J Am Oil Chem Soc (2017) 94:69–76
DOI 10.1007/s11746-016-2916-4

ORIGINAL PAPER

Development of Rigorous Fatty Acid Near-Infrared Spectroscopy Quantitation Methods in Support of Soybean Oil Improvement

Avinash Karn[1] · Crystal Heim[1] · Sherry Flint-Garcia[2] · Kristin Bilyeu[2] · Jason Gillman[2]

Received: 8 August 2016 / Revised: 31 October 2016 / Accepted: 1 November 2016 / Published online: 16 November 2016

Abstract The mature seeds of soybean (*Glycine max* L. Merr) are a valuable source of high-quality edible lipids and protein. Despite dramatic breeding gains over the past 80 years, soybean oil continues to be oxidatively unstable, due to a high proportion of polyunsaturated triacylglycerols. Until recently, the majority of soybean oil underwent partial chemical hydrogenation. Mounting health concerns over *trans* fats, however, has increased breeding efforts to introgress mutant and biotechnological genetic alterations of soybean oil composition into high-yielding lines. As a result, there is an ongoing need to characterize fatty acid composition in a rapid, inexpensive and accurate manner. Gas chromatography is the most commonly used method, but near-infrared reflectance spectroscopy (NIRS) can be calibrated to non-destructively phenotype various seed compositions accurately and at a high throughput. Here we detail development of NIRS calibrations using intact seeds for every major soybean fatty acid breeding goal over an unprecedented range of oil composition. The NIRS calibrations were shown to be equivalent to destructive chemical analysis, and incorporation into a soybean phenotyping operation has the potential to dramatically reduce cost and accelerate phenotypic analysis.

Keywords Oilseeds · Crop production and agronomy · Genetics/breeding · Lipid chemistry/lipid analysis

Abbreviations

NIRS	Near-infrared reflectance spectroscopy
GC	Gas chromatography
MS	Multiple scatter correction
SEP	Standard error of performance
RMSEP	Root mean square error of prediction
RPD	Relative prediction deviation
FAs	Fatty acid species
PLS	Partial least squares

Introduction

Soybean (*Glycine max* (L.) Merr) is one of the world's most important crop plants, accounting for 59% of total global oilseed production (http://soystats.com/2015-soystats/, accessed 5-24-2016). Soybean seeds are highly prized for the high protein and oil content. A proximate composition of whole soybean on a dry matter basis consists of ~40% protein, ~20% lipid, ~35% carbohydrate, and ~5% ash. These values vary depending on cultivar and specific growing conditions [1]. The soybean seed lipid pool is almost completely (~88%) in the form of triacylglycerols [2]. In the USA, soybean oil is the predominant edible oil, accounting for 55% of total oil consumption and internationally soybean oil is also a major consumed edible oil at 27% of the market (second only to palm oil, http://soystats.com/2015-soystats/, accessed, 5-24-2016).

The functionality and physical properties of plant oils are largely a consequence of their fatty acid composition. Soybean oil contains five principal fatty acid species: ~11.6% palmitic (C16:0), ~2.5% stearic (C18:0), ~21.1%

Electronic supplementary material The online version of this article (doi:10.1007/s11746-016-2916-4) contains supplementary material, which is available to authorized users.

✉ Jason Gillman
Jason.Gillman@ars.usda.gov

[1] Division of Plant Sciences, University of Missouri, Columbia, MO 65211, USA

[2] U.S. Department of Agriculture-Agricultural Research Service, Columbia, MO 65211, USA

oleic (C18:1), ~52.4% linoleic (C18:2), and ~7.1% linolenic (C18:3) acids [3]. The high proportion of oxidatively unstable polyunsaturated fatty acids has historically resulted in the need for chemical hydrogenation, however this practice has become less prevalent in recent years due to increasing consumer health concerns surrounding *trans* fatty acids, which are generated as an unavoidable side effect during partial hydrogenation [4]. Numerous epidemiological studies have shown that dietary intake of *trans* fats has strong deleterious effects on cardiovascular health due to the impact on low-density lipoprotein (LDL) levels in blood serum [5, 6]. These findings ultimately resulted in the 2016 ruling by the Food and Drug Administration (FDA) that partially hydrogenated oils are not "Generally Regarded as Safe" (http://www.fda.gov/ accessed 05-24-2016).

As a result, it is of interest to the market to develop soybean lines with altered seed oil composition that would reduce the need for chemical hydrogenation. Several genetic methods, including mutagenesis, and biotechnological and conventional breeding, have been used to meet this goal. This ongoing breeding effort was recently reviewed by Gillman & Bilyeu [7] and Medic et al. [8]. It is now possible to bring about directed, dramatic alterations in fatty acid composition of soybean oil. The primary method of quantifying fatty acid composition is direct chemical analysis by gas chromatography (GC) is widely accepted, but it is destructive, low-throughput, and is time and resource intensive. Consequently, there has also been an increased interest in novel, high-throughput, and non-destructive methods for quantification of soybean seed fatty acid composition.

