Development of Calibration Equations to Predict Oil Content and Fatty Acid Composition in Brassicaceae Germplasm by Near-Infrared Reflectance Spectroscopy

Leonardo Velasco*, Fernando D. Goffman, and Heiko C. Becker

Institut für Pflanzenbau und Pflanzenzüchtung, Georg-August-Universität, D-37075 Göttingen, Germany

ABSTRACT: The use of nondestructive analytical methods is critical for the evaluation of very small seed samples such as those from germplasm collections. The objective of this study was to evaluate the potential of near-infrared reflectance spectroscopy (NIRS) for the simultaneous analysis of seed oil content and concentration of major fatty acids in intact-seed samples of the family Brassicaceae. A total of 495 samples from 56 genera and 128 species were analyzed by NIRS. The fatty acid composition of the seed oil was determined in all the samples by gas–liquid chromatography (GLC). The total seed oil content was determined by solvent extraction in 129 samples from 22 genera. Calibration equations for oil content (n = 97) and individual fatty acids (n = 410) were developed and tested through external validation with the samples not included in the calibration sets. The calibration equations for oil content (r^2 = 0.97 in validation) and concentrations of $C_{18:1}$ (r^2 = 0.93), $C_{18:3}$ (r^2 = 0.95), and $C_{22:1}$ (r^2 = 0.94) showed very good performance and provided reliable estimations of these traits in the samples of the validation set. The calibration equations for $C_{16:0}$, $C_{18:0}$, and $C_{18:2}$ content were less reliable, with r^2 from 0.67 to 0.73. There was practically no response of NIRS to differences in $C_{20:1}$ (r^2 = 0.31). These results demonstrated that the oil content and concentrations of $C_{18:1}$, $C_{18:3}$, and $C_{22:1}$ can be estimated reliably within the family Brassicaceae by using NIRS calibration equations integrating broad taxonomic variability.

Paper no. J8886 in *JAOCS 76,* 25–30 (January 1999).

KEY WORDS: *Brassicaceae,* crucifers, fatty acid composition, intact seeds, near-infrared reflectance spectroscopy, NIRS, oil content.

The family Brassicaceae (synonymous with Cruciferae, common name crucifers) has traditionally attracted the attention of plant breeders and oil chemists because of its high seed oil content and special fatty acid composition. The oil content of wild crucifers ranges from about 5 to 50% (1–4). Two main fatty acid patterns are present in the family (1–6). The first one is characterized by linolenic acid ($C_{18:3}$) as the predominant fatty acid, accounting for up to more than 70% of the total fatty acids (e.g., *Matthiola* and *Malcomia*). The second pattern shows high levels of erucic acid ($C_{22:1}$), up to about 60% of the total fatty acids (e.g., *Brassica* and *Crambe*). There is a wide variation between both extreme patterns. Additionally, some genera are characterized by high levels of other fatty acids such as eicosenoic ($C_{20:1}$, *Leavenworthia*), nervonic ($C_{24:1}$, *Lunaria*), monohydroxy (*Lesquerella*), and dihydroxy (*Cardamine*) (2). Within this family, several species of *Brassica* (rapeseed and mustards) are cultivated on a large scale for oil production, representing the world's third-most important source of vegetable oil. The oil is used mainly as edible oil but it also has important industrial applications (7).

Detrimental environmental effects of intensive agriculture and the situation of overproduction in industrialized countries are leading to an increasing interest in alternative industrial oil crops, which should be better adapted to sustainable agricultural production than current major oilseed crops (8). Moreover, recent developments in biotechnology make it possible to transfer interesting genes from wild Brassicaceae species to the major *Brassica* crops either by protoplast fusion or by molecular gene transfer. Within this context, the evaluation of the available germplasm of underexploited cruciferous crops and wild forms of this family is an interesting task for the plant breeder. The evaluation of germplasm accessions for oil content and composition is frequently limited by the small number of seeds available. Although analytical techniques such as gas–liquid chromatography (GLC) of fatty acid methyl esters allow the simultaneous determination of oil content and its fatty acid composition in very small samples, the use of a representative sample requires the destruction of a high number of seeds. In *Brassica* species, recent studies have demonstrated that near-infrared reflectance spectroscopy (NIRS) may provide a reliable estimation of the fatty acid composition of the seed oil of intact seeds, with the main advantages that the analysis is nondestructive, simultaneous with the analysis of other seed constituents such as oil, protein, and glucosinolate content, and suitable for small samples (9,10).

