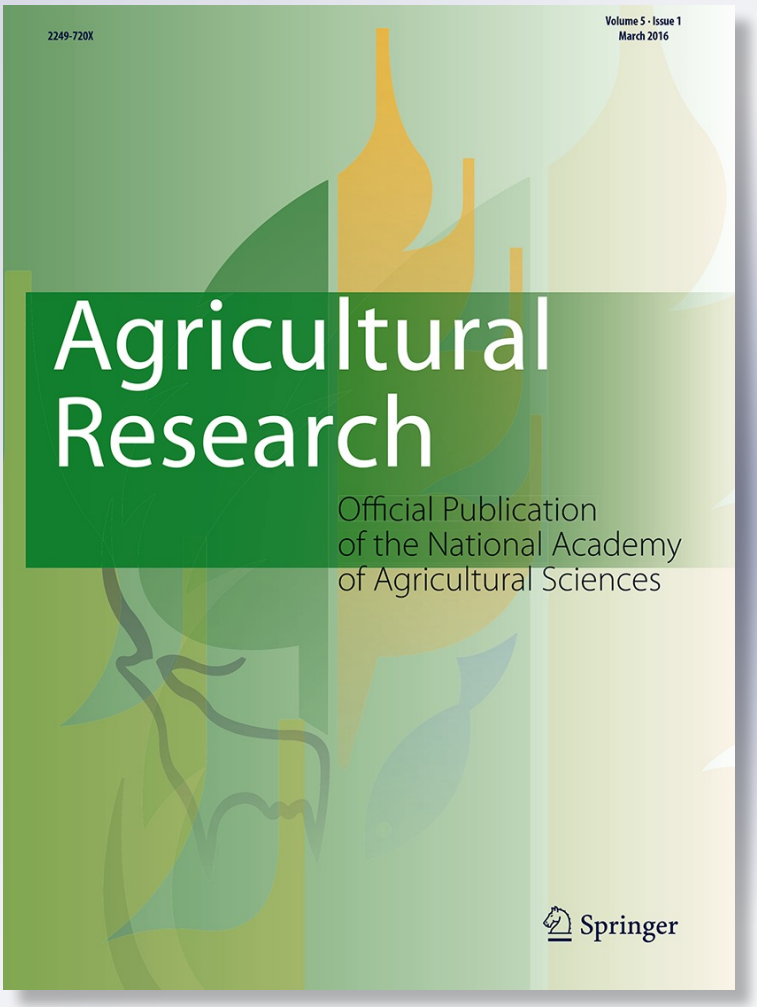
Near-Infrared Reflectance Spectroscopy for Protein Content in Soybean Flour and Screening of Germplasm Across Different Countries

Vaishali Mourya, Vineet Kumar, Anita Rani, Meeta Jain & S. M. Husain

Agricultural Research

ISSN 2249-720X
Volume 5
Number 1

Agric Res (2016) 5:29-34
DOI 10.1007/s40003-015-0196-0



Springer

Agric Res (March 2016) 5(1):29–34
DOI 10.1007/s40003-015-0196-0

FULL-LENGTH RESEARCH ARTICLE

Near-Infrared Reflectance Spectroscopy for Protein Content in Soybean Flour and Screening of Germplasm Across Different Countries

Vaishali Mourya[1] · Vineet Kumar[1] · Anita Rani[1] · Meeta Jain[2] · S. M. Husain[1]

Received: 31 August 2015 / Accepted: 27 November 2015 / Published online: 30 December 2015

Abstract Maintaining high protein content in soybean seeds is critical in view of the premium offered for protein content in soymeal in international market. Besides, quality of several soy-based products depends upon crude protein content of the initial raw material used. Determination of protein content in soybean seeds through wet chemistry involves use of chemicals and is time-consuming and labour-intensive. In the present investigation, a calibration model for determination of protein content in ground soyflour using non-destructive method of near-infrared reflectance spectroscopy (NIRs) was developed. A set of samples with wide variation in protein content as measured through wet chemistry were scanned by NIRs between 1100 and 2100 nm at 2-nm interval. High determination coefficient (R^2) and low values for root-mean-square error as well as standard error of prediction of cross-validation confirmed the utility of model for prediction of future samples. For external validation, a set of genotypes from different genetic background were tested and were predicted with good accuracy. Lower values of standard error of calibration and prediction were observed than reported previously for whole seeds. Subsequently, 1210 soybean accessions from 18 countries were screened through NIRS. Average protein content of soybean accessions from different countries varied significantly. Genotypes identified for high protein content from different countries may be used for development of genotypes containing further high protein content.