An alternative technology for measuring the constituents of biological materials is provided by near infrared reflectance spectroscopy (NIRS), which relies upon indirect measurement of seed chemical components via their interaction with infrared radiation [9]. NIRS is highly reproducible, non-destructive, less time-intensive than other methods, and has been suggested to have the potential to achieve the accuracy of reference analytical tests [10, 11]. NIRS calibration models have frequently been developed using ground seeds to avoid issues with scattering and non-homogenous distribution of seed constituents. However, grinding of seeds is destructive and has limited applicability in early stages of breeding when seeds are frequently limited. Other NIRS high-correlation calibration models ($r \geq 0.92$) have been developed for simple seed constituents (crude protein, oil, etc.) [11, 12].

For soybean fatty acid species, several NIRS calibrations have been developed [13–15], which generally show lower prediction correlations as compared to crude protein or oil. Although several models exist, none is ideal: Certain NIRS calibration models had been designed only with ground seed samples [15], used a small (<300) number of samples [14, 15], used only existing natural diversity with a narrow range of phenotypes [14, 16] or were not validated with an external sample set [15].

Certain seed compositional changes are too rare in natural germplasm pools to rely on existing diversity; e.g. the US Department of Agriculture-Germplasm Resources Information Network (USDA-GRIN) collection contains only a single entry which is listed as being either >50% oleic acid and only one entry >9% stearic acid (https://npgsweb.ars-grin.gov/gringlobal/descriptors.aspx, accessed 05-23-2016). As the breeding goal for stearic acid is currently >20%, there is a need for precise and accurate NIRS calibration. No existing calibration has featured a significant range of stearic acid concentrations; they are expected to be inaccurate or uninformative when evaluating seeds with significantly altered stearic acid composition. For oleic acid, bulk and single-seed calibrations have been designed to quantify oleic acid specifically [13, 17], but these calibrations were not designed to be comprehensive for all possible fatty acid breeding goals.

To improve on these methods, we specifically developed and characterized several unique soybean germplasm pools with one to several mutations in nine distinct fatty acid biosynthesis genes (Supplementary Table 1). This set encompasses all current major soybean fatty acid breeding targets: (1) increased oxidative stability via reduction of linoleic and linolenic acid content; (2) limited saturated fat content via reduction of palmitic acid; and (3) increased stearic acid content for solid fat applications. NIRS calibrations were then developed to predict the five major fatty acids in bulk, intact seeds in an unprecedentedly broad range of soybean seed oil compositional variation.

Experimental Procedures

Plant Germplasm and Growing Conditions

Elevated stearic acid content was represented by a population developed from crossing two elevated stearic acid lines: 'A6' and '194D'. Soybean line 'A6' has seed stearic acid levels of ~28%. 'A6' was reported to be a sodium azide-induced mutant line [18]. However, it bears a ~6221 kbp genomic deletion containing the *stearoyl acyl carrier desaturase isoform c* (*sacpd-c*) gene which is more consistent with radiation-induced mutagenesis [19]. Line '194D' is an ethyl methanesulfonate (EMS)-induced 'Williams 82' mutant line that bears a *sacpd-c* missense mutation in an ancestrally invariant amino acid residue (V211E) and has ~9–13% seed stearic acid [19]. 'Williams 82' typically has ~3% stearic acid (ARS-GRIN). 'A6' and '194D' were crossed in 2012 in a reciprocal cross, and advanced

to 176-RIL $F_{4:5}$ populations. No difference was noted in oil or stearic acid content between directions of cross (male to female), so populations were combined. The $F_{4:5}$ populations, along with both parents and conventional controls were planted June 11, 2015 in three replicate plots in a randomized complete block design in 2015 at a field location near Columbia (Hinkson Bottom, latitude 38.928°N, longitude −92.352°W). Each plot was composed of 10 seeds planted in a 1-foot long row with a 2-foot gap. Row spacing was 30 inches. When mature, each plot was harvested in bulk.

An additional set of soybean lines grown in a separate experiment (high oleic, low linolenic; HOLL) were selected to provide a comprehensive NIRS calibration which encompasses all other major breeding efforts in soybean oil improvement (reduced linolenic acid, reduced palmitic acid and increased oleic acid, Supplementary Table 1). Mutations of the FAD2-1A locus result in elevation of seed oleic acid content from ~20% in wild type lines to ~34–42% (depending on the severity of the specific mutation); mutations affecting the FAD2-1B locus elevate oleic acid content to ~28%. These mutations are synergistic, and when both mutations are homozygous, result in seed oleic acid levels from 65 to 86% oleic acid; the specific oleic acid level is dependent on the allelic combination. Most of the additional lines used were homozygous for *FAD21-A* and *FAD2-1B* mutant alleles, and some lines contained additional mutant alleles of other fatty acid related genes. Breeding lines were developed by crossing parents with defined alleles of genes controlling different fatty acids and selection with molecular markers in the F_2 or F_3 generations for different combinations of seed oil traits. The gene, allele, original source, and molecular reference are listed for each trait in supplementary Table 1.