The characteristics of NIRS technique make it optimal for the nondestructive evaluation of oil content and composition in germplasm accessions of cruciferous seeds. However, it would require the development of calibration equations integrating large taxonomic variation, since the development of specific equations for single species or even for single genera would be impracticable. The objective of this study was to evaluate the potential of NIRS to estimate the oil content and

*To whom correspondence should be addressed at Institute for Sustainable Agriculture (CSIC), Apartado 4084, E-14080 Cordoba, Spain.
E-mail: ia2veval@uco.es

TABLE 1
Genera, Number of Species, and Number of Samples of Brassicaceae Used in the Development and Validation of NIRS Calibration Equations[a]

Genus	# Species	# Samples	Genus	# Species	# Samples
Aethionema	2	3	*Fibigia*	1	4
Alyssoides	1	1	*Heliophila*	1	1
Alyssum	7	9	*Hesperis*	2	8
Alliaria	1	6	*Hirschfeldia*	1	1
Arabis	8	8	*Hornungia*	1	1
Aubrieta	1	1	*Hugueninia*	1	1
Barbarea	3	6	*Hutchinsia*	1	1
Besteroa	1	1	*Iberis*	2	4
Biscutella	1	1	*Kernera*	1	1
Brassica	17	254	*Lepidium*	5	12
Bunias	2	16	*Lobularia*	1	1
Calepina	1	1	*Lunaria*	2	8
Camelina	2	9	*Malcomia*	3	5
Capsella	1	1	*Matthiola*	2	2
Cardamine	2	3	*Moricandia*	1	1
Cardaminopsis	1	1	*Myagrum*	1	1
Cheiranthus	1	1	*Nasturtium*	1	2
Clypeeola	1	1	*Neslia*	1	2
Cochlearia	3	5	*Peltaria*	1	1
Conringia	1	6	*Raphanus*	2	13
Coronopus	2	14	*Rapistrum*	2	2
Crambe	3	6	*Rorippa*	2	2
Descurainia	1	1	*Shivereckia*	1	1
Diplotaxis	5	8	*Sinapis*	2	23
Draba	3	3	*Sisymbrium*	5	9
Eruca	1	3	*Succowia*	1	1
Erucastrum	2	4	*Thlaspi*	5	9
Erysimum	5	5			

[a]NIRS, near-infrared reflectance spectroscopy.

its fatty acid composition in a germplasm collection of *Brassicaceae* through the development of calibration equations integrating large taxonomic variability.

MATERIALS AND METHODS

Samples. The study was carried out with 495 intact-seed samples belonging to 56 genera and 128 species of the family Brassicaceae. Most of the samples were collected from germplasm banks worldwide. Thirty samples of *Brassica napus* were selected from breeding material showing genetically modified fatty acid composition of the seed oil. Table 1 shows the genera included in the study, the number of species, and the total number of samples of each genus.

Analyses by reference methods. The seed oil content was determined on milled samples of about 200 mg by Soxhlet oil extraction with petrol ether. The fatty acid composition of the seed oil was determined by GLC of fatty acid methyl esters. They were prepared following the procedure developed by Thies (11) and analyzed on a Perkin-Elmer gas chromatograph model 8600 (Perkin-Elmer Corporation, Norwalk, CT) equipped with a fused-silica capillary column FFAP, 25 m × 0.25 mm × 0.25 μm film thickness (Macherey & Nagel GmbH + Co. Kg, Düren, Germany). The oven, detector, and injector temperatures were 200, 250, and 250°C, respectively. The carrier gas was nitrogen at a pressure of 100 kPa. Two microliters of sample was injected at a split rate of 1:70. Identification of the fatty acids $C_{16:0}$, $C_{18:0}$, $C_{18:1}$, $C_{18:2}$, $C_{18:3}$, $C_{20:1}$, $C_{22:1}$, and $C_{24:1}$ was performed by comparison of retention times with standards. Individual fatty acids were expressed as percentages of the total fatty acids, also including other minor fatty acids.

