Use of Near-Infrared Spectroscopy for Estimating Lignan Glucosides Contents in Intact Sesame Seeds

Kwan Su Kim1*, Si Hyung Park1, Kang Bo Shim2, Su Noh Ryu3

1 Department of Medicinal Plant Resources, Mokpo National University, Muan 534-729, Korea
2 Yeongnam Agricultural Research Institute, National Institute of Crop Science, Miryang 627-803, Korea
3 Department of Agricultural Science, Korea National Open University, Seoul 110-791, Korea

Abstract

Near-infrared spectroscopy (NIRS) was used to develop a rapid and efficient method to determine lignan glucosides in intact seeds of sesame (Sesamum indicum L.) germplasm accessions in Korea. A total of 93 samples (about 2 g of intact seeds) were scanned in the reflectance mode of a scanning monochromator, and the reference values for lignan glucosides contents were measured by high performance liquid chromatography. Calibration equations for sesaminol triglucoside, sesaminol (1→2) diglucoside, sesamolinol diglucoside, sesaminol (1→6) diglucoside, and total amount of lignan glucosides were developed using modified partial least square regression with internal cross validation (n = 63), which exhibited lower SECV (standard errors of cross-validation), higher R2 (coefficient of determination in calibration), and higher 1-VR (ratio of unexplained variance divided by variance) values. Prediction of an external validation set (n = 30) showed a significant correlation between reference values and NIRS estimated values based on the SEP (standard error of prediction), r² (coefficient of determination in prediction), and the ratio of standard deviation (SD) of reference data to SEP, as factors used to evaluate the accuracy of equations. The models for each glucoside content had relatively higher values of SD/SEP(C) and r² (more than 2.0 and 0.80, respectively), thereby characterizing those equations as having good quantitative information, while those of sesaminol (1→2) diglucoside showing a minor quantity had the lowest SD/SEP(C) and r² values (1.7 and 0.74, respectively), indicating a poor correlation between reference values and NIRS estimated values. The results indicated that NIRS could be used to rapidly determine lignan glucosides content in sesame seeds in the breeding programs for high quality sesame varieties.

Key words: near-infrared spectroscopy (NIRS), lignan glucoside, sesame (Sesamum indicum L.)

Introduction

Sesame (Sesamum indicum L.) is one of the most important oilseed crops worldwide, and has been cultivated in Korea since ancient times for use as sesame oil and traditional health food stuffs. The sesame oil and seed are mainly used for commercial products. An advantage of both sesame seed and sesame oil is their resistance to oxidative deterioration, resulting in oxidative stability during storage and processing (Fukuda et al. 1986). Sesame seed contains a number of antioxidants including lipid-soluble lignans such as sesamin and sesamolin, and water-soluble lignan glucosides such as sesaminol triglucoside, sesaminol diglucoside, and sesamolin monoglucoside (Katsuzaki et al. 1994a; Moazzami et al. 2006a). These antioxidants are the principle constituents in sesame seeds, and their amounts are important factors in evaluating seed quality. Average contents of lignans and lignan glucosides of Korean sesame varieties were 408 and 248 mg/100 g of sesamin and sesamolin, respectively (Kim et al. 2004), and 68.4 and 11.6 mg/100 g of sesaminol triglucoside and sesaminol diglucoside, respectively (Ryu et al. 1998). Lignan glucosides isolated from sesame seeds were mainly pinoresinol glucosides (Katsuzaki et al. 1994b), sesaminol glucosides with mono-, di-, and tri-glucosides (Katsuzaki et al. 1994a; Moazzami et al. 2006a), and sesamolinol diglucoside (Moazzami et al. 2006b). Moazzami et al. (2006a) isolated sesaminol diglucoside with a 1→6 bond in sugar configuration, which was different with sesaminol diglucoside with a 1→2 bond isolated by Katsuzaki et al. (1994a). In the results of the...
HPLC analysis, sesaminol triglucoside was the most abundant lignan glucoside in sesame seeds (Ryu et al. 1998; Shyu and Hwang 2002; Moazzami et al. 2006a).