Samples of 23 experimental lines were produced at two locations: South Farm Research Center near Columbia, Missouri (SF), and the Greenly Memorial Research Center near Novelty, Missouri (Novelty). Seeds for another 16 experimental lines were produced only at South Farm Research Center, due to limited availability of seeds. The experimental lines were grown in three replicate plots in a randomized complete block design, and each plot was composed of 10 seeds planted in 1-foot rows, with row spacing of 30″. Each plot was harvested in bulk. For 2015, SF was planted June 7 and Novelty was planted June 10.

Two additional samples sets, (1) mutant lines (KK2, KK24, M25 and 194D); and (2) eight wild-type cultivars/breeding lines were also included (supplementary Table 1). These samples were produced as single-plot replications over a range of locations and years (2012–2015). These samples served to ensure the NIRS calibration could also quantify non-mutant phenotypes.

Gas Chromatography

Chemicals used in GC analysis were: chloroform–hexane–methanol (8:5:2, v/v/v, Fisher Scientific, Fair Lawn, NJ, USA), methylating reagent (0.25 M methanolic sodium methoxide-petroleum ether-ethyl ether, 1:5:2, v/v/v, Grace Discovery Sciences, Deerfield, IL, USA), and hexane (Sigma-Aldrich, St. Louis, MO, USA).

Analytical values for percent individual fatty acids (C16:0, C18:0, C18:1, C18:2 and C18:3) in the 'A6' × '194D' populations were obtained using five single seeds from each plot. For all samples except the experimental lines, seeds were not pooled, but five seeds were individually extracted. For experimental lines, three seeds were pooled per plot. Individual or pooled seeds were crushed and oils extracted in 1 mL of chloroform–hexane–methanol overnight. Derivatization of 150 µL of solvent was done with 75 µL of methylating reagent. Single seeds were diluted with hexane to 1 mL. An Agilent (Palo Alto, CA, USA) series 6890 capillary gas chromatograph fitted with a flame ionization detector (275 °C) was used with an AT-Silar capillary column (Alltech Associates, Deerfield, IL, USA). Standard fatty acid mixtures (Animal and Vegetable Oil Reference Mixture 6, AOACS, Matreya, LLC, State College, PA, USA) were used as reference standards. Values are expressed as percentage values of the total seed oil.

NIRS Data Collection

Pigmented seedcoats are known to complicate the development of accurate NIRS calibrations [11], and pigmented seedcoats are not typical of released soybean cultivars, except for certain small specialty markets. Therefore, all of the samples in this study featured yellow seedcoats.

Approximately 50–100 intact soybean seeds from each sample were scanned using the NIRS monochromator model FOSS 6500 (FOSS North America, Eden Prairie, MN, USA) using the transport quarter cup (dimension 97 × 55 mm). The NIRS reflectance (R) spectra were collected at 2-nm intervals in the NIRS region of 400–2500 nm at room temperature, and absorbance values were calculated as log (1/R). About 30 s were required to collect spectra from each sample. Collected raw NIRS spectra were reported using ISIscan® software and exported as text files via WinISI® monochromator instrument standardization IV software (FOSS North America, Eden Prairie, MN, USA) for spectral preprocessing and development of

Table 1 Fatty acid statistics for the NIRS calibration set

Fatty acid	*n*	mean	SD	CV	Range	Difference
C16:0	687	8.91	1.32	0.05	2.78–12.62	9.84
C18:0	687	12.30	6.25	0.24	1.85–28.04	26.19
C18:1	687	34.65	24.62	0.94	16.05–89.44	73.39
C18:2	687	38.43	18.37	0.70	1.24–58.66	57.42
C18:3	687	5.71	1.63	0.06	1.75–9.45	7.11

n number of samples, *SD* standard deviation, *CV* coefficient of variation

calibration models using the UnScrambler® software 6.11 (CAMO ASA, Trondheim, Norway).

