

# 豌豆品质性状近红外模型建立及区域差异分析

王姣姣, 刘浩, 任贵兴

(中国农业科学院作物科学研究所, 北京 100081)

**摘要:**探讨了傅立叶变换近红外光谱技术(FT-NIRS)检测豌豆蛋白质、淀粉、脂肪和总多酚含量的可行性。用化学方法测定 190 份豌豆种质的蛋白质、淀粉、脂肪以及总多酚含量,采集其子粒与粉末的近红外光谱,采用偏最小二乘法(PLS)分别建立两种光谱与成份含量预测模型。豌豆粉末模型结果优于子粒模型,其中蛋白质和淀粉的粉末模型的预测残差(RPD)为 5.88、5.82,相关系数  $r^2$  达到 0.99、0.99,具有很好的预测性能。对其中产地信息详细明确的 150 份豌豆种质的品质性状与产地进行两步聚类分析,明确得到 3 种类型,其特点分别为:类群 1 低蛋白质含量,类群 2 高总多酚含量,类群 3 高蛋白质、高淀粉和高脂肪含量。进一步分析了豌豆品质性状随播种期、经度、纬度、海拔高度的变化情况。结果表明,近红外光谱技术可对豌豆种质资源的部分品质性状进行快速筛选鉴定,聚类分析结论、地理坐标与播期对豌豆种质主要品质性状的影响规律,都可作为收集高品质性状豌豆种质资源提供可靠依据。

**关键词:**豌豆;种质资源;品质性状;傅立叶变换近红外光谱;聚类分析;地理坐标

## Using Fourier Transform Near-Infrared Spectroscopy for the Evaluation and Regional Analysis of Pea (*Pisum sativum* L.)

WANG Jiao-jiao, LIU Hao, REN Gui-xing

(Institute of Crop Sciences, Chinese Academy of Agricultural Sciences, Beijing 100081)

**Abstract:** Pea (*Pisum sativum* L.) is an important edible legume. Feasibility of the Fourier Transform Near-Infrared Spectroscopy (FT-NIRS) on estimating quality traits in pea was evaluated in current study. A total of 190 pea accessions involved with their protein, starch, oil, and total polyphenol content were chemically analyzed. After obtaining spectra of the samples in milled powder and intact seed forms, partial least squares (PLS) regression was applied for model development. Models of powder were generally superior to models in intact seed. The optimal models were powder-based for protein and starch with residual predictive deviation (RPD) of 5.88 and 5.82 as well as coefficients of correlation ( $r^2$ ) of 0.99 and 0.99, respectively. High values of correlation coefficient ( $r^2$ ) revealed that models had good predictive capacities for rapid germplasm analysis of pea. To explore the relationship between quality traits and producing regions, 150 pea varieties with specific information were analyzed by two-step cluster analysis. Three distinct groupings were obtained with obvious features. Group 1 was in low protein content. Group 2 was in high total polyphenol content. Group 3 was in high protein, starch, and oil content. The nutrition contents were affected by longitude, latitude, and altitude of planting location as well as seeding date. These results could provide reliable information for the collection of excellent germplasm resources with good quality traits.

**Key words:** *Pisum sativum* L.; germplasm resource; quality traits; FT-NIRS; cluster analysis; geographic coordinate

豌豆 (*Pisum sativum* L.,  $2n = 14$ ) 起源于亚洲西部、地中海、埃塞俄比亚、小亚细亚西部和外高加索

地区<sup>[1]</sup>, 属蝶形花亚科一年生或越年生草本植物。豌豆是世界第 4 大食用豆类作物<sup>[2]</sup>, 在我国已有

收稿日期: 2013-12-06 修回日期: 2013-11-19 网络出版日期: 2014-06-09

URL: <http://www.cnki.net/kcms/detail/11.4996.S.20140609.1408.005.html>

基金项目: 现代农业产业技术体系 (nyeyty-018); 国家农作物种质资源平台; 中国农业科学院科技创新工程

第一作者主要从事食用豆品质评价与研究。E-mail: doubleok@126.com; 刘浩为共同第一作者

通信作者: 任贵兴, E-mail: rengxing@caas.cn

2000 多年的栽培历史,并早已遍及全国。由于豌豆对土壤的适应能力较强,生育期较短,粮菜兼用,种植地理环境、自然生态条件的差异巨大,因此,我国的豌豆种质资源具有丰富的遗传多样性。另一方面,由于豌豆蛋白质含量在 24% 左右,高于水稻、小麦、玉米,是人类和畜禽植物蛋白质的重要来源之一;豌豆淀粉含量在 46% 左右,高于红小豆、菜豆、饭豆、木豆<sup>[3]</sup>,历来是传统粉丝、糕点、淀粉加工制作业的优选原料。随着市场多样化需求的不断增加,豌豆在食品加工工业、医药业、畜牧业等行业越来越显示出其重要作用。

近红外光谱(NIRS, near infrared spectroscopy)技术是光谱测量与化学计量学有机结合的产物,按照美国材料测试学会标准 ASTM 规定,NIRS 范围是介于 780 ~ 2526 nm(12820 ~ 3959/cm)之间的电磁波。NIRS 是通过利用有机物质 C-H、O-H、N-H 等化学键在这一光谱区倍频及合频的振动吸收,并基于不同基团产生的光谱在吸收峰的位置和强度上的差异,根据 Lambert-Beer 定律,吸收光谱特征随样品成分的组成和结构变化,从而实现对样品中多种化学成分进行定性、定量测定的一项技术。由于该技术具有检测速度快、可同时对多种化学成分进行检测分析,以及可对检测样品不破坏,不使用有机试剂,不会造成污染、测量信息数字化等特点,近年来广泛应用于农业、食品、化工、医药、化工、冶金、纺织、质量监督、环境保护等领域<sup>[4-6]</sup>。在近红外光谱仪器的发展方面,也已从 20 世纪 60、70 年代的光栅色散、滤光片发展到 90 年代的傅立叶技术,该技术以其光通量大和信息多通道传输大大提高了近红外光谱分析的灵敏度和分析的深度<sup>[7]</sup>。

NIRS 技术具有常规分析技术难以比拟的诸多优越性,在作物种质资源品质性状鉴定评价方面已有不少应用,是推进作物种质资源品质性状评价的有效鉴定分析工具。杨小红等<sup>[8]</sup>用傅立叶漫反射光谱技术,结合偏最小二乘法建立了玉米子粒 4 种脂肪酸含量和含油量的光谱校正模型。闵顺耕等<sup>[9]</sup>采用傅立叶变换技术测定大麦中蛋白质、淀粉的含量。张环等<sup>[10]</sup>用近红外光谱技术对大麦子粒水分、蛋白质、淀粉进行评价。孙君明等<sup>[11]</sup>采用傅立叶近红外反射光谱法测定大豆脂肪酸含量。朱志华等<sup>[12]</sup>用近红外光栅漫反射技术对 17 种作物 7240 份种质资源的主要营养品质性状进行鉴定评价。李楠等<sup>[13]</sup>用傅立叶变换近红外技术,结合偏最小二乘法建立校正模型,对 167 份大豆种质的 17 种氨基酸