Recently, more emphasis has been placed on evaluating qualitative traits in food processing and plant breeding, especially increasing bioactive lignan content in sesame. Qualitative determinations of sesame oils and seeds were performed by high performance liquid chromatography (HPLC) and gas liquid chromatography (GLC). Lignan is generally determined by an HPLC-UV/Vis system equipped with a reversed phase column (Kim et al. 2004; Moazzami et al. 2006a). However, this analysis method is time-consuming, expensive, labor-intensive, and also inefficient, therefore, it was not adequate for selecting superior lines from a number of sesame germplasm lines. Thus, a rapid and efficient method was required to evaluate seed quality for sesame breeding programs.

Near-infrared spectroscopy (NIRS) has been known as a powerful tool for analysis of chemical and physical properties without sample preparation, and it has been applied for the analysis of quality characteristics in food and agricultural commodities (Batten 1998; Williams and Noriss 2001). NIRS has been successfully used to determine diverse compounds in numerous foods and industrial crops such as soybean (Choung et al. 2005), perilla and peanut (Kim et al. 2007; Oh et al. 2000), sunflower (Fassio and Cozzolino 2004), rice (Kim et al. 2004), maize (Baye et al. 2006), rapeseed (Wu et al. 2006), and sesame (Sato et al. 2003). Kim et al. (2006) have reported NIRS could be used to determine lignans such as sesamin and sesamolin, and lignan glucosides such as sesaminol triglucoside and sesaminol (1–2) diglucoside in sesame, even though not including sesamolinol diglucoside and sesaminol (1–6) diglucoside as relatively abundant lignan glucosides in sesame seeds.

The objectives of this study were to improve NIRS application for estimating lignan glucosides, especially including newly separated two lignan di-glucosides, sesamolinol diglucoside and sesaminol (1–6) diglucoside, and to develop high-throughput screening techniques with intact seed samples for quality sesame breeding programs.

Materials and Methods

Sesame seed samples

A total of 93 samples of sesame breeding lines were obtained from Jeonnam Agricultural Research and Extension Services (Naju, Korea) including 20 Korean recommended sesame varieties, and were used to develop an NIRS prediction model for the determination of lignan glucosides. The sesame plants were grown in the greenhouse and harvested in 2004. The harvested seeds were cleaned, dried in the laboratory, and then stored in desiccators until NIRS and HPLC analysis.

Chemical analysis for lignan glucosides

About 2 g of each sample was homogenized using an Ultra-Turax T8 homogenizer (IKA-Werke GmbH & Co. KG, Staufen, Germany), and defatted three times in 30 mL of n-hexane for 1 day by shaking at 100 rpm using a VS-8480 SRN horizontal shaker (Vison Co., Bucheon, Korea). The defatted sesame meal (about 0.9 g) in a 50 mL conical tube was extracted with 40 mL 80% methanol/water for 1 day at room temperature by shaking at 150 rpm, and the supernatant was transferred to a 2 mL autosampler vial before HPLC injection for the determination of lignan glucosides. The Agilent 1100 Series HPLC instrument (Agilent Technologies Co., Palo Alto, CA, USA) was equipped with a 150 × 4.6 mm i.d. Develosil ODS-UG-5, reversed-phase column and HPLC detector operated at 290 nm. The mobile phase was a linear gradient from solvent A, methanol:water (30:70, v/v), to solvent B, methanol:water (80:20, v/v), in 40 min and flow rate was set at 1.0 mL /min. Running time was 60 min for each sample. Each peak of lignan glucosides was identified through further analyzing with LC-MS, and comparing HPLC retention time and UV spectra with each lignan glucoside, which was isolated by open chromatography and preparative HPLC before this study (data not reported).