Statistical Analyses

NIRS spectra from 687 soybean samples were collected and principal component analysis (PCA) was conducted in UnScrambler® to classify the spectral data and determine outliers [20]. The 687 samples were randomized in Microsoft Office Excel 2010, and separated into two sets: 596 samples for a calibration set and 93 samples for an external validation set (Tables 2 and 3). The NIRS calibration dataset had 1050 spectral data points (independent variables) and 5 individual fatty acid species (FAs, dependent variables). A partial least squares (PLS) regression [21] method was performed to derive calibration models for individual FA composition in the UnScrambler® software [22]. PLS regression models were obtained on pre-processed NIRS spectra as descriptors (*X* matrix) and the analytical data as response data set (*Y* matrix) for individual fatty acid composition. Absorbance spectra from 900 to 2500 nm were included in PLS regression analysis due to a higher level of noise in the spectral region at the lower wavelengths (400–900 nm). For all FAs, the NIRS spectra were preprocessed first with multiplicative scatter correction (MSC) [23], and then with the Savitzky–Golay first derivative [24] in the UnScrambler® software, simultaneously. For oleic acid/C18:1, most of the spectral variation was in the 1100–1800 nm region, and NIRS spectra from this spectral region were used to improve the calibration by preprocessing the raw spectra as described above as well as taking the second derivative [23] in the UnScrambler® software. The wavelength range and spectral preprocessing methods were chosen to optimize the calibration performance, and to minimize error during the cross-validation (CV) analysis. The significant PLS factors to build regression models for each fatty acids were determined by using segmented CV in UnScrambler®, where the 596 samples in the calibration set were divided into subgroups or 'segments'. One segment at a time was omitted from the calibration, and the calibration model was developed with the remaining samples. The omitted samples were then predicted, and the entire procedure was repeated until each sample in the calibration set was removed and predicted. For each calibration, a small number of samples were classified as outliers and were not included in the final NIRS calibration. Finally, an optimal calibration model was selected by using a significant number of PLS factors with the default setting of the software [25].

The performance of the NIRS calibration models for individual FAs was first evaluated by the multiple coefficient of correlation (*r*), standard error of calibration (SEC) and standard error of cross-validation (SECV) and root mean square of error for cross-validation CV [26]. The individual FA NIRS models were then tested by predicting external validation samples ($n = 93$). Each FA NIRS model was reported as the coefficient of correlation (*r*), standard error of performance (SEP), root mean square error of prediction (RMSEP), and the relative prediction deviation (RPD) [27]. RPD is the ratio of the SD for the validation samples to the SEP. In general, RPD values >2.4 indicate the model can be used to predict values for a trait, whereas, RPD values <2.4 suggest that the model can be used to group seeds according to high and low values for the trait [27]. These values (SEC, SECV, RMSEV, SEP and RMSEP) were automatically calculated by the UnScrambler® software.

Results and Discussion

Table 1 and Fig. 1 show the range of the fatty acid composition of the soybean accessions used in this study, which included both mutant and conventional lines. The calibration and validation sets had comparable ranges, means, and standard deviations for all fatty acids measured. The soybean accessions used to generate the NIRS calibration represented an extremely broad range of values for all fatty acids except palmitic (C16:0) acid. However, palmitic acid values in this set were representative of phenotypic variation in conventional soybean lines as well as common palmitic acid mutants.

Table 2 shows the partial least squares regression statistics for the NIRS calibration. The regression analysis was completed by comparing the *r* and SEC of the calibration data set to the *r* and SEP in the external validation set.