Keywords Determination coefficient · Germplasm · Near-infrared reflectance spectroscopy · Protein content · Soybean · Soy-based products

Introduction

Soybean is the numero uno oilseed crop, accounting for approximately 57 % of total oilseed production in the world. After crushing of soybean seeds for extraction of oil, the meal left being enriched in protein is valuable agricultural commodity for the feed purpose. Globally, China (25 %), USA (20 %), Argentina (17 %), Brazil (16 %), European Union (6 %) and India (4 %) are the major producers of soymeal [5]. Soymeal accounts for approximately 69 % of all the protein sources used in animal feeds, much ahead of rapeseed meal (13 %), cottonseed meal (6 %), sunflower meal (5 %) and peanut meal (2 %) [5]. India is the fifth largest exporter of soybean meal with shipment worth of $2.85 billion and accounted for 9 % of the total global export [2]. Export of soymeal from India has been hovering between 3 and 4 million tonne in recent years. Besides, domestic consumption of soymeal in India is steadily rising from 3.33 million tonne in 2011–2012 to 3.7 million tonne in 2012–2013 and increased to 4.7 million tonne in 2013–2014. In international market, three categories of soymeal, viz. high-protein (Hi-Pro), low-protein (Low-Pro) and Indian soymeal, are traded. High-protein soymeal of USA, Argentina, Brazil and European Union contains 48.6, 46.8 and 48.9 %

✉ Vineet Kumar
vineetksahni@yahoo.com

[1] ICAR-Directorate of Soybean Research, Indore, Madhya Pradesh 452001, India

[2] Devi Ahilya Vishwavidyalaya, Indore, Madhya Pradesh 452001, India

protein, respectively, while protein content of Indian meal is 46.98 %. Increment in protein content in soymeal fetches premium in international market. The content of the protein in the soymeal depends upon its level in the seeds. Besides, high-protein soybean genotypes are also preferred for soyfood processing in the South-east Asian countries, where high-protein non-GMO soybean has a potential market [9]. In general, soybean seeds of cultivars on dry weight basis contain 38–40 % protein content. A range of 28–55 % has been reported for protein content in soybean USDA germplasm collection [12]. In view of the fact that newly released soybean varieties do not sustain more than 5 years, it is important to continuously identify and develop new genetic combinations with high protein content through exhaustive germplasm screening in parallel with focused crossing programme. For this purpose, the conventional kjeldahl method employed for determination of N content for computing protein content in segregating population in a plant breeding programme involving use of chemicals is cumbersome and time-consuming. Lately, near-infrared reflectance spectroscopy (NIRS) which is quick and does not require use of chemicals has been employed for estimation of protein content in several field crops, viz. Brassica species [8], sunflower [11], maize [14], wheat [10] and foxtail millet [3]. NIRS has been employed for the determination of protein content in soybean seeds also [6, 7, 13]. In the present investigation, we calibrated ground soyflour for protein content through NIRS and, subsequently, screened one thousand two hundred soybean germplasm accessions from different countries for the identification of genotypes with high protein content.

Materials and Methods

Materials

Seeds of 52 soybean genotypes were raised in the field, and freshly harvested seeds were used for spectroscopic analysis in the near-infrared reflectance spectroscopy. Near reflectance spectroscope of Brimrose Luminar 5030 model was used. Forty soybean genotypes taken up for developing calibration and 12 genotypes for validation model were of Indian as well as exotic origin. One thousand two hundred soybean genotypes screened for the protein content, were from 18 countries, namely India (851), USA (94), Taiwan (48), Philippines (20), Sri Lanka (11), China (10), Brazil (9), Hungary (8), Nigeria (6), Argentina(5), Germany (5), Australia (4), Thailand (4), Myanmar (3), Japan(2), Nepal (2), Canada (2) and Russia (1). One hundred twenty-five genotypes were of unknown origin.