含量进行分析等,都获得了较为满意的结果。但到目前为止,尚未见到有关利用傅立叶变换近红外光谱技术对豌豆种质资源进行研究的报道。

李玲等<sup>[14]</sup>研究发现,来自我国各省区豌豆种质资源群体的遗传多样性指数,变异范围差异明显,并与资源的地理分布相关联。很多研究都表明产地对豆类品质性状有一定影响<sup>[15-16]</sup>。V. Kumar 等<sup>[17]</sup>的研究表明,纬度与大豆蛋白质含量具有显著的负相关性,而与脂肪含量具有显著的正相关性。H. Pizo 等<sup>[18]</sup>认为种植在海拔 300 m 的蚕豆和羽扇豆的蛋白质含量最高,而种植在海拔 700 m 的蛋白质含量最低。Y. Yao 等<sup>[19]</sup>研究发现海拔与赤小豆中总多酚含量呈显著正相关。R. Bressani 等<sup>[20]</sup>研究表明海拔(0 ~ 2256 m)对菜豆的蛋白质、淀粉、单宁和儿茶酚含量无显著影响。总体上,海拔和纬度对豆类成分含量的影响要大于经度的影响。有关豌豆品质性状与产地的关系研究尚未见报道。

本研究利用傅立叶变换近红外光谱技术对 190 份豌豆种质资源的蛋白质、淀粉、脂肪、总多酚的含量进行近红外光谱测定,采用偏最小二乘法(PLS, partial least squares)回归法对预测结果进行研究,并尝试应用两步聚类分析方法对豌豆种质的主要品质性状与产地之间的关系进行分析,进一步研究经度、纬度、海拔高度及播种期的变化对豌豆主要品质性状的影响,旨在为豌豆种质资源的筛选鉴定及收集利用提供支持。

## 1 材料与方法

### 1.1 试验材料和样品处理

试验所用 190 份豌豆种质来自国家种质资源中期库。已知样品的来源地信息见表 1,材料均为同期入库保存的种子且具有良好的代表性。样品于常温条件、干燥器中平衡 7 d,每份材料样品分为 2 份等量试样,其中 1 份用于子粒近红外光谱扫描,另 1 份

表 1 豌豆种质资源样品来源

Table 1 Origin information of pea (*Pisum sativum* L.) germplasm resources

地区 Region	份数 No.	地区 Region	份数 No.
甘肃 Gansu	22	山西 Shanxi	3
河北 Hebei	12	西藏 Tibet	32
江苏 Jiangsu	4	青海 Qinghai	1
内蒙古 Innernengolia	2	重庆 Chongqing	18
宁夏 Ningxia	21	四川 Sichuan	21
陕西 Shaanxi	14	其他 Others	40

经旋风磨研磨成粉,过 60 目筛后分装于封口塑料袋中,用于化学分析与粉末近红外光谱扫描。

## 1.2 化学分析方法与主要仪器

**1.2.1 蛋白质含量** 采用农业行业标准 NY/T 3-1982(谷类、豆类作物种子粗蛋白质测定法-半微量凯氏法),蛋白质-氮的换算系数为 5.70,使用 FOSS2300 型全自动定氮仪(丹麦福斯特卡托公司)测定。

**1.2.2 淀粉含量测定** 采用农业行业标准 NY/T 11-1985(谷物子粒粗淀粉测定法),使用 WZZ-1S 数字式旋光分析仪(上海物理光学仪器厂)测定。

**1.2.3 脂肪含量测定** 采用国家标准 GB/T 5512-2008(粮油检验 粮食中粗脂肪含量测定)中索氏抽提法测定。

**1.2.4 总多酚含量测定** 参考 Y. Yao 等<sup>[21]</sup>没食子酸(gallic acid)—福林试剂紫外分光光度法进行,根据实际情况略有改动。准确称取没食子酸标准品 11.60 mg,加蒸馏水定容至 100 mL,配成 116  $\mu\text{g}/\text{mL}$  的备用液,吸取上述备用液 2 mL、3 mL、4 mL、5 mL、6 mL、7 mL、8 mL、10 mL 加蒸馏水稀释至 10 mL,得到每 1 mL 分别含 23.2  $\mu\text{g}$ 、34.8  $\mu\text{g}$ 、46.4  $\mu\text{g}$ 、58.0  $\mu\text{g}$ 、69.6  $\mu\text{g}$ 、81.2  $\mu\text{g}$ 、92.8  $\mu\text{g}$ 、116.0  $\mu\text{g}$  的没食子酸标准品溶液;准确吸取 1 mL 上述标准品溶液,加 1 mL Folin-Ciocalten (1 mol/L) 显色剂,加  $\text{Na}_2\text{CO}_3$  (1 mol/L) 溶液 3 mL,静置反应 15 min,反应液用蒸馏水定容至 10 mL,混匀,用 1 mL 蒸馏水代替标准品溶液同法操作作为空白,在 765 nm 波长处测定吸光度值;以吸光度为纵坐标,浓度( $\mu\text{g}/\text{mL}$ )为横坐标,绘制标准曲线,回归方程为  $y = 86.002x + 9.4282$  ( $r^2 = 0.9998$ );称取 1 g 样品(精确到 0.001g)置于离心管中,固液比 1:20 加入 80% 乙醇-水溶液,10 mL 6 mol/L HCl,80 mg 抗坏血酸,溶液 72  $^{\circ}\text{C}$  水浴振荡提取 2.5 h,5000 r/min 离心 10 min,取上清液用 80% 乙醇-水溶液定容至 30 mL 制得供试液。测定方法如标准曲线的制备,空白为 80% 乙醇-水溶液。所用没食子酸标准品购于中国药品生物制品检定所,Folin-Ciocalten 试剂购于 Sigma 公司,无水乙醇、无水碳酸钠及其他试剂均为分析纯,购于北京化工厂。总多酚的结果以每克干样品中含有的等效没食子酸的毫克数( $\text{mg}/\text{g}$ )表示。