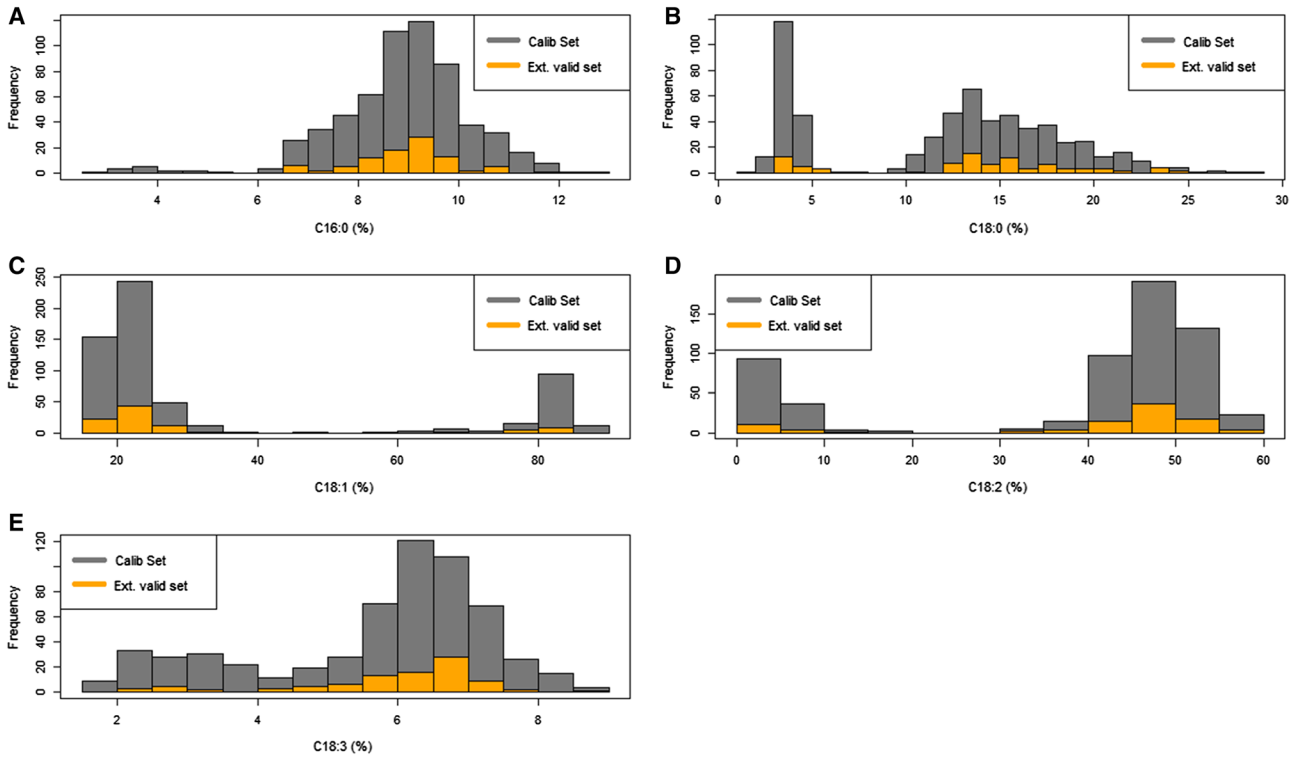

Fig. 1 Histogram distribution of individual fatty acids content (% seed oil) in samples from the NIRS calibration (*gray*) set and in the NIRS external validation (*orange*) set: **a** C16:0 (palmitic acid), **b** C18:0 (stearic acid), **c** C18:1 (oleic acid), **d** C18:2 (linoleic acid) and **e** C18:3 (alpha-linoleic acid) (color figure online)

Table 2 Calibration statistics for partial least squares regression models for individual fatty acids

Fatty acid	*n*	Spectral range (nm)	NIRS pretreatment	PLS factors	SEC	SECV	RMSECV	R
C16:0	583	900–2500	MSC; 1 Der	7	0.64	0.67	0.67	0.82
C18:0	588	900–2500	MSC; 1 Der	12	1.78	2.17	2.17	0.95
C18:1	586	1100–1800	MSC; 1 Der; 2 Der	5	3.89	4.47	4.46	0.98
C18:2	591	900–2500	MSC; 1 Der	6	3.61	3.73	3.73	0.98
C18:3	584	900–2500	MSC; 1 Der	6	0.64	0.66	0.66	0.92

n number of samples, *NIRS* near-infrared specgtroscopy, *PLS* partial least squares, *SEC* standard error of calibration, *SECV* standard error of cross-validation, *RMSECV* root mean square error for cross-validation, *r* coefficient of correlation

Several spectral pretreatments including the first derivative (1 Der), second derivative (2 Der), MSC as well as combinations between pretreatments were evaluated. NIRS spectra from both the calibration and validation sets (treated with combinations of MSC and 1 Der) had similar correlation coefficients, as well as similar SEC and SEP values for each fatty acid. These results suggested that the spectral pretreatments were successful in improving the accuracy of the PLS regression.

The NIRS calibrations showed a strong correlation between GC-measured reference values and NIRS-predicted values for all fatty acids examined (Table 2). *r* values were 0.82, 0.95, 0.98, 0.98, and 0.92 for C16:0, C18:0, C18:1, C18:2, and C18:3, respectively (Fig. 2). These results were consistent with those obtained from segmented CV (Table 2).

In order to determine the applicability and predictability of the NIRS calibration models for all fatty acids, we tested an external validation set of randomly selected samples that were not included in the original calibration sets. The 93-sample external validation set spanned the same fatty acid composition ranges as seen in the calibration set (Table 3). The calibrations were successful in determining fatty acid composition in the external validation set, with *r* values ranging from 0.77, 0.95, 0.98, 0.98, and 0.82 for C16:0, C18:0, C18:1, C18:2, and C18:3, respectively