Methods

Seeds of 52 genotypes were finely ground and passed through a sieve of 100 meshes and completely dried in an oven at 40 °C till the flour became moisture free. Moisture content of each genotype was calculated. Finely ground dried flour was subjected to nitrogen estimation for the triplicate determination of crude protein content through kjeldahl method, and the value of protein content so computed was converted on fresh weight basis. Freshly ground flour approximately (12 g) was layered into the rotating cup of NIRS with a uniform thickness, and the rotating button was pressed. The monochromator was set for generating 5 spectra of each sample. The wavelength range was set between 1100 and 2100 nm with a resolution of 2 nm in ratio mode. Scanning was done in timer mode with an interval of 5 s, and acquisition was stopped after 5 scans. Scanning procedure for one sample was completed in 25 s.

Of the 52 soybean samples, 40 were used for calibration, while the remaining 12 genotypes were set for external validation. Both the calibration set and validation set contained wide range of variability for protein content as determined through wet chemistry. Average protein content (fresh weight basis) of each genotype estimated through kjeldahl method was entered into an MS excel file 5 times. The calibration was carried out through partial least square mean regression (PLS1) at full cross-validation. Scans of the spectra saved were merged using Prospect SNAP 32 into U5 data file. The merged file of the spectra was processed, viz. spectroscopic transformation from reflectance to absorbance mode, and thereafter selection of processing operation Savitzky–Golay (SG) derivative in first derivative with smoothing filter 11. The processed file was exported into Unscrambler and saved into .uns file, which was imported into Unscrambler U5 data file, to obtain a matrix, which was finally subjected to partial least square mean regression (PLS1) at full cross-validation.

The prediction of the models was evaluated by computing the root-mean-square error of calibration (RMSEC), root-mean-square error of prediction (RMSEP), standard error of calibration (SEC), standard error of prediction (SEP) and determination of coefficient (R^2). R^2 value which explains the variation in the difference between the actual value and the predicted value attributed to the calibration and validation. In the calibration model, R^2 value in near-infrared spectroscopy needs to be near zero [1]. Lower values for SEC and SEP, which exhibit the difference between the predicted and measured values in calibration and validation, ascertain the efficiencies for future prediction for the protein content. The outliers at X and Y variables were marked and deleted by residual × influence plot.

Results and Discussion

Fifty-two genotypes taken up for developing calibration and validation models were categorized into classes with three per cent variation in protein content (Fig. 1). Three genotypes exhibited protein content more than 44 %. Maximum numbers of genotypes (21) were in the range of 38–40 %. Five genotypes exhibited protein content below 38 %. Sufficient variation for protein content was present in both the sets earmarked for calibration and validation. Protein content in samples (40) selected for calibration ranged from 36.81 to 44.21 % with average value of 40.21, while 12 samples selected for validation were in range of 36.72–45.76 with average value of 40.45 % (Table 1). Statistic values for calibration and external validation for the protein content are also given in Table 1. R^2 value for calibration and validation model was 0.95 and 0.98, while root-mean-square error was 0.32 and 0.36, respectively. Thus, the indices of evaluating the capabilities of the model for predicting the protein content were better, i.e. high values for R^2 and low value for root-mean-square error for validation test compared to the calibration model. This showed the suitability of use of models developed for protein predictability. Figure 2 shows the plots for reference versus predicted value for protein content for calibration as well as validation model. Lately, in a near-infrared spectroscopic analysis of seeds of 40 soybean genotypes for protein content, determination coefficient (R^2) of 0.88 has been reported [7]. Similarly, in a study carried out in soybean seed samples of 310 germplasm lines through NIRs, standard error of prediction and R^2 value to the magnitude of 0.568 and 0.927, respectively, have been reported [4]. In another investigation of whole kernels of soybean for crude protein using NIRS, SEC, SECV and R^2 to the magnitude of 0.610, 0.766 and 0.971, respectively, have been reported [6]. Our results showed lower values of SEC/SEP than the above-mentioned studies. This may be attributed to the finely ground soybean flour used in our study in contrast to the whole seeds in the previous studies. In rapeseed, SEC, SECV and R^2 to the magnitude of 0.38, 0.74 and 0.96, respectively, have been reported in an investigation for protein content through NIRS [8]. Higher standard error of validation compared to the standard error of calibration and R^2 value of >0.9 is similar to the observation in the above-mentioned studies.