Spectra collection and pretreatment

The NIR spectroscopic analysis was performed using a NIRSystem model 6500 near-infrared scanning monochromator (Foss NIRSystems Inc., Silver Spring, MD, USA) in the reflectance mode. Intact seed samples (about 2 g) were placed in a standard ring cup and then scanned. Reflectance energy readings were references to corresponding readings from an internal ceramic disc. Each spectrum was recorded once from each sample, and was obtained as the average of 32 successive scans over the sample, plus 16 scans over the standard ceramic before and after scanning the samples. All spectral data were recorded as the logarithm of the reciprocal of reflectance (log 1/R) in the wavelength range from 400 to 2500 nm, at 2 nm intervals. The scanning procedure could be finished within about 1.5 min per sample. The NIRS manipulation for scanning, mathematical processing, and statistical analysis was performed with the WinISI II software (Windows version 1.60) (Foss and Infrasoft International LLC, Stage College, PA, USA).

The samples (n = 93) were randomly split into two sets for calibration and validation using the WinISI program. The calibration set (63 samples) was used to calibrate and cross-validate
the equation derived, and internal cross validation was used to avoid overfitting of the equations (Shenk and Westerhaus 1996). The other 30 samples as an external validation set were used to test the efficacy of the developed equations by using random samples that were not included in the calibration sample set.

Data processing

The equations for NIRS prediction were developed using the Global program in WinISI software with the regression method of modified partial least squares (MPLS) using wavelengths of entire visible (400-1100 nm) and near-infrared (1100-2500 nm) region at every 8 nm. Various mathematical treatments using the raw optical spectrum (log 1/R), or first or second derivatives of the 1/R data, were applied for calibration equation development, with several combinations of smoothing and gap size. For instance, 2,5,5,1; second-order derivative, gap size, first smoothing, and second smoothing (Shenk and Westerhaus 1991). In addition to derivatives, scatter correction using standard normal variate and detrending (SNVD) was applied for the calibration to reduce the differences in spectra related to physical characteristics such as particle size and path length of samples (Barnes et al. 1989).

The best predicted equations for each chemical component were selected on the basis of minimizing the standard error of cross validation (SECV) and increasing the coefficient of determination ($R^2$) (Windham et al. 1989). The ratios of standard deviation (SD) of reference data to SECV and the corrected standard error of prediction (SEP(C)) were used as criteria to evaluate the performance of calibrations and the accuracy of equations, respectively (Williams and Sobering 1996). The developed equations were monitored with the Monitor program in WinISI software, using the validation set.
Results and Discussion

Variation of lignan glucosides contents in Korean sesame germplams

Four lignan glucosides, sesaminol triglucoside, sesaminol (1→2) diglucoside, sesamolinol diglucoside, and sesaminol (1→6) diglucoside isolated by the authors before this study were used as standards for HPLC analysis, and a total of 93 sesame germplasm accessions were evaluated for estimation of each lignan glucoside. Total lignan glucoside (TLG) was the total amount of four lignan glucosides determined here. Chemical structures of lignan glucosides were shown in Fig. 1. Average contents of lignan glucosides were 68.8 ± 32.3 mg/100 g of seeds of sesaminol triglucoside, 3.6 ± 2.1 mg/100 g of sesaminol (1→2) diglucoside, 19.9 ± 12.2 mg/100 g of sesamolinol diglucoside, and 14.9 ± 7.9 mg/100 g of sesaminol (1→6) diglucoside, and 107.1 ± 53.5 mg/100 g of TLG. Variation in the amount of each lignan glucoside was relatively large among sesame germplasm accessions as shown in Fig. 2. These mean values for lignan glucosides in sesame seeds were lower than those reported earlier (Shyu and Hwang 2002; Moazzami et al. 2006a).