蛋白质、淀粉、脂肪化学测定结果均以样品干基百分含量表示。

## 1.3 近红外光谱数据采集及模型建立

**1.3.1 粉末和子粒样品的光谱数据采集** 试验采

用德国 Bruker 公司的 Matrix-I 型傅立叶变换近红外光谱分析仪。将 30 g 左右豌豆子粒样品装满于直径 50 mm 的旋转样品池中,在 4000 ~ 12500/ $\text{cm}$  (800 ~ 2500 nm) 谱区范围内扫描 64 次,分辨率设置为 16/ $\text{cm}$ ,采集样品的吸收光谱。为了克服样品粒度差异造成的光谱漂移,每个样品重复装样扫描 3 次,以消除粒度大小、均匀性不一致等因素对光谱的影响,可用 Bruker 公司 OPUS 6.5 近红外处理软件计算平均光谱。样品子粒光谱采集方法与粉末光谱采集方法相同。

**1.3.2 近红外模型的建立与优化** 将样品集(190 份)按 4:1 的比例分成校正集和检验集。采用偏最小二乘法(PLS)作为建立模型的化学计量学方法。利用 Bruker 公司 OPUS 6.5 定量分析软件对方法 1.3.1 中所采集的校正集样品试样光谱与方法 1.2 中化学法检测的相同样品蛋白质、淀粉、脂肪和总多酚含量进行关联回归运算,建立校正模型。利用该分析软件对所建校正模型进行近红外光谱预处理、谱区选择、剔除异样样品以及回归统计分析等优化处理,向校正模型输入检验集样品光谱进行待测成分含量的预测,对所得预测值与相同样品通过方法 1.2 测得的化学值进行相关性分析验证,以检验定标模型的预测能力。应用该软件的近红外 OPTIMIZE 自动优化功能优化谱区范围和数据预处理方法组合,自动剔除建模含量范围中间的异常样品点,对多种组合进行内部交叉验证和外部验证,从中选择最佳预处理方法、光谱谱区范围和优化维度。根据外部验证的相关系数( $r^2$ )、外部验证决定系数( $R_p^2$ )、交叉验证决定系数( $R^2$ )、预测均方根误差(RMSEP, root mean square error of prediction)、交叉验证均方根误差(RMSECV, root mean square error of cross validation)、残余预测偏差(RPD, residual predictive deviation)和维度(rank)等指标确定最优校正模型。再经比较预测值和化学测定值的符合程度最终评价其模型的预测效果。

## 1.4 聚类分析

从参试的 190 份豌豆种质中选择具有明确产地信息的样品 150 份,应用 SPSS 13.0 数据处理系统的两步聚类分析(two-step cluster analysis)法,以对数似然函数测距法(log-likelihood)为聚类依据,以贝叶斯标准(BCI, Schwarz's Bayesian Criterion)为判定聚类数目的准则,按照豌豆品质特性对其不同产地的材料进行分析,分析得到每个大类的组成特点。

## 1.5 豌豆种质的地理坐标与播种期对其品质性状的影响

豌豆原产地地理坐标来源于金农网全国县区分站 (<http://www.agri.com.cn/host/province/china.htm>) 的数据,为了便于统计和作图,将地理坐标经纬度转化为十进制。豌豆的播种期信息由中国农业科学院作物科学研究所提供。

## 1.6 数据处理

通过 SPSS 13.0 数据处理系统进行 Pearson 相关性检验(2-tailed)。采用独立样本均值差异性检

表 2 豌豆蛋白质、淀粉、脂肪和总多酚含量的化学检测统计结果

Table 2 Statistics results of protein, starch, oil, and total polyphenol contents in pea determined by chemical methods

品质性状 Quality traits	均值 ± 标准差 Mean ± SD	变幅 Range	变异系数 CV	偏度 Skewness	峰度 Kurtosis
蛋白质 (%) Protein	25.19 ± 2.04	19.85 ~ 30.25	0.08	-0.01	-0.16
淀粉 (%) Starch	51.31 ± 3.44	33.13 ~ 57.10	0.07	-2.06	8.69
脂肪 (%) Oil	1.33 ± 0.48	0.49 ~ 2.62	0.36	0.47	-0.23
总多酚 (mg/g) Total polyphenol	1.91 ± 0.45	0.83 ~ 3.45	0.24	0.43	0.31

从表 2 可见,豌豆种质淀粉平均含量  $51.31\% \pm 3.44\%$ ,变幅在  $33.13\% \sim 57.10\%$ ;蛋白质平均含量  $25.19\% \pm 2.04\%$ ,变幅在  $19.85\% \sim 30.25\%$ ;脂肪平均含量  $1.33\% \pm 0.48\%$ ,变幅在  $0.49\% \sim 2.62\%$ ;总多酚平均含量  $1.91 \pm 0.45 \text{ mg/g}$ ,变幅在  $0.83 \sim 3.45 \text{ mg/g}$ 。从豌豆种质主要品质性状的变异看,以脂肪的变异程度最大( $CV = 0.36$ ),总多酚次之( $CV = 0.24$ ),蛋白质和淀粉含量变异较小( $CV$ 值分别为 0.08 和 0.07)。豌豆种质主要品质性状的含量化学测定结果与文献研究报道的蛋白质含量( $20.0\% \sim 24.9\%$ )、淀粉含量( $28.70\% \sim 58.69\%$ )、脂肪含量( $1.00\% \sim 2.70\%$ )、总多酚含量( $1.83 \sim 4.60 \text{ mg/g}$ )的结果基本一致<sup>[22-27]</sup>。

就其化学测定结果而言,目前国内尚未建立统一完善的豌豆种质主要品质性状评价标准,以豌豆 4 项品质含量的平均值及其标准差( $\bar{x} \pm s$ )为分类依据,筛选出蛋白质含量超过 27.23% 的豌豆种质 31 份,占参试种质的 16.34%;筛选出淀粉含量超过 54.8% 的豌豆种质 16 份,占参试种质的 8.42%;筛选出脂肪含量超过 1.81% 的豌豆种质 29 份,占参试种质的 15.26%;筛选出总多酚含量低于  $1.46 \text{ mg/g}$  的豌豆种质 39 份,占参试种质的 20.5%。以豌豆种质蛋白质、淀粉、脂肪含量较高且总多酚含量较低作为优良豌豆种质的筛选依据,兼顾高总多酚含量作为抗氧化功能食品开发原料的要求,筛选出蛋白质、淀粉含量较高而总多酚含量较低