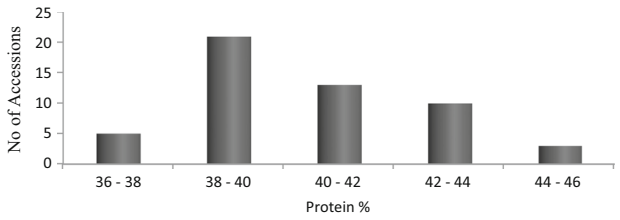


Fig. 1 Range for protein content among 52 genotypes through wet chemistry method

A close scrutiny of the absorbance spectra of average of 5 high-protein samples and 5 low-protein samples in the first derivative (Fig. 3) showed that the transmittance in the high-protein samples exceeded 1.6, while low-protein samples showed transmittance lesser than 1.4. Compared to the low-protein genotypes, high-protein soybean genotypes exhibited higher absorbance at regions, viz. 1325–1475, 1625–1725, 1875–1925 and 2025–2125 nm regions. On the contrary, low-protein genotypes exhibited higher absorbance samples at 1210–1325, 1500–1625 and 1950–2000 nm than the high-protein genotypes.

Of the 1210 soybean genotypes analysed for crude protein content, the lowest protein content (30.58 %) was observed in IC1576, while the highest (47 %) in Pant 7 (47 %). Both these genotypes are Indian soybean accessions. Average protein content of the 1210 germplasm lines was 39.5 %. The total numbers of genotypes were categorized into classes with three per cent variation in protein content (Fig. 4). Thirty-five genotypes exhibited protein content more than 45 %. Maximum numbers of genotypes (355) were found in the class of 38.0–40.0 % followed by the class of 36–38 % (341). The least number of accessions (11) were in the class of 30–32 % (Fig. 4). Among 851 Indian soybean germplasm accessions, protein content in 96 Indian soybean varieties released for cultivation ranged from 34 to 45 % with 39.5 % average protein content and in 410 advanced breeding lines, it ranged from 32 to 46.8 % with 39.5 average protein content. Three hundred forty-five accessions from Indian germplasm collection exhibited protein content in range of 32.0–47.0 % with 39.8 % as average protein content. In an investigation of screening of 135 soybean accessions for protein content by near-infrared reflectance spectroscopy, a range of 32.18–48.20 % has been reported for protein content [13]. The lowest and the highest ranges reported for the protein content by the authors in this study are in the proximity of values observed for the protein content in the present study.

Table 2 exhibits the range and average protein content in soybean accessions from different countries. Protein content in 94 germplasm accessions from USA ranged from 32.15 to 45.14 % with average protein content of 38.5 %, which was similar to the average protein content of 851 Indian soybean accessions. Average protein content of 851 accessions from India was lower compared to the average protein content of soybean accessions from Taiwan, Sri Lanka, China, Brazil, Hungary, Nigeria, Australia and Myanmar, though the number of accessions investigated from these countries was very less (Table 2).