All 495 samples included in this study were analyzed for fatty acid composition. However, only 129 of them, belonging to 22 different genera, were analyzed for total seed oil content because of the limiting availability of seeds.

NIRS method. The samples were scanned on a monochromator NIR Systems model 6500 (NIR Systems, Inc., Silver Spring, MD) equipped with sample autochanger. The standard ring cup (Ø 4.7 cm), which requires a seed volume of about 6 cm^3, was fitted with a special adapter for small samples (Ø 1.5 cm), which reduced the sample volume to about 0.6 cm^3. For each sample the reflectance spectrum (log 1/*R*) from 400 to 2500 nm was recorded at 2-nm intervals. Calibration and validation procedures were carried out with ISI software, version 3.10 (Infrasoft International, Port Matilda, PA). Original reflectance spectra were corrected prior to calibration by applying a second derivative transformation, a standard normal variate transformation, and a de-trend scatter correction. The second derivative was calculated from the log

TABLE 2
Number of Samples Analyzed, Mean, Standard Deviation, and Range of Oil Content (%) and Major Fatty Acids (% of total fatty acids), with Indication of the Species to Which the Samples with the Minimum and Maximum Values Belong

Trait	#	Mean	SD[a]	Min	Species	Max	Species
Oil	129	30.1	9.3	8.3	*Bunias orientalis*	50.8	*Brassica napus*[b]
$C_{16:0}$	495	4.9	2.5	1.5	*Crambe abyssinica*	14.3	*Fibigia clypeata*
$C_{18:0}$	495	1.3	0.7	0.2	*Lunaria annua*	4.3	*Lobularia maritima*
$C_{18:1}$	495	17.9	12.5	5.4	*Sisymbrium loeselii*	80.8	*Brassica napus*[b]
$C_{18:2}$	495	18.0	6.4	5.4	*Brassica napus*[b]	45.5	*Brassica napus*[b]
$C_{18:3}$	495	19.7	15.4	1.0	*Lunaria annua*	69.8	*Alyssum alyssoides*
$C_{20:1}$	495	7.0	4.8	0.0	Several	36.1	*Biscutella auriculata*
$C_{22:1}$	495	26.3	18.1	0.0	Several	60.2	*Crambe abyssinica*
$C_{24:1}$	495	1.8	2.6	0.0	Several	23.4	*Lunaria annua*

[a]SD, standard deviation.
[b]Entries selected from breeding material.

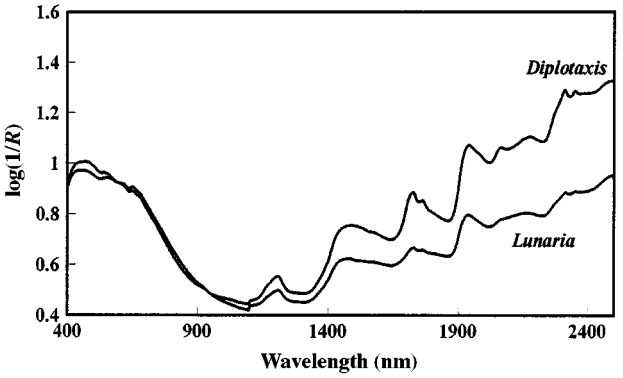

FIG. 1. Average spectra of the samples belonging to the genera *Diplotaxis* and *Lunaria,* which are characterized by very different seed types.

1/*R* spectra every 10 nm, with smoothing over segments of 10 nm in length. The original sets (n = 495 for individual fatty acids and n = 129 for oil content) were randomly divided into a calibration and a validation set. The number of samples included in the validation set was 32 for oil content and 85 for individual fatty acids. In both cases the validation set contained samples from genera that were not present in the calibration set. These genera were *Bunias* and *Moricandia* in the validation set for oil content and *Aubrieta*, *Capsella*, *Cheiranthus*, *Descurainia*, *Hugueninia*, *Hutchinsia,* and *Peltaria* in the validation set for fatty acids. Calibration equations were developed by using the spectral information from 1100 to 2500 nm and modified partial least squares (MPLS) regression. They were tested through external validation.