的种质白豌豆 G0005559 (蛋白质 28.82%、淀粉 54.87%、总多酚  $1.40 \text{ mg/g}$ ) 和豌豆 G0005579 (蛋白质 28.25%、淀粉 54.97%、总多酚  $1.28 \text{ mg/g}$ )、筛选出蛋白质、脂肪含量较高而总多酚含量较低的种质白豌豆 G0005571 (蛋白质 27.64%、脂肪 2.24%、总多酚  $1.37 \text{ mg/g}$ )、筛选出蛋白质和总多酚含量较高的种质豌豆 G0004550 (蛋白质 27.44%、总多酚  $2.54 \text{ mg/g}$ )、筛选出淀粉和脂肪含量较高且总多酚含量较低的种质豌豆 G0005457 (淀粉 55.31%、脂肪 1.86%、总多酚  $1.02 \text{ mg/g}$ )、以及脂肪和总多酚均较高的种质豌豆 G0004303 (脂肪 2.11%、总多酚  $3.45 \text{ mg/g}$ )、小麻豌豆 G005401 (脂肪 2.16%、总多酚  $3.19 \text{ mg/g}$ ) 和白豌豆 G0005403 (脂肪 1.85%、总多酚  $3.15 \text{ mg/g}$ ) 等。

## 2 结果与分析

### 2.1 蛋白质、淀粉、脂肪和总多酚含量及其相关性

2.1.1 蛋白质、淀粉、脂肪和总多酚含量 190 份豌豆种质的蛋白质、淀粉、脂肪、总多酚含量的化学测定值统计结果见表 2。

的种质白豌豆 G0005559 (蛋白质 28.82%、淀粉 54.87%、总多酚  $1.40 \text{ mg/g}$ ) 和豌豆 G0005579 (蛋白质 28.25%、淀粉 54.97%、总多酚  $1.28 \text{ mg/g}$ )、筛选出蛋白质、脂肪含量较高而总多酚含量较低的种质白豌豆 G0005571 (蛋白质 27.64%、脂肪 2.24%、总多酚  $1.37 \text{ mg/g}$ )、筛选出蛋白质和总多酚含量较高的种质豌豆 G0004550 (蛋白质 27.44%、总多酚  $2.54 \text{ mg/g}$ )、筛选出淀粉和脂肪含量较高且总多酚含量较低的种质豌豆 G0005457 (淀粉 55.31%、脂肪 1.86%、总多酚  $1.02 \text{ mg/g}$ )、以及脂肪和总多酚均较高的种质豌豆 G0004303 (脂肪 2.11%、总多酚  $3.45 \text{ mg/g}$ )、小麻豌豆 G005401 (脂肪 2.16%、总多酚  $3.19 \text{ mg/g}$ ) 和白豌豆 G0005403 (脂肪 1.85%、总多酚  $3.15 \text{ mg/g}$ ) 等。

2.1.2 蛋白质、淀粉、脂肪和总多酚含量间的相关性 不同品质性状含量的相关性分析结果(表 3)表明,蛋白质含量与淀粉含量呈极显著负相关( $P < 0.01$ ),淀粉含量与总多酚含量呈极显著负相关( $P < 0.01$ ),其他品质性状间相关性均不显著( $P > 0.05$ )。

### 2.2 品质性状模型的建立、优化及分析

从豌豆种质主要品质性状含量扫描光谱图(图 1)可见,豌豆种质的粉末样品和子粒样品图谱相似,同时发现两组样品扫描图谱在  $4200/\text{cm}$ 、 $4600/\text{cm}$ 、 $5200/\text{cm}$ 、 $5500/\text{cm}$ 、 $6900/\text{cm}$ 、 $8300/\text{cm}$  波段中均出现 6 个较强的吸收峰。

表 3 豌豆蛋白质、淀粉、脂肪和总多酚含量间的相关性

Table 3 Correlation of contents of protein, starch, oil, and total polyphenol in pea

品质性状 Quality traits	蛋白质 Protein	淀粉 Starch	脂肪 Oil
淀粉 Starch	-0.34 **	—	—
脂肪 Oil	0.05	-0.03	—
总多酚 Total polyphenol	0.04	-0.25 **	-0.05

\*: 在 0.05 水平上有显著相关性; \*\*: 在 0.01 水平上有显著相关性。下同

\*: Correlation is significant at the 0.05 level (2-tailed), \*\*: Correlation is significant at the 0.01 level (2-tailed). The same as below

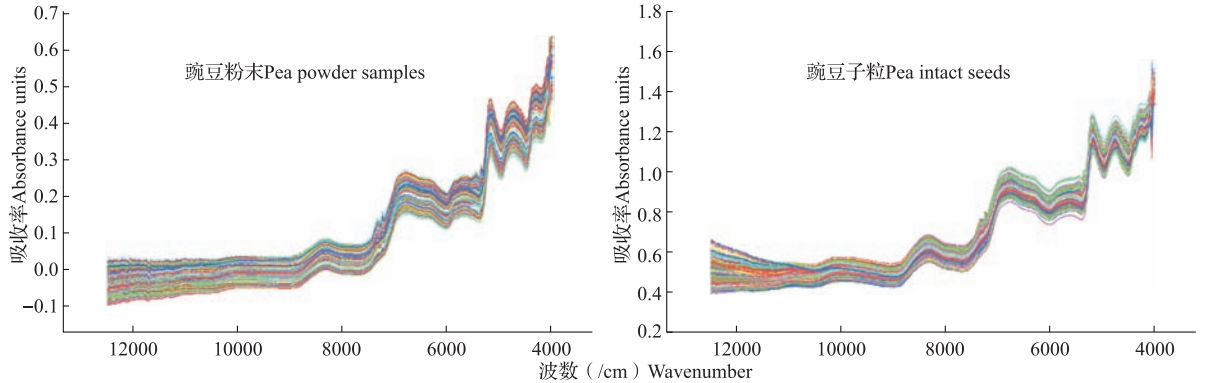


图 1 豌豆样品粉末和子粒近红外光谱图

Fig. 1 Typical NIR spectra obtained from 190 pea in powder samples and intact seeds

表 4 豌豆蛋白质、淀粉、脂肪和总多酚近红外 PLS 模型的参数

Table 4 Parameters of NIR PLS model of protein, starch, oil, and total polyphenol in pea

样品形式 Sample status	品质性状 Quality traits	样品数量 Sample size		平均值 ± 标准差 Mean ± SD		变幅 Range	
		Cal	Val	Cal	Val	Cal	Val
粉末 Milling powder	蛋白质 (%) Protein	149	35	25.23 ± 1.99	24.95 ± 2.07	19.85 ~ 30.25	20.79 ~ 30.22
	淀粉 (%) Starch	152	35	50.99 ± 3.63	51.39 ± 2.75	33.13 ~ 57.10	44.58 ~ 56.89
	脂肪 (%) Oil	141	35	1.21 ± 0.41	1.21 ± 0.39	0.49 ~ 2.56	0.57 ~ 2.24
	总多酚 (mg/g) Total polyphenol	151	35	1.89 ± 0.43	1.86 ± 0.39	1.00 ~ 3.19	1.08 ~ 2.66
完整子粒 Intact seed	蛋白质 (%) Protein	149	35	25.13 ± 1.90	24.28 ± 2.16	19.87 ~ 30.25	20.79 ~ 30.13
	淀粉 (%) Starch	141	35	51.40 ± 2.75	51.32 ± 2.57	36.55 ~ 57.10	44.58 ~ 56.89
	脂肪 (%) Oil	141	35	1.16 ± 0.37	1.26 ± 0.36	0.55 ~ 2.56	0.58 ~ 2.17
	总多酚 (mg/g) Total polyphenol	141	35	1.86 ± 0.38	1.87 ± 0.38	1.03 ~ 3.04	1.34 ~ 2.57