In international market, Indian soybean meal, which is traded at 46.98 % protein content, is manufactured from

Table 1 Comparison between the measured and the predicted value by the regression model for the protein content (%) in soybean seed samples scanned through NIRs

	Max.	Min.	SD	Mean	Slope	RMSEC	SEC	RMSEP	SEP	R^2
Calibration	44.21	36.81	2.80	40.21	0.95	0.32	0.32	–	–	0.95
Validation	45.76	36.72	2.17	40.45	0.96	–	–	0.36	0.36	0.98

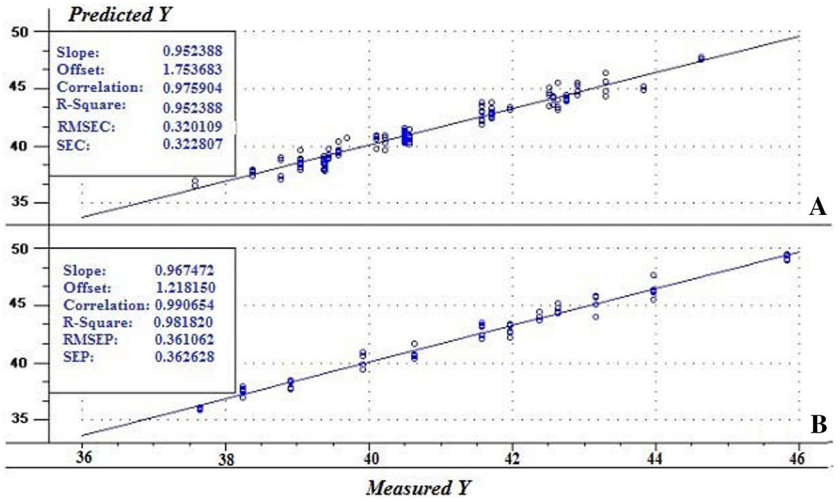


Fig. 2 Measured (wet chemistry) versus predicted (NIRS) values of protein content (%) in ground soybean samples for developing combined calibration (**a**) and validation (**b**) models

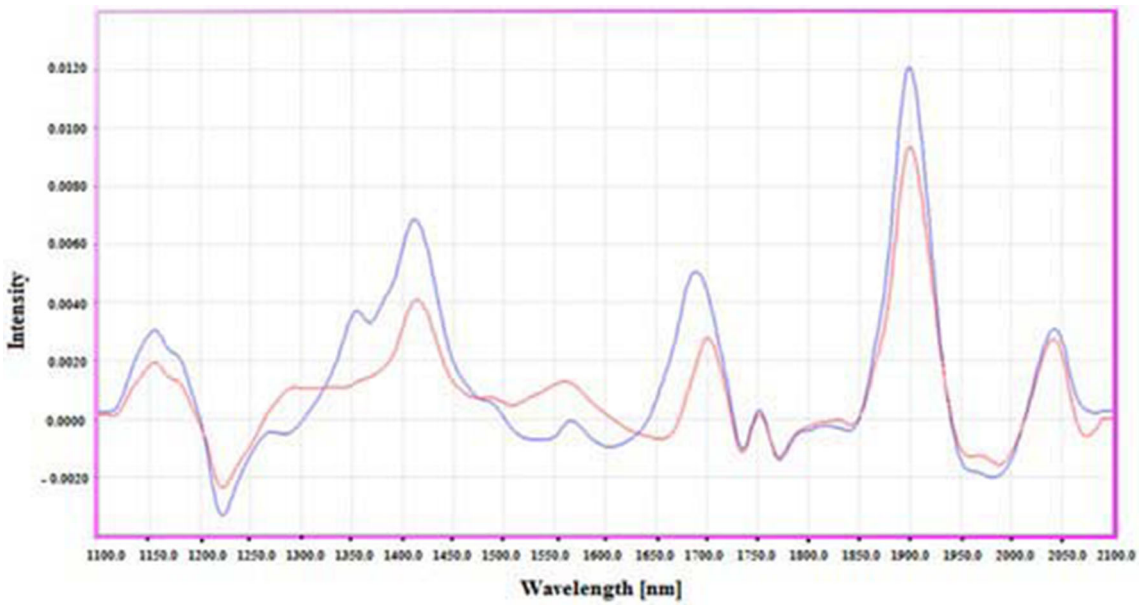


Fig. 3 Processed mean of 5 NIR spectra of high-protein (*blue line*) genotypes and low-protein (*red line*) soybean genotypes, obtained between 1100 and 2100 nm (color figure online)

the mixture of soybean varieties in the commercial cultivation. Average protein content of 94 Indian soybean varieties of the 851 Indian soybean accessions used in the present study was found to be 39.5 %, which was higher compared to the average protein content of accessions from USA (38.5 %), Nepal (38.73 %) and Russia (38.58 %) and lower compared to the accessions from Sri Lanka (41.5 %), China (40.62 %), Australia (40.92 %), Myanmar (41.01 %),