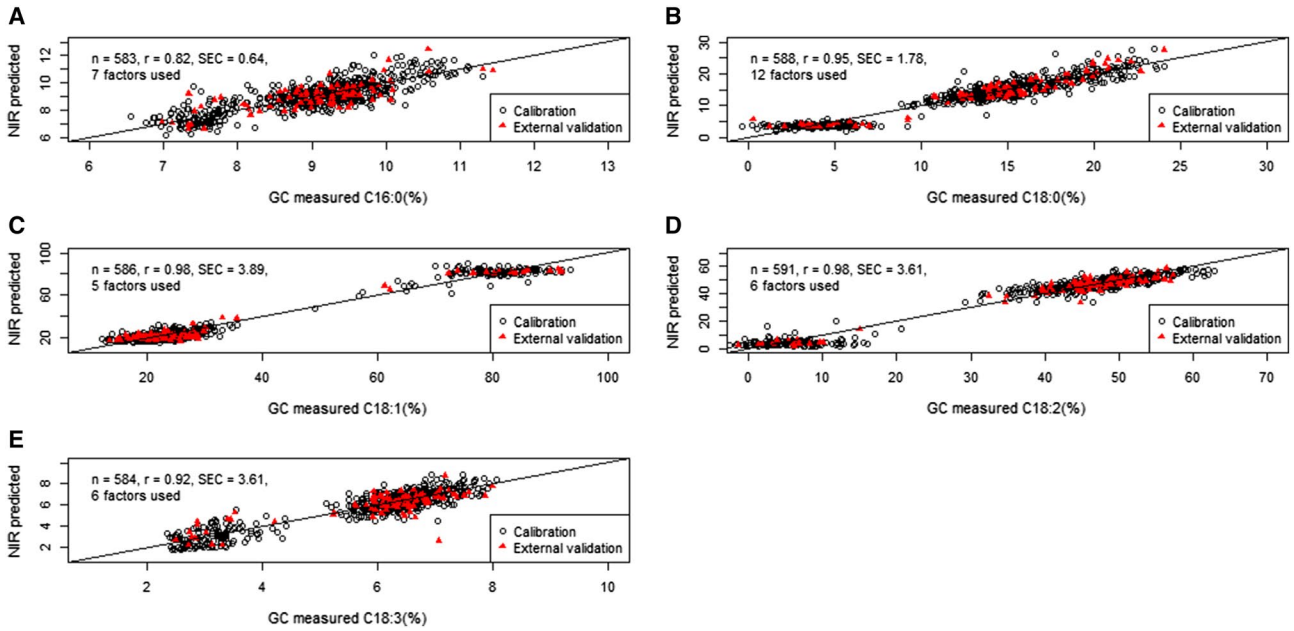

Fig. 2 *Scatter plots* of fatty acid content (% seed oil) of the NIRS-predicted (*Y*-axis) and analytically determined (GC method; *X*-axis) for a calibration set (*black open circles*) and for an external validation set (*red closed triangles*) for individual fatty acids with their corresponding coefficient of correlation (*r*), standard error of calibration (SEC) and partial least squares (PLS) factors for each fatty acid in the calibration set: **a** C16:0 (palmitic acid), **b** C18:0 (stearic acid), **c** C18:1 (oleic acid), **d** C18:2 (linoleic acid) and **e** C18:3 (linolenic acid). Each plot shows a linear regression trend line (color figure online)

Table 3 External validation statistics in NIRS models for the estimation of individual fatty acids

Fatty acid	*n*	mean	range	SD	SEP	RMSEP	RPD	*R*	*T* Stat
C16:0	93	8.97	6.58–12.44	1.04	0.66	0.65	1.57	0.77	0.74
C18:0	93	13.45	3.24–27.57	6.17	1.85	1.84	3.34	0.95	−0.38
C18:1	99	33.14	17.26–84.86	23.6	4.27	4.27	5.53	0.98	1.35
C18:2	93	40.34	1.24–58.66	16.41	3.38	3.39	4.85	0.98	1.34
C18:3	93	5.91	2.20–8.78	1.36	0.82	0.81	1.69	0.82	0.60

n number of samples, *SD* standard deviation, *SEP* standard error of performance, *RMSEP* root mean square error for prediction, *RPD* ratio of standard deviation of data to standard error of performance, *r* coefficient of correlation, *T* Stat *t* test statistic

(Table 3). Scatter plots comparing the GC-based fatty acid composition and the NIRS-predicted values showed a highly similar prediction error, with a very slight bias toward overestimating fatty acid content (Fig. 2).

Igne et al. developed methods to predict the major fatty acids in whole soybean seeds using a near-infrared transmittance spectroscopy (NITS) instrument (Infratec Grain Analyzers 1229 and 1241, FOSS North America, Eden Prairie, MN, USA) and two Bruins OmegAnalyzerG instruments (Bruins Instruments, Puchheim, Germany) with a transmittance module [28]. Raw spectra were collected from a much narrower wavelength band than in the current study (850–1048 nm at an increment of 2 nm). Igne et al. used approximately 900 calibration samples, and used both linear (partial least squares) and non-linear (artificial neutral networks and a least squares support vector machine) regression models to develop prediction models for palmitic acid (C16:0) [$r^2 = 0.97$, SEP = 0.64%], stearic acid (C18:0) [$r^2 = 0.85$, SEP = 0.30%], oleic acid (C18:1) [$r^2 = 0.59$, SEP = 1.62%], linoleic acid (C18:2) [$r^2 = 0.77$, SEP = 1.51%] and linolenic acid (C18:3) [$r^2 = 0.95$, SEP = 0.64%] using PLS regression. In comparison to the NIRS/PLS regression results, the prediction capability of the NITS spectral derived model has lower SEP for all fatty acids, i.e. lower prediction error, and higher prediction accuracy. The PLS regression results obtained using these two different instruments suggest that NITS is efficient in predicting C16:0 and C18:3 fatty acids with high accuracy and low error, but were less effective at the other fatty acids in soybean seeds.

