

构建整粒油菜籽脂肪酸成分近红外反射光谱分析模型的研究

吴建国, 石春海*, 张海珍

浙江大学农业与生物技术学院农学系, 浙江 杭州 310029

摘要 选用具有多年份、多地点、变异大的497份油菜籽育种材料组成原始样品集, 光谱经散射和数学预处理利用改良偏最小二乘法(MPLS)构建各脂肪酸近红外反射光谱(NIRS)校正模型, 同时采用三种不同用量的样品杯进行NIRS建模分析。结果表明, 以8g样品校正建模效果最好, 六种脂肪酸的校正决定系数为0.74~0.98。同时以3和0.6g样品分别发展的校正模型效果也较好, 两者分析效果相近。各项决定系数(RSQ1, 1-VR)高, 相应的各项误差(SEC, SECV)较低。该研究以完整油菜籽为样品所建立的脂肪酸NIRS模型, 可直接用于育种材料选择、突变体筛选和种质资源的评价等研究。

主题词 近红外反射光谱; 油菜籽; 脂肪酸; 校正模型

中图分类号: O657.1 **文献标识码:** A **文章编号:** 1000-0593(2006)02-0259-04

引言

油菜是世界重要的油料作物之一, 脂肪酸成分的改良是油菜优质育种中的重要内容。现有的脂肪酸分析往往采用气相色谱方法, 必须破坏样品种子, 具有分析过程复杂、耗时长等不利因素, 严重影响油菜品质改良的进度。近红外反射光谱(near infrared reflectance spectroscopy, NIRS)技术是近年迅速崛起的一种新颖物理测试技术, 具有测定速度快、无损伤分析等特点^[1-4]。NIRS技术所具有的特点也正是育种所需的, 因此该技术已较早地应用于油菜育种的品质分析。在脂肪酸分析方面, 有学者成功地发展芥酸和油酸含量的NIRS测定技术, 但尚未发展能够同时进行分析各主要脂肪酸含量的NIRS分析模型^[5]。Daun等尝试构建同时分析多种脂肪酸的近红外校正模型, 虽取得一定效果, 但准确性欠佳^[6]。利用突变体材料, 首先在埃塞俄比亚芥中构建能够准确预测和分析各主要脂肪酸的校正模型^[7]; 之后, 利用脂肪酸突变体材料参与校正集建模, 完成了能同时准确分析各主要脂肪酸的整粒分析模型^[8], 但能否用于育种实践有待进一步研究。

目前, 国内在油菜籽NIRS分析领域也取得了一些进展。开展了油菜籽硫甙、蛋白质和油份等常规品质指标^[9, 10]和各脂肪酸^[11]NIRS分析研究, 取得较好效果。但由于样品数少, 模型的稳定性尚须进一步研究。有学者应用多年多点的大量样品, 发展了能准确、稳定分析油菜籽含油量、硫甙

和芥酸品质指标的NIRS分析模型^[12, 13]。李延莉等比较了NIRS法与传统方法测定油菜籽常规品质指标的效果, 结果显示NIRS法测定具有快速、简便、样品用量少等特点, 且两者分析误差相近^[14]。

本研究根据育种实际的需要, 选用亲本、分离群体、高代品系等材料组成NIRS建模校正集, 建立稳定、准确、快速的油菜籽脂肪酸NIRS分析模型, 促进该技术在油菜育种中的应用。

1 材料与amp;方法

1.1 油菜籽样本和脂肪酸的分析

选用遗传源广、脂肪酸含量差异大的甘蓝型油菜品种, 相互杂交构成一个具有156份样品组成的遗传分离群体, 该分离群体经种植两年共有312份样品。外加1999和2000年的育种高代品系和品种材料185份, 共有497份样本用于NIRS建模和检验分析。这些材料的遗传差异大, 信息丰富, 是构建近红外回归模型的理想样品集。