Cal: 定标集; Val: 检验集

Cal: Calibration set, Val: Validation set

含量在使用不同的预处理方法下的  $R_p^2$  与 RMSEP 值的大小选择模型, 得到建立近红外模型的最佳条件 (表 5)。

豌豆品质性状的优化 PLS 回归模型的内部交叉验证和外部验证的检验统计量结果见表 6。总体上豌豆粉末品质性状最优, NIRS 回归模型优于子粒模型, 豌豆 4 项品质性状模型的相关系数  $r^2$  均在 0.94 以上, 内部交叉检验和外部检验的决定系数  $R_c^2$  和  $R_p^2$  均在 0.86 以上, RMSCV 和 RMSEP 达到最小值, 其中粉末蛋白质和淀粉模型的相关系数 ( $r^2 = 0.99, r^2 = 0.99$ ) 和决定系数 ( $R_p^2 = 0.97, R_p^2 = 0.97$ ) 最高, 其次是子粒蛋白质 ( $r^2 = 0.97, R_p^2 = 0.94$ ) 和淀

近红外光谱通过自动剔除异常点参数值, 最终获得了用于建立模型的定标集和检验集样品。豌豆种质粉末和子粒样品主要品质性状定标集和检验集的基础统计参数见表 4。定标集和检验集数据组成可见, 豌豆种质主要品质性状含量的变幅涵盖范围较大, 含量分布较均匀, 具有较好的代表性, 符合近红外光谱建模的要求。

本研究将化学方法测定值与获得的光谱数据, 导入 OPUS 6.5, 通过自动优化给出的不同品质性状

粉 ( $r^2 = 0.95, R_p^2 = 0.89$ ), 豌豆总多酚的粉末 ( $r^2 = 0.95, R_p^2 = 0.90$ ) 和子粒 ( $r^2 = 0.94, R_p^2 = 0.88$ ) 模型接近, 脂肪模型的相关系数 ( $r^2 = 0.94, r^2 = 0.94$ ) 和决定系数 ( $R_p^2 = 0.87, R_p^2 = 0.86$ ) 最低。豌豆粉末蛋白质和淀粉模型内部交叉验证和外部验证的 RPD 值都接近 6, 表明该模型具有非常优良的预测性和检测准确度, 子粒蛋白质和淀粉模型以及粉末总多酚模型的内部交叉验证和外部验证的 RPD 值在 3.0 ~ 4.0 之间, 也说明这些模型具有一定的预测能力, 子粒总多酚、粉末脂肪和子粒脂肪的 RPD 值在 2.5 ~ 3.0 之间, 说明这些模型的预测能力比较有限。

表 5 豌豆近红外模型建立的优化条件

Table 5 Optimal conditions of NIR PLS models of pea varieties

样品形式	品质性状	优化光谱范围(/cm)	优化光谱前处理
Sample status	Quality traits	Wavenumber region	Data pre-processing
粉末 Milling powder	蛋白质 (%) Protein	6102 ~ 4242. 8	线性补偿差减法 Constant offset elimination
	淀粉 (%) Starch	7506 ~ 5446. 3	矢量归一法 Multiplicative Scattering Correction
	脂肪 (%) Oil	6804 ~ 6094. 3; 454 ~ 4242. 8	最小-最大归一法 Min-Max normalization
	总多酚 (mg/g) Total polyphenol	7722 ~ 4242. 8	二阶导数法 Second derivative(17ps)
完整子粒 Intact seed	蛋白质 (%) Protein	12489. 4 ~ 6078. 8; 5592. 8 ~ 4597. 7	一阶导数 + 乘法散射矫正法 First derivative + MSC(17ps)
	淀粉 (%) Starch	12265. 7 ~ 7498. 3; 6102 ~ 4597. 7	线性补偿差减法 Constant offset elimination
	脂肪 (%) Oil	8663. 1 ~ 7737. 4; 6595. 7 ~ 5438. 6	一阶导数 + 矢量归一法 First derivative + SNV(17ps)
	总多酚 (mg/g) Total polyphenol	12489. 4 ~ 7498. 3; 6102 ~ 4597. 7	线性补偿差减法 Constant offset elimination

MSC:乘法散射矫正;SNV:矢量归一法

MSC: Multiplicative scattering correction, SNV: Vector normalization

表 6 豌豆品质性状粉末和子粒优化 PLS 回归模型的内部交叉验证和外部验证统计量

Table 6 Statistics of PLS regression in cross-validation and test set validation for quality traits by OPUS 6. 5

样品形式	品质性状	相关系数	维度	内部交叉验证			外部验证		
				$R_C^2$	RMSCV	RPD	$R_p^2$	RMSEP	RPD
Sample status	Quality traits	$r^2$	Rank	Cross validation			Test set validation		
粉末 Milling powder	蛋白质 Protein	0.99	8	0.97	0.341	5.83	0.97	0.351	5.88
	淀粉 Starch	0.99	10	0.97	0.627	5.78	0.97	0.472	5.82
	脂肪 Oil	0.94	8	0.86	0.152	2.68	0.87	0.136	2.86
	总多酚 Total polyphenol	0.95	8	0.89	0.141	3.02	0.90	0.124	3.11
完整子粒 Intact seed	蛋白质 Protein	0.97	7	0.94	0.472	4.01	0.94	0.542	4.03
	淀粉 Starch	0.95	8	0.89	0.930	2.95	0.89	0.867	3.07
	脂肪 Oil	0.94	7	0.86	0.138	2.65	0.86	0.135	2.76
	总多酚 Total polyphenol	0.94	8	0.86	0.141	2.70	0.88	0.129	2.93

$r^2$ :外部验证相关系数; $R_C^2$ :交叉验证决定系数; $R_p^2$ :外部验证决定系数;RMSCV:交叉验证均方根误差;RMSEP:预测均方根误差;RPD:残余预测偏差

$r^2$ :Coefficient of correlation in test set validation, $R_C^2$ :Coefficient of determination of cross validation, $R_p^2$ :Coefficient of determination of test set validation, RMSCV:Root mean square error of cross validation,RPD:Residual predictive deviation, RMSEP:Root mean square error of prediction