Spectroscopic analysis

The NIR reflectance spectrum and standard deviation of absorbance of the intact sesame seed samples were shown in Fig. 3. Main absorption bands are observed at 1208 nm related to C-H stretching 2nd overtone (-CH2), 1496 nm related to C-H stretching 1st overtone, 1724 nm related to C-O (oil) and C-H stretching 1st overtone (-CH2), 1942 nm related to O-H bending 2nd overtone (water), and 2308 nm related to C-H bending 2nd overtone (-oil). The information of functional group in spectrum was searched from WinISI II software (Windows version 1.60), which gave chemical information associated with absorption bands in the specific NIR region. The overall spectrum showed strong absorption bands related to oil and water, and were similar to other oil crops such as perilla, peanut, and soybean, espe-
especially in the near-infrared region (Oh et al. 2000; Choung et al. 2005). The second derivative spectra had a trough corresponding to each peak in the original spectra, removing the overlapping peaks and baseline effects (Osborne et al. 1993). Average spectrum of the second derivative (Fig. 3B) showed absorption bands at 1680 nm related to C-H stretching 1st overtone (aromatic), and 1212, 1726, and 1762 nm related to hydrocarbon (-CH) in the NIR region.

Reference analysis of lignan glucosides

The descriptive statistics including mean and standard deviation (SD) for lignan glucosides contents of sesame samples used in the calibration and validation sets showed relatively large variation, being expected to be applicable to wide ranges of samples (Table 1). Each reference value of lignan glucosides in a validation sample set was similar to those in the calibration sample set. Mean values of lignan glucosides were 68.2 ± 32.6 mg/100 g of sesaminol triglucoside, 3.5 ± 2.2 mg/100 g of sesaminol (1→2) diglucoside, 19.5 ± 11.8 mg/100 g of sesamolinol diglucoside, and 14.6 ± 7.9 mg/100 g of sesaminol (1→6) diglucoside in the calibration set, and 70.0 ± 32.3 mg/100 g of sesaminol triglucoside, 3.7 ± 1.8 mg/100 g of sesaminol (1→2) diglucoside, 20.4 ± 13.1 mg/100 g of sesamolinol diglucoside, and 13.1 ± 8.1 mg/100 g of sesaminol (1→6) diglucoside in the validation set, respectively.

Table 1. Descriptive statistics for lignan glucosides in intact sesame samples used in both calibration and validation.

<table>
<thead>
<tr>
<th>Constituents</th>
<th>Calibration (n = 63)</th>
<th>Validation (n = 30)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean* Range</td>
<td>SD Mean Range</td>
</tr>
<tr>
<td>Sesaminol triglucoside</td>
<td>68.2 22.4-148</td>
<td>32.6 70.0 7.4-128</td>
</tr>
<tr>
<td>Sesaminol (1→2) diglucoside</td>
<td>3.5 0.85-9.5</td>
<td>2.2 3.7 0.45-7.9</td>
</tr>
<tr>
<td>Sesamolinol diglucoside</td>
<td>19.5 3.6-54.9</td>
<td>11.8 20.4 1.1-51.0</td>
</tr>
<tr>
<td>Sesaminol (1→6) diglucoside</td>
<td>14.6 3.3-33.2</td>
<td>7.9 13.1 0.71-32.1</td>
</tr>
<tr>
<td>TLG</td>
<td>105.9 31.6-243</td>
<td>53.5 109.5 10.1-209</td>
</tr>
</tbody>
</table>

*Expressed as milligrams per gram (mg/100 g of seeds); SD, standard deviation of mean; TLG, total lignan glucosides, total amount of four lignan glucosides: sesaminol triglucoside, sesamolinol diglucoside, and sesaminol (1→6) diglucoside.