RESULTS AND DISCUSSION

The entries in the germplasm collection showed a similar variability for seed oil content and fatty acid composition to that reported to date within the family Brassicaceae (Table 2). Oil content ranged from about 8% in an entry of *Bunias orientalis* to more than 50% in breeding material of *B. napus*. The sum of the fatty acids listed in Table 2 ranged from 91% of the total fatty acids to more than 99%. The range of variation for the major fatty acids was very similar to that summarized by Kumar and Tsunoda (4), excluding the extreme values of $C_{18:2}$ and the highest levels of $C_{18:1}$, which were found in breeding material of *B. napus*. Only species with very high $C_{20:1}$ content (>50%), which have been reported previously in this family (4), were not included. The maximum $C_{20:1}$ content (36%) corresponded to an entry of *Biscutella auriculata* (Table 2). The large variability in the calibration set concerned not only oil content and fatty acid composition but also seed color, size, texture and shape, resulting in large differences in NIRS spectra. Figure 1 shows, as an example, the average spectra of the samples belonging to *Diplotaxis* and *Lunaria*, which are characterized by very different seed types.

Table 3 shows the calibration and external validation statistics of the calibration equations for seed oil content and concentration of the major fatty acids. The calibration equation for oil content showed a high accuracy in the estimation of this trait, with an r^2 of 0.97 in external validation and standard error of performance (SEP) of 1.92%. This SEP was higher by about 0.6% than that obtained in calibration equations for oil content in single species such as *B. napus* (12,13). The reduction of accuracy associated with the integration of large variability in the calibration set is a well-known effect in NIRS analysis (14). Figure 2A shows the validation plot for this trait. Two of the samples of the validation set, indicated with asterisks, corresponded to genera not included in the calibration set (*Bunias* and *Moricandia*). The performance of the calibration equation in the prediction of oil content in both samples was similar to the other samples from genera represented in the calibration.

The results obtained in the development of calibration equations for predicting the concentration of individual fatty acids were very heterogeneous (Table 3). The calibration equations for $C_{18:1}$, $C_{18:3}$, and $C_{22:1}$ showed a high r^2 in external validation, from 0.93 to 0.96 (Fig. 2), indicating that a reliable estima-

TABLE 3
Calibration and Validation Statistics of NIRS Calibration Equations for the Analysis of Oil Content and Major Fatty Acids in Intact-Seed Samples of Brassicaceae

	Calibration					External validation				
Trait	#	Mean	SD[a]	SEC[a]	r^2	#	Mean	SD	SEP[a]	r^2
Oil	97	30.80	8.85	1.85	0.96	32	28.48	10.51	1.92	0.97
$C_{16:0}$	410	4.88	2.45	1.19	0.76	85	5.23	2.82	1.47	0.73
$C_{18:0}$	410	1.32	0.68	0.38	0.69	85	1.37	0.75	0.43	0.67
$C_{18:1}$	410	17.60	12.31	3.79	0.90	85	22.10	16.19	4.33	0.93
$C_{18:2}$	410	18.29	6.37	2.95	0.78	85	17.46	6.83	3.70	0.70
$C_{18:3}$	410	19.35	15.19	3.02	0.96	85	21.25	16.26	3.44	0.95
$C_{20:1}$	410	7.00	4.69	3.73	0.37	85	7.12	4.34	3.61	0.31
$C_{22:1}$	410	26.76	18.07	4.33	0.94	85	24.61	18.25	4.54	0.94
$C_{24:1}$	410	1.82	2.53	0.96	0.85	85	1.58	2.81	1.54	0.70[b]

[a]SD, standard deviation; SEC, standard error of calibration; SEP, standard error of performance. For other abbreviation see Table 1.
[b]The number of samples with intermediate and high values (>5%) was very low (see Fig. 2 part I).