图 2、图 3 分别给出了豌豆种质粉末和子粒近红外 PLS 回归模型的外部验证曲线。粉末样品蛋白质、淀粉、脂肪、总多酚含量的决定系数  $R_p^2$  值分别为 0.97、0.97、0.87、0.90,脂肪、总多酚含量决定系数明显低于蛋白质、淀粉含量。豌豆种质子粒样品蛋白质、淀粉、脂肪、总多酚含量的决定系数  $R_p^2$  值分

别为 0.94、0.89、0.86、0.88,结合表 6 模型处理获得的预测值和化学值间相关性外部验证拟合曲线结果,豌豆子粒样品的近红外模型预测能力未达到预测要求,有关豌豆种质子粒样品的近红外预测模型的优化还有待进一步研究。

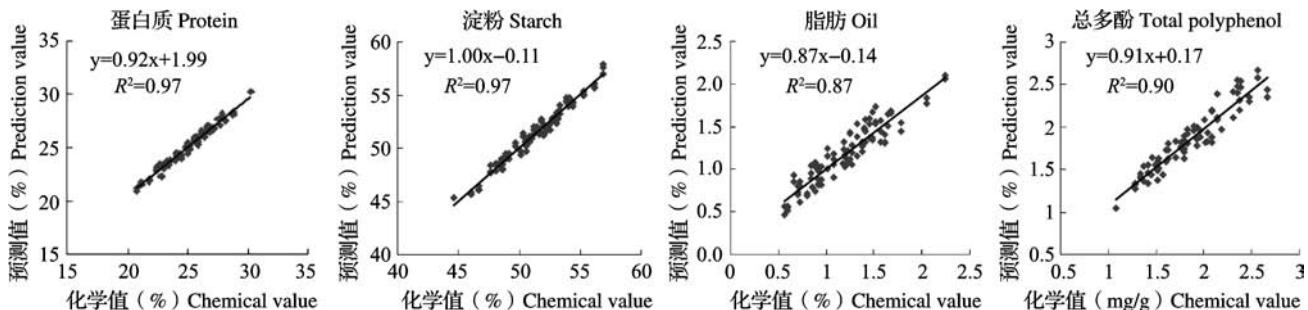


图 2 豌豆品质性状粉末近红外 PLS 优化模型外部验证回归线

Fig. 2 Optimized regression line in test set validation of PLS models for the quality traits of milling powder obtained by FT-NIRS

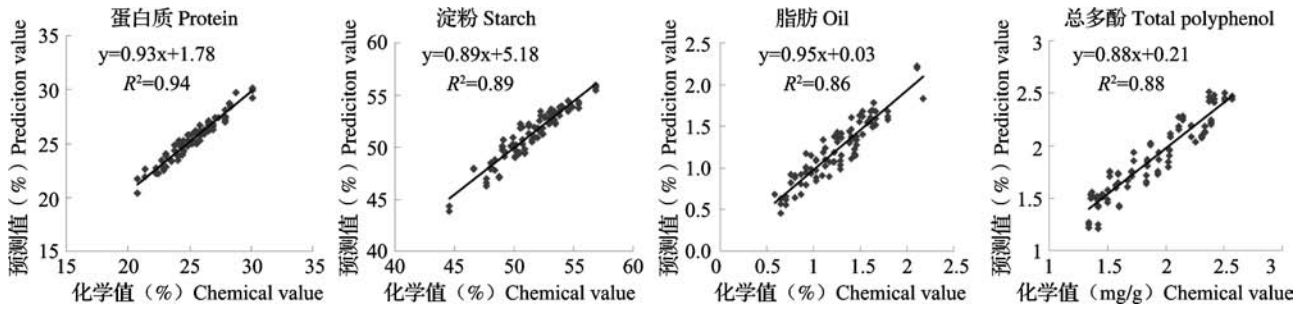


图 3 豌豆品质性状子粒近红外 PLS 优化模型外部验证回归线

Fig. 3 Optimized regression line in test set validation of PLS models for the quality traits of intact seeds obtained by FT-NIRS

2.3 主要品质性状的两步聚类分析

以豌豆种质主要品质性状的均值为变量进行两步聚类分析,获得聚类特征(表 7),结果表明,不同产地的豌豆种质大致可以聚类划分为 3 类。其中,

分布在类群 1 中的有 56 份,所占比例为 37.3%;分布在类群 2 中的有 55 份,所占比例为 36.7%;分布在类群 3 中的有 39 份,占 26.0%。

两步聚类分析区别于其他常规聚类分析法的一

表 7 两步聚类分析中豌豆品质性状的聚类特征

Table 7 Features of three groups of quality traits in pea varieties by two-step cluster analysis

品质性状 Quality traits	样品数量 Sample size	蛋白质 (%) Protein	淀粉 (%) Starch	脂肪 (%) Oil	总多酚 (mg/g) Total polyphenol
类群 1 Group 1	56	24.52 ± 1.46	52.67 ± 2.56	1.34 ± 0.49	1.84 ± 0.52
类群 2 Group 2	55	25.30 ± 1.94	50.20 ± 2.72	1.34 ± 0.52	1.93 ± 0.46
类群 3 Group 3	39	26.81 ± 1.50	52.82 ± 2.04	1.51 ± 0.29	1.74 ± 0.37
合计 Total	150	25.40 ± 1.93	51.84 ± 2.77	1.38 ± 0.46	1.85 ± 0.47

个特点是在丰富的数据资料中发现不明显的聚类组的特性,根据 SPSS 统计结果,对于每个类群,通过对各项变量(蛋白质、淀粉、脂肪、总多酚)均值和标准偏差的分析结果表明,类群 1 中豌豆种质的蛋白质含量、脂肪含量低(24.52% ± 1.46%, 1.34% ± 0.49%),豌豆种质来自河北、内蒙古、江苏、宁夏、陕西、山西。类群 2 中豌豆种质淀粉含量、脂肪含量低(50.20% ± 2.72%, 1.34% ± 0.52%)、总多

酚含量高(1.93 ± 0.46 mg/g),豌豆种质来自西藏、青海、甘肃。类群 3 中豌豆种质的蛋白质含量、淀粉含量、脂肪含量高(26.81% ± 1.50%, 52.82% ± 2.04%, 1.51% ± 0.29%)、总多酚含量低(1.74 ± 0.37 mg/g),豌豆种质来自重庆、四川。