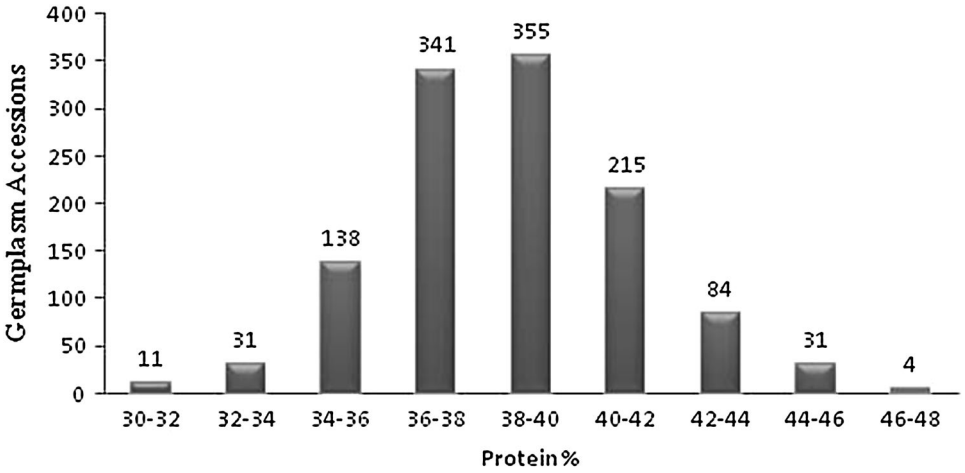


Fig. 4 Range of protein content in 1210 soybean germplasm lines as analysed by NIRS

Table 2 Range and mean protein content (%) of soybean accessions from different countries

Country	Accessions	Range	Mean	Accessions (>42 % protein)
India	851	30.48–47	38.50	98
USA	94	32.13–45.14	38.50	6
Taiwan	48	30.87–46.74	39.50	8
Philippines	20	33.33–38.58	36.50	0
Sri Lanka	11	40.13–43.27	41.5	4
China	10	36.67–45.01	40.62	2
Brazil	9	35.27–41.60	39.69	0
Hungry	8	35.44–41.48	39.61	0
Nigeria	6	38.05–43.82	40.01	2
Germany	5	33.07–37.75	35.62	0
Argentina	5	34.54–36.53	35.33	0
Australia	4	38.39–42.18	40.92	1
Thailand	4	37.17–39.12	38.5	0
Myanmar	3	43.91–39.12	41.01	1
Canada	2	41.83–38.39	40.11	0
Nepal	2	38.35–39.12	38.73	0
Japan	2	39.36–40.13	39.74	0
Russia	1	38.58	38.58	0

Canada (40.11 %) and Nigeria (40.01 %). Average protein content of soybean accessions from Taiwan (39.5 %), Brazil (39.69 %), Hungary (39.61 %) and Japan (39.74 %) was almost at par with that of Indian soybean varieties (39.5 %). Of the 1210 germplasm lines, 83 soybean accessions showed protein content less than 35 %. Average protein content of 5 germplasm accessions from Argentina and Germany was the lowest, followed by 20 accessions from Philippines, with magnitude of 35.33, 35.62 and 36.50, respectively. All the 11 accessions from Sri Lanka exhibited protein content more than 40 %, with average protein content of 41.5 %, which was the highest among all the accessions from 18 countries. Of the 122 germplasm accessions which exhibited protein content more than 42 %, 98 were from India, 6 from USA, 8 from Taiwan, 4 from Sri Lanka, 2 from China and 1 each from Australia and Myanmar, respectively. These accessions exhibiting high protein content from different countries in the present study may be harbouring diverse quantitative trait loci (QTLs) contributing to protein content. In brief, NIRS calibration and validation models developed in the study can be applicable for the estimation of protein content in the soy-flour samples. An exhaustive screening of soybean genotypes from different countries using this technique exhibited genotypes with high protein content from diverse genetic background. These genotypes can be combined to obtain transgressive segregants with protein content 48 % or more.