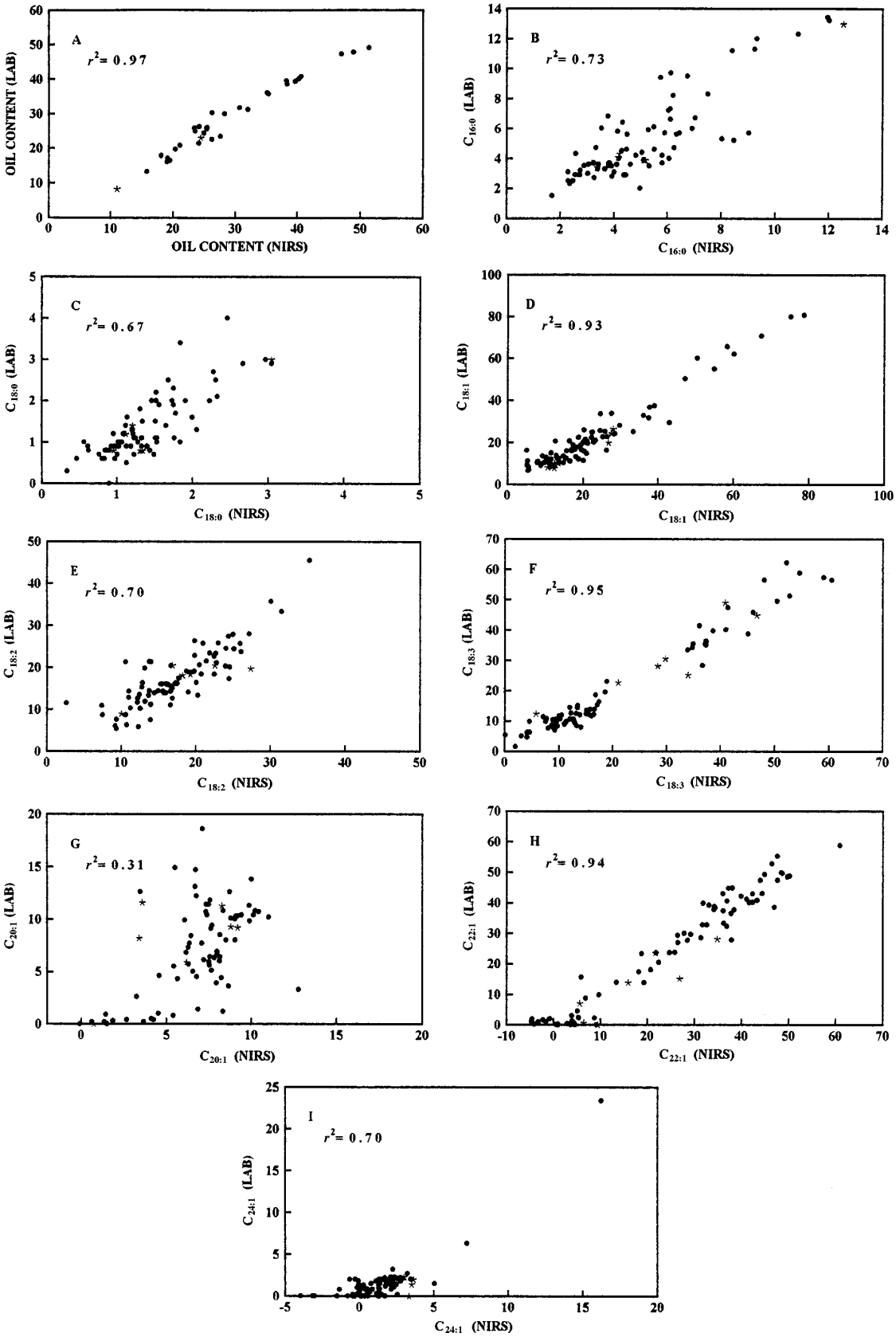

FIG. 2. External validation plots for oil content, expressed as % seed weight, and individual fatty acids, expressed as % of the total fatty acids. The samples represented with asterisks belong to genera that were not included in the calibration set.

tion of the concentration of these fatty acids can be obtained in samples of crucifers by applying these equations. These values are similar to those reported for calibration equations integrating samples from different species of *Brassica* (15). The calibration equations for the fatty acids $C_{16:0}$, $C_{18:0}$, and $C_{18:2}$ were characterized by poorer validation statistics, with r^2 between 0.67 and 0.73. The validation plots shown in Figure 2 suggest that the calibration equations for these three fatty acids could be useful only to make preliminary selections or to get an approximate classification of the entries; the SEP was too high (Table 3) to consider them as adequate for reliable analyses. Previous studies developing either single- (9,10) or multispecies calibration equations (15) for these fatty acids in the genus *Brassica* reported considerably lower SEP values. For example, Velasco *et al.* (15) obtained a SEP of 2.25% (mean 16.6%) for $C_{18:2}$ in a multi-species calibration equation, as compared with a SEP of 3.7% (mean 17.5%) obtained in this study.