采用气相色谱法进行分析, 使用日本岛津生产的GC-9A气相色谱仪分析, 其原理和方法参见孙传经^[15]以及周永明^[16]的报道。

1.2 近红外光谱收集和样本选择

运用近红外扫描单色器(5000型, NIRSystems, Silver Spring, MD, USA), 在波长为1100~2498nm的范围内采集反射量(reflectance readings, R)。测定时将待测材料装入

收稿日期: 2004-06-06, 修订日期: 2004-10-26

基金项目: 教育部基金项目“高校骨干教师计划”, 浙江省教育厅项目和国际合作项目计划资助

作者简介: 吴建国, 1971年生, 浙江大学农业与生物技术学院农学系副教授, 博士 * 通讯联系人

1/4 方形样品杯后,置于仪器插口;小样品则采用直径为 35 mm 或直径为 18 mm 的圆形样品杯,外加一个接件可置于相应的插口。每隔 2 nm 采集反射强度(R),每份样品重复扫描 32 次,取平均值,并转化为 $\log(1/R)$,贮存于计算机中。

样本选择则是指在某一群体中,选择若干有代表性的样本用于校正(Calibration)。样本选择前通常先进行群体界定,用“集中”程序(CENTER)计算出给定群体内每个样品的谱带与平均谱带的差异,即 Mahalmlobis 距离(或称为全局距离, GH),一般 $GH < 3$ 的样本看作是同一个群体。然后,用选择 SELECT 程序计算出该群体内样品之间的相似性,即邻居距离(neighborhood H, NH),以便在定标时剔除相似的谱带,一方面可以避免在校正时造成过适(over fit),另一方面则能减少测定工作量。“集中”与“选择”程序的工作原理参见 NIR 技术培训手册^[17]。

1.3 光谱数据预处理与回归模型构建

光谱分析时,采用光谱散射处理和数学处理两种方法。本研究中,前者采用趋势变换法(detrending D)和标准正态变量转换(transformation of standard normal variate, SNV)两种方法联合进行散射处理。后者则通过不同的导数(derivative)、波段一间隙(segment-gap)和平滑(smooth)组合进行优化处理。

通过校正和检验两方面来发展 NIRS 模型。采用改良偏最小二乘法(modified PLS, MPLS)发展校正模型;检验则有交叉检验和外部检验两种方法,一般用交叉检验就可以判断所创建校正模型的效果。衡量 NIRS 模型的主要参数有校正决定系数(RSQ1)、校正标准误差(SEC)、交叉检验标准误差

(SECV)和交叉检验决定系数(1-VR)。对于同一样品集所构建的回归模型而言,各类标准误差和差值越小越好,而决定系数越大越好。

以上各光谱预处理方法、数学处理方法和回归统计方法都通过 WinISI V1.04 软件在 PC 兼容机上运行完成。

2 结果与分析

2.1 脂肪酸分析群体

将这一原始群体所有样本的 NIRS 光谱都汇集在一个 NIRS 文件中,使用 CENTER 程序重新定义群体,剔除马氏距离(H 值)大于 3 的样本。利用 SELECTER 程序选择邻居距离(NH)为 0.35 的样品,重新组成一个群体,即校正集,用于进一步构建 NIRS 模型。选择后重组的新群体各脂肪酸组成列于表 1。由于采用 3 种规格不同样品盒,所收集的 NIR 光谱存在少量的差异。用相同邻居距离,其选择的样本数不同,随着样品用量的减少,样品数增大,8 g 样品用量的群体样本数为 298,小样品 3 和 0.6 g 的群体分别有 336 和 386 个样品。尽管样品数有所不同,但所选择的样本组成还是非常一致的。不同群体中各脂肪酸的均值和标准误差非常接近。各脂肪酸的变异范围较大,基本上覆盖了育种和生产中所用材料脂肪酸的变异。各脂肪酸中以油酸含量的变异性最大,其变幅为 12.12%~70.98%。其余几个不饱和脂肪酸的变异也较突出,如二十烷烯酸(0.85%~29.08%)和芥酸(0.01%~51.72%)的变幅也特别明显,这非常有利于校正分析。