Calibration model for lignan glucosides

In developing NIRS models for contents of lignan glucosides, the statistics of calibrations and cross-validations are shown in Table 2. The MPLS regression model using the second derivative transformation with scatter correction (SNVD) of raw reflectance spectra yielded the equations of each lignan glucoside. The best equations for sesaminol triglucoside and TLG were developed using mathematical treatment 2,8,6,1 in the whole visible and NIR spectra range (400-2500 nm). The best equations for two isomers of sesamolin diglucosides were obtained using mathematical treatment 2,5,5,1 in near infrared region (1100-2500 nm) showing higher R² and 1-VR, and lower SEC and SECV values than the different mathematical treat-
ments tested, while that for sesamolinol diglucoside was obtained by mass treatment 2,3,3,1 in the whole range. The equations for each lignan glucoside were selected considering $R^2$ and SD/SECV values with the highest values, 0.91 and 2.47 of sesaminol triglucoside, 0.90 and 1.83 of sesaminol (1→2) diglucoside, 0.97 and 2.37 of sesamolinol diglucoside, 0.93 and 2.29 of sesaminol (1→6) diglucoside, and 0.92 and 2.59 of TLG, respectively, as the selection criteria of models. Although showing relatively low SD/SECV value in sesaminol (1→2) diglucoside containing the lowest amount among lignan glucosides, the best prediction models for lignan glucosides were successfully developed with the mathematical approach over the visible and/or near infrared segment (400-2500 nm), and the results showed that the equations could be used for screening lignan glucosides contents in intact sesame seeds.

**External validation**

The robustness of calibration models developed by NIRS analysis was tested through external validation (prediction) with 30 samples, which were not included in the calibration process. The statistics of external validation for lignan glucosides in intact sesame seeds included bias, $r^2$, SEP(C) (the corrected standard error of prediction), and SD/SEP(C) values, which were factors used to evaluate the reliability of the calibration model. The $r^2$ values for sesaminol triglucoside, sesaminol (1→2) digluco-
Determination of Lignan Glucosides in Sesame Seeds by NIRS

coside, sesamolinol diglucoside, sesaminol (1–6) diglucoside, and TLG were 0.82, 0.74, 0.80, 0.80, and 0.81, respectively. The SEP(C) values for each compound were 14.1, 0.74, 5.23, 3.58, and 25.0 mg/100 g, respectively. Based on lower SEP(C), and higher $r^2$ and SD/SEP(C) values, accurate prediction can be monitored with the reliability of the established calibration models. We obtained relatively higher values (more than 2.0) of SD/SEP(C), which were 2.28 for sesaminol triglucoside, 2.23 for sesamolinol diglucoside, 2.20 for sesaminol (1–6) diglucoside, and 2.19 for total lignan glucoside indicating good correlation between reference values and NIRS predicted values in the application of the calibration equations, though their SD/SEP(C) values were lower than the cut-off point (3.0) recommended for screening purposes (Williams and Sobering 1996). However, the relatively low value of SD/SEP(C), 1.67 for sesaminol (1–2) diglucoside showing a minute quantity in sesame seeds, indicated a poor relationship between the reference values and the NIRS estimated values.

Figure 4 displays NIRS predicted values against laboratory reference values in the validation set for sesaminol triglucoside, sesaminol (1–2) diglucoside, sesamolinol diglucoside, sesaminol (1–6) diglucoside, and TLG, showing also the relationship between NIRS and the reference.

These results demonstrated the accurate prediction capacities of the calibration models for lignan glucosides using a non-destructive NIRS method in sesame seeds. This study is the first in which lignan glucosides including most glucosides in Korean sesame germplasm were quantitatively screened using HPLC, and contents of each lignan glucoside, especially three diglucosides, were estimated by NIRS method. The sample set used in this study was not very large and included limited variation in environmental and genetic factors. Therefore, the developed models may need to be improved by applying new samples from different germplasm accessions collected under various environments.

Acknowledgement

This study was carried out with the support of "Specific Joint Agricultural Research-promoting Projects (Code No. 20070101033050)"), Rural Development Administration, Republic of Korea.

References


Katsuzaki H, Osawa T, Kawakishi S. 1994b. Chemistry and antioxidative activity of lignan glucosides in sesame seed. In: Food Phytochemicals for Cancer Prevention II (Eds. Ho CT, Osawa T, Huang MT, Rosen RT), American Chemical Society


Osborne BG, Fearn T, Hindle PH. 1993. Practical NIR spec-