Compliance with Ethical Standards

Conflict of interest Authors declare that they have no conflict of interest.

Authors contribution Vineet Kumar and Anita Rani planned the work. The authors calibrated and validated the near-infrared spectroscope and Vaishali Mourya determined protein content in germplasm lines. S.M. Husain provided the germplasm. Meeta Jain was helpful in the standardization.

References

1. Burns DA, Ciurczak EW (1992) Handbook of near-infrared analysis, vol 13, in practical spectroscopy series. Marcel Dekker, Inc, New York, pp 393–395
2. Business Line (2014) Why India has lost mass hold in the global market? http://m.thehindubusinessline.com/markets/commodities/why-india-has-lost-its-hold-in-theglobal-oilmeals-market/article6583909.ece. Accessed 11 Dec 2015
3. Chen J, Ren X, Zhang Q, Diao X, Shen Q (2013) Determination of protein, total carbohydrates and crude fat contents of foxtail millet using effective wavelengths in NIR spectroscopy. J Cereal Sci 58:241–247

4. Choung MG, Baek IY, Kang ST, Han WY, Shin DC, Moon HP, Kang KH (2001) Non-destructive method for selection of soybean lines contained high protein and oil by near infrared reflectance spectroscopy. Korean J Crop Sci 46(5):401–406
5. Cromwell GL (2012) Soybean meal: an exceptional protein source. www.soymeal.org/reviews/papers/SBMExceptionalproteinfat
6. Dong Y, Qu SY (2012) Nondestructive method for analysis of the soybean quality. Int J Food Eng 8(4):1556–3758
7. Ferreira DS, Galao OF, Pallone JAL, Poppi RJ (2014) Comparison and application of near-infrared and mid-infrared spectroscopy for determination of quality parameters in soybean samples. Food Control 35:227–232
8. Hom NH, Becker HC, Mollers C (2007) Non-destructive analysis of rapeseed quality by NIRS of small seed samples and single seed. Euphytica 153(1–2):27–34
9. Hosoi T, Kan K (2008) Natto: A soybean food made by fermenting cooked soybeans with *bacillus subtilis* (natto). In: Farnworth ER (ed) Handbook of fermented functional foods, 2nd edn. CRC Press, Boca Raton, pp 267–290
10. Lindedam J, Bruun S, Demartini J, Jorgensen H, Felby C, Yang B, Wyman CE, Magid J (2010) Near infrared spectroscopy as a screening tool for sugar release and chemical composition of wheat straw. J Biobased Mater Bioenergy 4:1–6
11. Loggerenberg DV, Pretorious A (2011) Determining oil and protein content of striped and black sunflower seeds using NIR technology. Commun Soil Sci Plant Anal. doi:10.1081/CSS-200043362
12. National Genetic Resources Program, USDA-ARS (2001) Germplasm resources information network (GRIN). Online database of the NGR Laboratory, Beltsville. http://www.arsgrin.gov/npgs/descriptors/soybean
13. Peric V, Nikolic A, Drinic SM, Kravic N, Kovacevic D (2012) Characterization of soybean varieties based on protein and oil content. http://semenarska.rs/UNS-PSU/radovi/2/12%20Characterization%2064-67.pdf
14. Rosales A, Galicia L, Oviedo E, Ezequiel O, Islas C, Palacios-Rojas N (2011) Near-infrared reflectance spectroscopy (NIRS) for protein, tryptophan, and lysine evolution in quality protein maize (QPM) breeding programs. J Agric Food Chem 59(20): 10781–10786