Table 1 Reference value of fatty acid constituent of calibration set in rapeseed

Sample size	Stastics parameter	Palmitic C16 : 0	Oleic C18 : 1	Linoleic C18 : 2	Linolenic C18 : 3	Eicosenoic C20 : 1	Erucic C22 : 1
8 g	Min	2.83	12.12	11.60	4.90	0.85	0.01
	Max	6.13	70.98	27.80	16.75	29.08	51.72
	Mean	4.15	40.28	17.33	10.10	8.95	19.82
	SD	0.62	20.05	3.09	1.53	6.18	17.47
	Sample No.	298	298	298	298	298	281
3 g	Min	1.8	12.12	11.33	4.9	0.85	0.01
	Max	6.13	70.98	27.8	15.8	25.17	51.72
	Mean	4.12	39.20	17.16	10.09	9.35	20.65
	SD	0.63	19.95	3.02	1.40	6.12	17.41
	Sample No.	336	336	336	336	336	319
0.6 g	Min	1.8	12.14	11.33	4.9	0.85	0.01
	Max	6.13	70.98	27.8	15.8	25.17	51.72
	Mean	4.13	39.36	17.21	10.12	9.24	20.59
	SD	0.63	20.10	3.06	1.45	6.15	17.46
	Sample No.	386	386	386	386	386	365

2.2 脂肪酸校正模型

表 2 中列出了以不同样品用量获得的 NIRS 校正方程结果。在 8 g 大样品的校正方程中,油酸、亚油酸、亚麻酸和芥酸 NIR 回归方程都表现出较好的稳定性和准确性,校正和检

验的决定系数较高,尤其是油酸和芥酸的回归模型具有很高的 RSQ1 和 1-VR,分别为 0.98, 0.97 和 0.97, 0.96, 其标准误差(SEC 和 SECV)均小。以大样品光谱发展的 NIRS 分析模型,在分析测定一些大量样品如品种或高世代品系时,

使用较为方便。但在实际育种中,许多样本是分离世代的小材料或是套袋得到的一些杂交材料,很难满足大样品的分析

用量。因此,发展小样品用量的 NIR 分析方法对于近红外技术能否在育种实践中推广和应用将起到关键作用。

Table 2 Calibration and cross validation statistics in NIRS equations for percentage of individual fatty acids in the intact rapeseed samples scanned with different sample size

Sample size	Constituent	Calibration			Cross validation	
		Mean	SEC	RSQ1	SECV	1-VR
8 g	C16 : 0	4.14	0.28	0.77	0.30	0.75
	C18 : 1	40.29	3.03	0.98	3.54	0.97
	C18 : 2	17.13	1.16	0.84	1.42	0.76
	C18 : 3	10.08	0.62	0.79	0.71	0.73
	C20 : 1	8.80	3.09	0.74	3.28	0.71
	C22 : 1	19.72	2.87	0.97	3.54	0.96
3 g	C16 : 0	4.11	0.29	0.76	0.30	0.74
	C18 : 1	39.22	3.01	0.98	3.69	0.97
	C18 : 2	17.02	1.31	0.79	1.45	0.75
	C18 : 3	10.07	0.60	0.78	0.69	0.72
	C20 : 1	9.30	2.55	0.82	3.08	0.74
	C22 : 1	20.64	3.23	0.97	3.73	0.95
0.6 g	C16 : 0	4.12	0.28	0.79	0.29	0.78
	C18 : 1	39.56	3.06	0.98	3.48	0.97
	C18 : 2	17.10	1.36	0.78	1.46	0.75
	C18 : 3	10.09	0.67	0.75	0.73	0.70
	C20 : 1	9.06	3.31	0.70	3.47	0.67
	C22 : 1	20.59	3.14	0.97	3.57	0.96

Note: Scatter correction; NSV+D; Mathematical treatment: "2, 4, 4, 1"