These results are most likely due to the limited genetic diversity available at the time. Germplasm were available which featured excellent phenotypic distributions for palmitic acid (2.89–13.64% seed oil) and linolenic acid (0.89–11.08% seed oil), but only limited genetically—controlled variation was available for oleic acid (19.42–36.91% seed oil) and only a very limited range for stearic acid (2.62–6.81% seed oil). Although useful, the NITS models for oleic acid and stearic acid had relatively low coefficients of determination. As they did not include samples with elevated stearic acid content or extremely elevated oleic acid content (>50%), the NITS models may have greater error as compared to our NIRS-based models when predicting samples that meet current market demands.

Our NIRS calibrations performed best in predicting C18:0, C18:1, C18:2, and C18:3, and the RPD values suggested that these models can be used to predict these fatty acids in bulk, intact soybean seeds. Palmitic acid (C16:0) predictions had the lowest RPD values, which suggested that the calibration could be used to group or classify seeds as high or low relative values for the trait, though the specific percentage reported is likely to have some associated error. We observed a lack of linear response at the extremely high end for C18:1 (>75% seed oil), which limited the ability to differentiate samples via NIRS. A linear response for C18:1 was observed below 75%.

The SEPs of the external validation set for most of the fatty acids were slightly higher than the SECs, which could be due to fewer seeds in the samples and/or error associated with GC measurement. RMSECV provides an estimate of the magnitude of error expected in independent samples using the calibration model; a reliable model should have a lower RMSEP as compared to its RMSECV. Finally, a paired *t* test revealed that the fatty acid composition data obtained from the GC and NIRS methods were not statistically different ($P > 0.05$) for all comparisons (Table 3).

Conclusion

The goal of this study was to develop NIRS calibration models for the five major fatty acids in soybean, and we demonstrated that NIRS can be used to phenotype the major fatty acids in bulk intact seeds in a high-throughput manner, which will assist in accelerating soybean breeding programs. The robustness and accuracy of predicting fatty acid composition using the NIRS models was confirmed by means of both CV and external validation. In four of the five fatty acids examined, the *r* value was greater than 0.90. The *r* value for the calibration model of palmitic acid was slightly lower than the rest of the fatty acid species at 0.82. Phenotyping individual fatty acids in bulk, intact seeds using NIRS does not require extensive sample processing or the use of harsh chemicals and solvents. As a result, sample analysis can be fully automated, non-destructive, and non-toxic as compared to chemical analytical methods. Using the NIRS method, accurate selections and/or predictions of soybean seeds with favorable fatty acid composition can be made directly, rapidly and efficiently. It has been proven possible to readily transfer calibrations between instruments with similar optical and mechanical characteristics [11, 29], and we anticipate the calibrations developed in this study can be readily transferred. These methods can be used regardless of the specific biotechnological or mutational fatty acid genetic alteration(s). One intriguing possibility is the incorporation of these NIRS calibrations after the breeding process has concluded; i.e. when farmers deposit seeds at an elevator. Although not previously practical, the development of rapid, in-line NIRS (reviewed in [30]) opens the possibility that grain elevators could directly ensure seed purity, provide incentives for farmers who choose to grow value-added seed, and potentially introduce constituent-based pricing for soybean seed, rather than on a dry weight basis.

Acknowledgements We acknowledge the excellent technical assistance of two USDA-ARS technicians Alexandria Berghaus and Jeremy Mullis who were primarily responsible for all field work and gas chromatography in support of this work.

Compliance with Ethical Standards

Conflict of interest The authors declare that they have no pertinent conflict of interest related to this study, which was partially supported by a grant from United Soybean Board (Project No. 1420-632-6605), and by USDA-Agricultural Research Service internal funding. Mention of any trademarks, vendors, or proprietary products does not constitute a guarantee or warranty of the product by the USDA and does not imply its approval to the exclusion of other products or vendors that may also be suitable. The USDA, Agricultural Research Service, Midwest Area, is an equal opportunity, affirmative action employer and all agency services are available without discrimination.

References

1. Rotundo JL, Westgate ME (2009) Meta-analysis of environmental effects on soybean seed composition. Field Crop Res 110:147–156
2. Harwood JL (1980) 1—Plant acyl lipids: structure, distribution, and analysis A2. In: Stumpf PK (ed) Lipids: structure and function. Academic Press, New York, pp 1–55
3. Wilson RF (2004) Seed composition. In: Boerma HR, Specht J (eds) Soybeans: improvement, production, and uses. American Society of Agronomy, Madison, pp 621–677
4. Remig V, Franklin B, Margolis S, Kostas G, Nece T, Street JC (2010) Trans fats in America: a review of their use, consumption, health implications, and regulation. J Acad Nutr Diet. 110:585–592
5. Hunter JE, Zhang J, Kris-Etherton PM (2010) Cardiovascular disease risk of dietary stearic acid compared with trans, other