图 4 显示了 3 组类群之间差异的显著性(P = 0.05),其差异显著性由高至低依次为类群 3、类群 2、

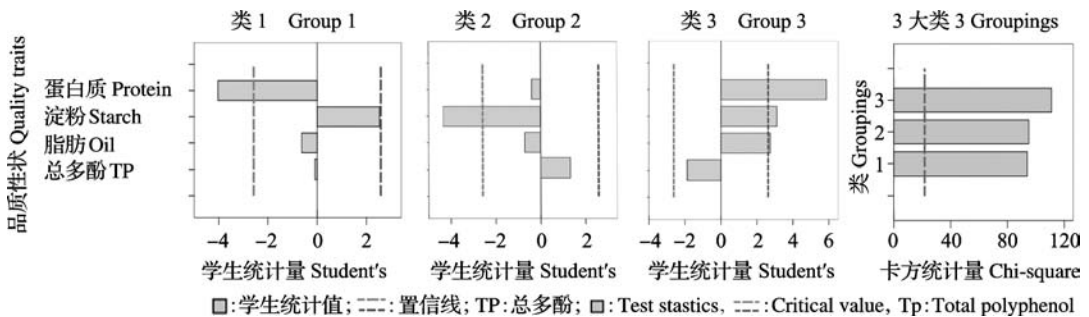


图 4 类群 1、2 和 3 中特征重要性和三类群间的显著性差异

Fig. 4 Significance of principal constituents in Group 1, Group 2, and Group 3 as well as significant difference existed among three groupings by two-step cluster analysis

类群 1。图 4 还分别给出了 3 个类群中的蛋白质、脂肪、淀粉和总多酚的重要性,横坐标为检验的学生统计量  $t(\alpha=0.05)$ ,其中  $t$  统计量正值表示该含量大于平均值,负值则小于平均值,当统计量的绝对值超过虚线则该含量具有显著影响,图内纵向虚线表示各个品质性状变量对本类群具有的决定性临界值 ( $\alpha=0.05$ )。类群 1 中蛋白质具有显著影响性,淀粉、脂肪、总多酚的影响性不显著,淀粉高于平均值,脂肪、总多酚均低于平均值;类群 2 中淀粉有显著影响性,蛋白质、脂肪、总多酚影响不显著,蛋白质、脂肪低于平均值,总多酚高于平均值;类群 3 中蛋白质、淀粉和脂肪有显著影响性,总多酚低于平均值。不同豌豆种质主要品质性状在不同类群中的影响显著性结果与表 7 类群特征结果吻合。

## 2.4 播种期和地理坐标对豌豆品质性状的影响

### 2.4.1 播种期对豌豆品质性状的影响

参试的 150 份豌豆种质包括春播豌豆种质 110 份,秋播豌豆种质 40 份。不同播种期(春播、秋播)豌豆种质品质性状变化结果见表 8。

表 8 豌豆种质资源品质性状的春秋播均值

Table 8 Mean content of protein, starch, oil, and total polyphenol of spring and winter seeded pea

品质性状 Quality traits	春播 Spring sowing date	秋播 Winter sowing date
蛋白质 (%) Protein	24.91 ± 1.83	26.77 ± 1.51
淀粉 (%) Starch	51.44 ± 2.92	52.81 ± 2.02
脂肪 (%) Oil	1.34 ± 0.50	1.51 ± 0.29
总多酚 (mg/g) Total polyphenol	1.89 ± 0.49	1.73 ± 0.36

由表 8 可见,不同播种期对豌豆种质主要品质的含量有影响。其中春播豌豆蛋白质平均含量为 24.91% ± 1.83%,秋播豌豆种质蛋白质平均含量为 26.77% ± 1.51%,秋播豌豆种质蛋白质含量极显著高于春播蛋白质含量 ( $F=3.380, P < 0.01$ );春播豌豆种质淀粉平均含量为 51.44% ± 2.92%,秋播豌豆种质淀粉平均含量为 52.81% ± 2.02%,秋播豌豆种质淀粉含量极显著高于春播淀粉含量 ( $F=8.785, P < 0.01$ );秋播豌豆种质脂肪平均含量为 1.51% ± 0.29%,春播豌豆种质脂肪平均含量为 1.34% ± 0.50%,秋播豌豆种质脂肪含量显著高于春播脂肪含量 ( $F=16.009, P < 0.05$ );春播总多酚平均含量为 1.89 ± 0.49 mg/g,秋播总多酚平均含量为 1.73 ± 0.36 mg/g,春播豌豆种质总多酚含量显著高于秋播总多酚含量 ( $F=4.273, P < 0.05$ )。

## 2.4.2 地理坐标与豌豆种质品质性状的相关性

豌豆种质品质性状与产地地理坐标的统计结果见表 9。结果表明,不同的地理坐标的经度、纬度和海拔因素对豌豆种质的主要品质性状均有影响。其中,蛋白质含量与纬度呈极显著负相关 ( $P < 0.01$ ),即随着纬度的升高,蛋白质含量递减;淀粉含量与经度呈极显著的正相关 ( $P < 0.01$ ),与海拔呈极显著的负相关 ( $P < 0.01$ ),即随着经度增加,海拔降低,淀粉含量呈上升的趋势;总多酚含量与海拔呈极显著正相关 ( $P < 0.01$ ),即随着海拔高度上升,总多酚含量有增加的趋势;未发现脂肪含量与其地理坐标的显著相关性。从地理坐标对豌豆种质品质性状的影响看,海拔的影响比经度、纬度的影响更广泛。

表 9 豌豆品质性状与产地地理坐标的相关系数

Table 9 Coefficients of correlation ( $r^2$ ) between quality traits and geographic coordinates of pea varieties

地理坐标 Geographic coordinates	蛋白质 Protein	淀粉 Starch	脂肪 Oil	总多酚 Total polyphenol
经度 Longitude	0.01	0.34 **	0.04	-0.11
纬度 Latitude	-0.32 **	0.07	-0.07	0.03
海拔 Elevation	-0.12	-0.45 **	-0.07	0.26 **

## 3 讨论

近红外光谱分析是复杂、重叠光谱分析,测定样品不经预处理,故样品的状态、测定方式以及测定的条件都会影响测定结果<sup>[28]</sup>。因此得到的光谱模型是由样品待测成分、样品背景成分、测量条件、测定环境等信息变量的复杂参数构成<sup>[29]</sup>。本研究为了使模型达到理想的分析效果,在试验材料的选择上注意了种质产地来源的代表性,在主要品质性状的含量上关注到样品含量的涵盖幅度,同时注意不同子粒大小、不同子粒饱满度、不同颜色子粒等样品的采用比例,以提高模型的稳健性,拓宽模型的适用范围。为了减少试验误差,采用了平均化学值、平均光谱为基础参数的方法,在进行建模软件的优化、对异常值剔除等选用了合适的预处理方法以尽量扣除干扰信息。建立的豌豆种质粉末模型的决定系数 ( $R^2$ ) 在 86% ~ 97% 之间,获得了比较理想的预测效果。

由于近红外光谱中信息强度低,因此组分吸收较弱<sup>[28]</sup>。本试验中蛋白质、淀粉预测模型的决定系数明显高于脂肪、总多酚的预测模型。此外,豌豆子粒模型比其粉末模型的预测性差,究其原因,与豌豆