对于小样品用量的 NIR 分析模型,表现为与大样品用量的校正效果相近。3 g 样品用量的 NIR 模型的决定系数与大样品量的模型相同,各类误差大小也相近。在更小的样品用量(0.6 g)构建的校正方程,各类决定系数(RSQ1 和 1-VR)与前面两种一样高,如棕榈酸、油酸、亚油酸和芥酸的检验决定系数分别为 0.78, 0.97, 0.75, 0.70 和 0.96,误差与前两种样品用量的回归模型相仿。仅二十烷烯酸的模型效果略差,但其校正和检验误差均较小。油菜籽的品质性状随着生产和栽培条件的变化,也会发生一定的波动。因此,环境的差异往往会给近红外品质分析模型带来一些误差,影响模型分析的准确度。在本试验中,校正组包括了 3 个不同年份和 2 个地点的样本,其定标模型具有较大的适用范围。而且本模型具有遗传分离世代,与育种中间材料的状态相仿,适宜于进行育种中未纯合材料的分析和筛选。

3 讨 论

种子是一种特殊的农产品,种子品质受年份等环境条件的影响较大,不同年份的气候、栽培条件、管理方法等都会影响种子的品质组分。在确定种子校正集时,要考虑年份的因素,尽可能的收集多年的样品参与校正集发展 NIR 建模。徐广通等(2001)提出将因子分析与 Mahalanobis 距离相结合

的方法来判断定量模型适用性,为样品的准确分析提供依据^[18]。而 WinISI 提出的有关 GH 和 NH 理念,非常有利于模型的调整和扩充。当增加的样品与校正集具有 $GH < 3$ 且 $NH > 0.35$ 的关系,则该类样品加入到校正集时,原校正模型预测的适用性及准确性会进一步地提高。

本研究在收集多年份、多产地的大量样品基础上,系统地开展了构建油菜籽各项品质指标 NIR 校正模型的研究。旨在创建油菜籽各品质性状 NIR 校正分析模型,发展适合育种筛选的 NIR 快速、高效的测定技术;并研究了样品用量、样品年份等因素对建立 NIR 分析模型的影响以及优化建立校正模型的条件。样品的来源对近红外分析具有很重要的作用,不同年份间存在较大差异,需要尽可能收集多年份、多产地、遗传变异大的样品。在本研究中,选用的样品脂肪酸成分变异范围大,基本上覆盖了生产上应用的甘蓝型油菜品种和育种中间材料的品质分析范围。因此,本研究所创建的油菜籽脂肪酸成分 NIR 分析模型,具有稳定性好、分析准确度高、适用性广、实用性强等特点。尤其是所创建的小样本(0.6 g)NIR 分析模型,相应决定系数高,标准误差小,可用于育种早代分离世代材料的选择、突变体库和种质资源筛选以及遗传分析等研究。