saturated, and unsaturated fatty acids: a systematic review. Am J Clin Nutr 91:46–63
6. Yu S, Derr J, Etherton TD, Kris-Etherton PM (1995) Plasma cholesterol-predictive equations demonstrate that stearic acid is neutral and monounsaturated fatty acids are hypocholesterolemic. Am J Clin Nutr 61:1129–1139
7. Gillman JD, Bilyeu K (2012) Genes and alleles for quality traits on the soybean genetic/physical map. In: Willson RF (ed) Designing soybeans for 21st century markets. AOCS Press, Urbana, pp 67–96
8. Medic J, Atkinson C, Hurburgh CR (2014) Current knowledge in soybean composition. J Am Oil Chem Soc 91:363–384
9. Osborne BG (2006) Near-Infrared Spectroscopy in food analysis. Encyclopedia of analytical chemistry. Wiley, New York
10. Williams P (2007) Grains and seeds. In: Ozaki Y, McClure WF, Christy AA (eds) Near-infrared spectroscopy in food science and technology. Wiley, New York, pp 165–217
11. Hacisalihoglu G, Gustin JL, Louisma J, Armstrong P, Peter GF, Walker AR, Settles AM (2016) Enhanced single seed trait predictions in soybean (*Glycine max*) and robust calibration model transfer with near-infrared reflectance spectroscopy. J Agric Food Chem 64:1079–1086
12. Orman BA, Schumann RA (1991) Comparison of near-infrared spectroscopy calibration methods for the prediction of protein, oil, and starch in maize grain. J Agric Food Chem 39:883–886
13. Han S-I, Chae J-H, Bilyeu K, Shannon JG, Lee J-D (2014) Nondestructive determination of high oleic acid content in single soybean seeds by near infrared reflectance spectroscopy. J Am Oil Chem Soc 91:229–234
14. Pazdernik DL, Killam AS, Orf JH (1997) Analysis of amino and fatty acid composition in soybean seed, using near infrared reflectance spectroscopy. Agron J 89:679–685
15. Roberts CA, Ren C, Beuselinck PR, Benedict HR, Bilyeu K (2006) Fatty acid profiling of soybean cotyledons by near-infrared spectroscopy. Appl Spectrosc 60:1328–1333
16. Patil AG, Oak MD, Taware SP, Tamhankar SA, Rao VS (2010) Nondestructive estimation of fatty acid composition in soybean [*Glycine max* (L.) Merrill] seeds using near-infrared transmittance spectroscopy. Food Chem 120:1210–1217
17. Tillman BL, Gorbet DW, Person G (2006) Predicting oleic and linoleic acid content of single peanut seeds using near-infrared reflectance spectroscopy. Crop Sci 46:2121–2126
18. Hammond EG, Fehr WR (1983) Registration of A6 germplasm line of soybean. Crop Sci 23:192–193
19. Gillman JD, Stacey MG, Cui Y, Berg HR, Stacey G (2014) Deletions of the SACPD-C locus elevate seed stearic acid levels but also result in fatty acid and morphological alterations in nitrogen fixing nodules. BMC Plant Biol 14:143
20. Cantor SL, Hoag SW, Ellison CD, Khan MA, Lyon RC (2011) NIR Spectroscopy applications in the development of a compacted multiparticulate system for modified release. AAPS PharmSciTech 12:262–278
21. Martens M, Martens H (1986) Partial least squares regression. Elsevier Applied Science, London
22. Brown SD (1995) The Unscrambler®, Version 5.5. Multivariate analysis software for MS-DOS. J Chemom 9:527–529
23. Geladi P, MacDougall D, Martens H (1985) Linearization and scatter-correction for near-infrared reflectance spectra of meat. Appl Spectrosc 39:491–500
24. Spielbauer G, Armstrong P, Baier JW, Allen WB, Richardson K, Shen B, Settles AM (2009) High-throughput near-infrared reflectance spectroscopy for predicting quantitative and qualitative composition phenotypes of individual maize kernels. Cereal Chem 86:556–564
25. Baye TM, Pearson TC, Settles AM (2006) Development of a calibration to predict maize seed composition using single kernel near infrared spectroscopy. J Cereal Sci 43:236–243
26. Martens H, Naes T (1989) Assessment, validation and choice of calibration method. Wiley, New York
27. Williams PC, Norris KH (2001) Near-infrared technology in the agricultural and food industries. American Association of Cereal Chemists, Saint Paul
28. Igne B, Rippke GR, Hurburgh CR (2008) Measurement of whole soybean fatty acids by Near Infrared Spectroscopy. J Am Oil Chem Soc 85:1105–1113
29. Igne B, Roger J-M, Roussel S, Bellon-Maurel V, Hurburgh CR (2009) Improving the transfer of near infrared prediction models by orthogonal methods. Chemometr Intel Lab 99:57–65
30. Porep JU, Kammerer DR, Carle R (2015) On-line application of near infrared (NIR) spectroscopy in food production. Trends Food Sci Tech. 46:211–230