子粒大小、表面光滑度、粗糙和褶皱等有直接关系,这些因素对样品的装样状态、装样条件等造成的影响,导致测定条件不完全一致,造成了光谱的不确定性。因此,样品外观和内部的差异性,包括子粒大小、表面光滑度、水分含量和样品内部各成份之间的相互作用等,都会带来近红外光谱吸收的波动变化,影响模型的预测准确性<sup>[30-31]</sup>。有关如何提高豌豆子粒模型方案仍有待进一步研究。

宗绪晓等<sup>[32-33]</sup>的研究表明豌豆核心种质组群分布与豌豆种植区域分布有一定关系。特别是中国栽培豌豆多在边远山区种植,地理交通闭塞,气候差异大,由此形成的长期相互隔离的农业生态系统有可能对中国地方品种的基因库形成产生影响。本研究用两步聚类分析的结果将品种-产地关系大致分为3个区,即东部中部区、西部区和西南部区,从品质性状的测定结果看,不同区域来源的品种其品质特征不同,如四川和重庆地理环境特殊,传统种植区域封闭,种质的品质性状表现独特;西藏、青海和甘肃地理海拔高气候环境相似,种质的品质性状表现相似;中部和东部地区种植区域相对开放,种质的品质性状表现具有一致性,可能与其基因特殊性和种植环境均有关。

豆类营养成分的含量会受产地地理坐标(经度、纬度和海拔高度)的影响,但不同豆类受影响的方式可能不同。由于海拔增高常常相应地增加紫外线UV辐射强度,细胞脂类会因UV辐射强度增加促进过氧化。而具有抗氧化性的总多酚含量增加可能有防止细胞脂类过氧化,保护细胞组织正常运行的作用。有研究发现,与酚类代谢有关的途径,有脯氨酸的磷酸戊糖途径(PPP, pentose phosphate pathway)模型很可能与应激UV辐射的保护性抗氧化代谢有关<sup>[34-35]</sup>。纬度对蛋白质含量影响显著,经度和海拔高度对淀粉含量影响显著,海拔高度对总多酚含量影响显著,但这些影响产生的原因尚有待进一步的研究探讨。

## 4 结论

以不同样品状态的豌豆种质资源为材料,建立基于傅立叶变换近红外光谱的模型,并对不同产地豌豆种质主要品质性状进行分析。结果表明,利用近红外光谱技术对豌豆种质资源的主要品质性状进行鉴定评价可以获得理想的预测结果,该技术是有效、简便、环保、实用的快速种质资源鉴定方法;对豌豆种质资源主要品质性状及其不同产地的影响进行

分析,为豌豆优良品质资源的选择、收集提供了可能的途径。

## 参考文献

- [1] 贺晨昉,宗绪晓. 豌豆种质资源形态标记遗传多样性分析[J]. 植物遗传资源学报,2011,12(1):42-48
- [2] 宗绪晓,Ford R,Redden R R,等. 豌豆属(*Pisum*) SSR 标记遗传多样性结构鉴别与分析[J]. 中国农业科学,2009,42(1):36-46
- [3] 朱志华,李为喜,张晓芳,等. 食用豆类种质资源粗蛋白及淀粉含量的评价[J]. 植物遗传资源学报,2005,6(4):427-430
- [4] Liu D,Zeng X A,Sun D W. NIR Spectroscopy and imaging techniques for evaluation of fish quality-a review[J]. Appl Spectrosc Rev,2013,48(8):609-628
- [5] Tsuchikawa S,Schwanninger M. A review of recent near-infrared research for wood and paper(part 2)[J]. Appl Spectrosc Rev,2013,48(7):560-587
- [6] Vahmeijer A L,Hutteman M, Van der Vorst J R, et al. Image-guided cancer surgery using near-infrared fluorescence[J]. Nat Rev Clin Oncol,2013,10(9):507-518
- [7] 陈文杰,谭小力,王竹云,等. 用傅立叶变换近红外光谱仪测定油菜种子品质指标的研究[J]. 陕西农业科学,2002(8):6-9
- [8] 杨小红,郭玉秋,傅旻,等. 利用近红外光谱分析玉米籽粒脂肪酸含量的研究[J]. 光谱学与光谱分析,2009(1):108-111
- [9] 闵顺耕,覃方丽,李宁,等. 傅里叶变换近红外光谱法测定大麦种蛋白质、淀粉和赖氨酸含量[J]. 分析化学,2003(7):843-845
- [10] 张环,车宗贤,李玉芳. 用近红外光谱法测定大麦品质的研究[J]. 甘肃农业科技,2009(3):6-9
- [11] 孙君明,韩粉霞,闫淑荣,等. 傅里叶近红外反射光谱法快速测定大豆脂肪酸含量[J]. 光谱学与光谱分析,2008(6):92-97
- [12] 朱志华,王文真,刘三才,等. 近红外漫反射光谱分析技术在作物种质资源品质性状鉴定中的应用[J]. 现代科学仪器,2006(1):63-66
- [13] 李楠,许韵华,宋雯雯,等. 利用近红外光谱技术快速检测大豆氨基酸含量[J]. 植物遗传资源学报,2012,13(6):1037-1044
- [14] 李玲. 国内豌豆种质资源形态性状多样性分析[D]. 北京:中国农业科学院,2009
- [15] Nikolopoulou D,Grigorakis K,Stasini M, et al. Differences in chemical composition of field pea (*Pisum sativum*) cultivars: Effects of cultivation area and year[J]. Food Chem,2007,103(3):847-852
- [16] Wang N,Hatcher D W,Warkentin T D, et al. Effect of cultivar and environment on physicochemical and cooking characteristics of field pea (*Pisum sativum*) [J]. Food Chem,2010,118(1):109-115
- [17] Kumar V,Rani A,Solanki S, et al. Influence of growing environment on the biochemical composition and physical characteristics of soybean seed[J]. J Food Compos Anal,2006,19:188-195
- [18] Pizo H,Bobrecka-Jamro D,Rzasa B. Nutrient contents in horse bean and lupin seeds grown in the foothill areas of southeastern Poland[J]. Zeszyty Naukowe Akademii Rolniczej im Hugona Kozatajaw Krakowie,Sesja Naukowa,1992,32(260):267-276
- [19] Yao Y,Cheng X Z,Wang S H, et al. Influence of altitudinal variation on the antioxidant and antidiabetic potential of adzuki bean (*Vigna angularis*) [J]. Int J Food Sci Nutr,2012,63:117-124
- [20] Bressani R,Chon C. Effects of altitude above sea level on the cooking time and nutritional value of common beans[J]. Plant Food Hum Nutr,1996,49:53-61