NIRS estimations of $C_{20:1}$ content were scarcely correlated with GLC data (Fig. 2G), showing an r^2 of 0.31. The low correlation between NIRS and GLC for $C_{20:1}$ content in species of *Brassica* has been reported previously. Velasco *et al.* (10) explained this on the basis of the particular relationship between $C_{20:1}$ and $C_{22:1}$, also described by Appelqvist (16). The calibration equation for $C_{24:1}$ was very promising, showing an r^2 of 0.85 between NIRS and GLC values in the calibration set (Table 3). Nevertheless, it could not be adequately validated because of the low number of samples with $C_{24:1}$ content of above 5%. Two of these samples were included in the validation set and their $C_{24:1}$ content was correctly identified, although more entries are needed to have a more complete validation of the equation.

This study included intact-seed samples from 56 different genera of Brassicaceae. The calibration equations for fatty acids included samples from 49 of them, while the samples of the other seven genera were included in the validation set. As shown in Figure 2 (samples represented with asterisks), NIRS performance in the prediction of the fatty acid composition of these samples was similar to that shown in the prediction of samples from genera represented in the calibration. Although the number of samples was rather low, the calibration equations may have a similar accuracy to that reported here in the analysis of future entries of Brassicaceae, even of those taxonomically different.

The calibration equations developed in this study will be very useful for future evaluations of germplasm accessions of crucifers, but some limitations also have to be taken into account. For example, the performance of NIRS was very good for the analysis of oil content and $C_{18:1}$, $C_{18:3}$, and $C_{22:1}$ content, but not for the other fatty acids. Furthermore, some fatty acids, such as monohydroxy and dihydroxy fatty acids that can be found in species of this family (2), were not included in the study. Taking into account these limitations, the application of NIRS technique should be limited to a preliminary screening of the germplasm. Since $C_{18:3}$ and $C_{22:1}$ are the most abundant fatty acids in the family, NIRS screening would allow a high percentage of the samples to be characterized, while other samples should be further analyzed by GLC.

The best performance in calibration equations for individual fatty acids corresponded to those fatty acids for which the variability in the calibration set was higher: $C_{18:1}$ [standard deviation (SD) = 12.3%], $C_{18:3}$ (SD = 15.2%), and $C_{22:1}$ (SD = 18.1%) (Table 3). This suggests that the variability for fatty acid composition in the calibration set is one of the main factors determining successful calibration equations. Because additional variability has not been reported within the family Brassicaceae for some of the other fatty acids such as $C_{16:0}$, $C_{18:0}$, and $C_{18:2}$, it would be interesting to evaluate the performance of calibration equations integrating wider variability by using samples from other botanical families.

This study demonstrated that the use of calibration equations integrating large taxonomic diversity is an effective method for the evaluation of seed oil content and composition in intact-seed samples of the family Brassicaceae. The accuracy of such estimations is lower than that reported in single-species and in single-genus (*Brassica*) calibration equations. Therefore, this approach is advisable only for the evaluation of germplasm collections including large taxonomic diversity, where the development of single-species or single-genus calibrations is impracticable. For individual fatty acids, highly reliable calibration equations were developed for $C_{18:1}$, $C_{18:3}$, and $C_{22:1}$. The use of further variability, even by including samples from other families, might lead to the development of more accurate calibration equations for other fatty acids.

ACKNOWLEDGMENTS

Christine Reuter and Uwe Ammermann provided technical assistance in GLC analyses. L. Velasco received a grant from the Dirección General de Enseñanza Superior (Ministerio de Educación y Cultura, Spain). F.D. Goffman was funded by the Deutscher Akademischer Austauschdienst (DAAD, Germany).