致谢:浙江省农科院作核所的张冬青和张尧锋两位老师在脂肪酸测定方面提供了帮助,在此深表谢意。

参 考 文 献

- [1] LI Qin-chen, WANG Wang-zheng, ZHANG Yu-liang, et al. (李庆春, 王文真, 张玉良, 等). *Acta Agronomica Sinica*(作物学报), 1992, 18(3): 235.
- [2] JIN Tong-ming, CUI Hong-chang(金同铭, 崔洪昌). *Acta Agriculturae Boreali-Sinica*(华北农学报), 1994, 9(4): 104.
- [3] Batten G D. *Australian Journal of Experimental Ariculture*, 1998, 38: 697.
- [4] CHU Xiao-li, YUAN Hong-fu, WANG Yan-bin, LU Wan-zhen(褚小立, 袁洪福, 王艳斌, 陆婉珍). *Spectroscopy and Spectral Analysis* (光谱学与光谱分析), 2004, 24(6): 666.
- [5] Rernhardt T D, Röbbelen G. In: GCIRC (Ed.), *Proc 8th Int Rapeseed Congr*, Saskatoon, Canada, 1991. 1380.
- [6] Daun J K, Clear K M, Williams P. *Journal of American Oil Chemists Society*, 1994, 71: 1063.
- [7] Velasco L, Fernandez-Martinez J M, Haro A D. *Crop Science*, 1996, 36: 1068.
- [8] Velasco L, Becker H C. *Euphytica*, 1998, 101: 221.
- [9] GU Wei-zhu, WANG Yan-xiang(顾伟珠, 汪延祥). *Journal of the Chinese Cereals and Oils Association*(中国粮油学报), 1995, 10(2): 57.
- [10] ZHANG Ye-hui, ZHAO Long-lian, LI Xiao-wei, et al. (张晔晔, 赵龙莲, 李晓薇, 等). *Acta Laser Biology Sinica*(激光生物学报), 1998, 7(2): 138.
- [11] CHENG Wen-jie, TAN Xiao-li, WANG Zhu-yun, et al(陈文杰, 谭小力, 王竹云, 等). *Shanxi Agriculture Science*(陕西农业科学), 2002, (8): 6.
- [12] WU Jian-guo, SHI Chun-hai, ZHANG Hai-zheng, et al(吴建国, 石春海, 张海珍, 等). *Acta Agronomica Sinica*(作物学报), 2002, 28(3): 421.
- [13] WU Jian-guo, SHI Chun-hai, FAN Long-jiang(吴建国, 石春海, 樊龙江). *Journal of the Chinese Cereals and Oils Association*(中国粮油学报), 2002, 17(2): 59.
- [14] LI Yan-li, SUN Chao-cai, QIAN Xiao-fang, et al(李延莉, 孙超才, 钱小芳, 等). *Acta Agriculturae Shanghai*(上海农业学报), 2003, 19(1): 11.
- [15] SUN Chuan-jing(孙传经). *Gas Chromatography Mechanism and Technique*(气相色谱分析原理与技术). Beijing: Chemical Industry Press(北京: 化学工业出版社), 1979.
- [16] ZHOU Rong-ming, LIU Hou-li(周永明, 刘后利). *Acta Agronomica Sinica*(作物学报), 1987, 13: 1.
- [17] Shenk J S, Westerhaus M O. *Infrasoft International*, Port Matilda, PA, USA, 1993.
- [18] XU Guang-tong, YUAN Hong-fu, LU Wan-zhen(徐广通, 袁洪福, 陆婉珍). *Spectroscopy and Spectral Analysis* (光谱学与光谱分析), 2001, 21(4): 459.

Study on Developing Calibration Models of Fat Acid Composition in Intact Rapeseed by Near Infrared Reflectance Spectroscopy

WU Jian-guo, SHI Chun-hai*, ZHANG Hai-zhen

Agronomy Department, College of Agriculture and Biotechnology, Zhejiang University, Hangzhou 310029, China

Abstract Four hundred ninety seven rapeseed samples, which feature multi-year, multi-loci and highly variant characteristics, were collected as a raw set. The NIR spectra were pretreated by scatter correction and mathematics treatments, and calibration models of fat acid composition of intact rapeseed were developed by using the algorithm method of modified partial least square (MPLS). Meanwhile, three types of sample cups with different capacity were used to screen the suitable calibration model for rapeseed quality breeding. The results showed that the calibration model of 8-gram-sample was the best, and the calibration determination coefficient was in the range of 0.74-0.98. The calibration effects of 3-gram-sample, which were similar to those of 0.6-gram-sample, were good with high determination coefficient (RSQ1, 1-VR) and low error (SEC, SECV). Therefore, the calibration set with multi-year and multi-loci samples can improve the accuracy and repeatability of NIRS models. The fat acids NIRS models of intact rapeseed developed could be introduced into breeding lines' selection, mutant screening and germplasm evaluation.

Keywords Near infrared reflectance spectroscopy (NIRS); Rapeseed; Fat acid; Calibration model

* Corresponding author

(Received Jun. 6, 2004; accepted Oct. 26, 2004)