REFERENCES

1. Mikolajczak, K.L., T.K. Miwa, F.R. Earle, I.A. Wolff, and Q. Jones, Search for New Industrial Oils. V. Oils of Cruciferae, *J. Am. Oil Chem. Soc. 38:*678–681 (1961).
2. Miller, R.W., F.R. Earle, and I.A. Wolff, Search for New Industrial Oils. XIII. Oils from 102 Species of Cruciferae, *Ibid. 42:*817–821 (1965).
3. Kumar, P.R., and S. Tsunoda, Fatty Acid Spectrum of Mediterranean Wild Cruciferae, *Ibid. 55:*320–323 (1978).
4. Kumar, P.R., and S. Tsunoda, Variation in Oil Content and Fatty Acid Composition Among Seeds from the Cruciferae, in *Brassica Crops and Wild Allies,* edited by S. Tsunoda, K. Hinata, and C. Gómez-Campo, Japan Scientific Societies Press, Tokyo, 1980, pp. 235–252.
5. Appelqvist, L.-A., Lipids in Cruciferae: VIII. The Fatty Acid Composition of Seeds of Some Wild or Partially Domesticated Species, *J. Am. Oil Chem. Soc. 48:*740–744 (1971).
6. Jart, A., The Fatty Acid Composition of Various Cruciferous Seeds, *Ibid. 55:*873–875 (1978).
7. Friedt, W., and W. Lühs, Development in the Breeding of Rapeseed Oil for Industrial Purposes, in *Proceedings of the 9th International Rapeseed Congress*, Cambridge, United Kingdom, 4–7 July 1995, edited by GCIRC, Henry Ling Limited, Dorchester, pp. 437–448.
8. Merker, A., and P. Nilsson, Some Oil Crop Properties in Wild *Barbarea* and *Lepidium* Species, *Swedish J. Agric. Res. 25:*173–178 (1995).

9. Velasco, L., J.M. Fernández-Martínez, and A. De Haro, Determination of the Fatty Acid Composition of the Oil in Intact-Seed Mustard by Near-Infrared Reflectance Spectroscopy, *J. Am. Oil Chem. Soc. 74:*1595–1602 (1997).
10. Velasco, L., and H.C. Becker, Estimating the Fatty Acid Composition of the Oil in Intact-Seed Rapeseed (*Brassica napus* L.) by Near-Infrared Reflectance Spectroscopy, *Euphytica 101:*221–230 (1998).
11. Thies, W, Schnelle und einfache Analysen der Fettsäurezusammensetzung in einzelnen Raps-Kotyledonen I. Gaschromatographische und Papierchromatographische Methoden, *Z. Pflanzenzüchtg. 65:*181–202 (1971).
12. Biston, R., P. Dardenne, M. Cwikowski, J.-P. Wathelet, and M. Severin, Analysis of Quality Parameters of Whole Rapeseed by NIRS, in *Glucosinolates in Rapeseeds: Analytical Aspects*, edited by J.-P. Wathelet, Martinus Nijhoff, Dordrecht, The Netherlands, 1987, pp. 163–172.
13. Daun, J.K., K.M. Clear, and P. Williams, Comparison of Three Whole Seed Near-Infrared Analyzers for Measuring Quality Components of Canola Seed, *J. Am. Oil Chem. Soc. 71:*1063–1068 (1994).
14. Shenk, J.S., and M.O. Westerhaus, Near Infrared Reflectance Analysis with Single- and Multiproduct Calibrations, *Crop Sci. 33:*582–584 (1993).
15. Velasco, L., F.D. Goffman, and H.C. Becker, Variability for the Fatty Acid Composition of the Seed Oil in a Germplasm Collection of the Genus *Brassica*, *Genet. Rcs. Crop Evol. 45:*371–382 (1998).
16. Appelqvist, L.-A., Lipids in Cruciferae. IV. Fatty Acid Patterns in Single Seeds and Seed Populations of Various *Cruciferae* and in Different Tissues of *Brassica napus* L., *Hereditas 61:*9–44 (1969).

[Received May 26, 1998; accepted September 4, 1998]