

文章编号: 1005-0906(2007)03-0062-05

# 高油玉米群体油分、蛋白质和淀粉含量 近红外分析模型的构建

张 俊, 张义荣, 卢宝红, 宋秀芳, 宋同明, 李建生

(中国农业大学, 国家玉米改良中心, 北京市作物遗传育种重点实验室, 北京 100094)

**摘 要:** 以 BHO 高油玉米  $F_{2:3}$  家系为材料, 应用主成分空间和傅里叶变换近红外光谱分析技术, 采用偏最小二乘回归法(PLS), 建立了测定高油玉米子粒的油分、蛋白质和淀粉含量的近红外校正模型。预处理分别采用一阶导数 + 矢量归一化、一阶导数 + 多元散射校正及直线相减等方法, 主成分维数分别为 5、9、9。验证分析表明, 所建立的油分、蛋白质和淀粉含量的校正模型的校正和预测效果最好, 其校正决定系数( $R^2_{cal}$ )分别为 0.950、0.973、0.976, 交叉验证决定系数( $R^2_{cv}$ )和外部验证决定系数( $R^2_{val}$ )在 0.918 ~ 0.948, 各项误差(RMSEE、RMSECV、RMSEP)在 0.305% ~ 0.721%。结果表明, 所建立的高油玉米完整子粒品质性状三成分模型的准确度和精确度均较高, 可以满足高油玉米群体大量样品无损品质分析的需要。

**关键词:** 高油玉米; 子粒品质; 近红外光谱; 校正模型

**中图分类号:** S513.01

**文献标识码:** A

## A Model for Determination of Kernel Oil, Protein and Starch Contents with High-oil Maize by Near Infrared Reflectance Spectroscopy

ZHANG Jun, ZHANG Yi-rong, LI Jian-sheng, et al.

(China Agricultural University, National Maize Improvement Center of China,

Key Lab of Crop Genetics and Breeding of Beijing, Beijing 100094, China)

**Abstract:** Using near infrared reflectance spectroscopy(NIRS) and partial least square(PLS) regression, a model for determination of oil, protein and starch contents in intact kernel was established with the high-oil materials of  $F_{2:3}$  family lines in maize. The results showed that the calibration models for oil, protein, and starch contents with the first derivative + vector normalization, the first derivative + multiplication scattering correction, and the straight line subtraction were the best. The coefficients of determination of the three models for calibration( $R^2_{cal}$ ) were 0.950, 0.973 and 0.976 respectively while the  $R^2_{cv}$  and  $R^2_{val}$  were ranged from 0.918 to 0.948. The root mean square error of estimation, the root mean square error of cross validation, and the root mean square error of prediction(RMSEE, RMSECV, and RMSEP) for oil, protein, and starch contents varied from 0.305% to 0.721%. These results demonstrated that it was feasible to use NIRS as a rapid, accurate, and non-destructive technique to estimate oil, protein, and starch contents in intact kernel of the high-oil maize.

**Key words:** High-oil maize; Kernel components; Near infrared reflectance spectroscopy(NIRS); Calibration models

收稿日期: 2007-04-18

基金项目: 国家自然科学基金项目(30571165)

作者简介: 张 俊(1973-), 博士研究生, 主要从事特用玉米分子育种方面的研究。

李建生为本文通讯作者。Tel: 010-62732422

E-mail: lijiansheng@cau.edu.cn

随着我国经济和科技的发展以及人们生活水平的提高, 玉米生产由单一的粮用转变为饲料、食品工业和加工原料等多元的利用。高油玉米以其优良的加工品质和多种用途而备受人们青睐, 具有广阔的发展前景<sup>[1]</sup>。油分、蛋白质和淀粉是玉米子粒的重要组成部分, 普通玉米子粒的油分、蛋白质和淀粉含量分别为 4% ~ 5%、8% ~ 10% 和 70% ~ 75%<sup>[2]</sup>。高油玉

米子粒的油分含量一般大于 6%，是一种高附加值玉米，它不仅提高了含油量和总能量水平，也相应提高了子粒的蛋白质含量，改善了氨基酸平衡<sup>[3]</sup>。用高油玉米饲喂家畜和家禽，可以显著提高生长水平和生产潜力<sup>[4,5]</sup>。1896 年美国 Illinois 大学的 Hopkins 开始对玉米子粒的油分选择试验，经过 100 多年纯理论研究，含油量由 4.69% (cycle-0) 提高到 20.37% (Cycle-100)<sup>[6]</sup>。我国宋同明教授经过 20 多年的努力，创造了北农大高油(BHO)玉米群体，含油量由 4.71% (cycle-0) 提高到 15.5% (cycle-18)<sup>[7]</sup>。

常用高油玉米分析方法是核磁共振技术(nuclear magnetic resonance, NMR)<sup>[8~10]</sup>，它仅仅能分析子粒油分的含量，无法分析蛋白质和淀粉的含量。近红外反射光谱(near infrared reflectance spectroscopy, NIRS)是 20 世纪 80 年代后期迅速发展起来的一种现代分析新技术，已经成为农作物品质分析的一种重要手段<sup>[11]</sup>。Orman 等<sup>[12]</sup>应用偏最小二乘法建立了普通玉米混合子粒样品的油分、蛋白质和淀粉含量的校正模型。魏良明等<sup>[13]</sup>利用我国常用普通玉米及杂交种样品为材料，建立了测定玉米子粒的 NIRS 定量分析模型，解决大量普通玉米样品的品质分析问题。但由于该模型的光谱覆盖范围、品质含量的浓度梯度局限等问题，无法满足对高油玉米群体完整子粒的油分、蛋白质和淀粉含量进行分析。本研究以我国高油玉米种质(BHO)与普通玉米(B73)组配的高油玉米群体为材料，建立高油玉米完整子粒品质性状的近红外分析模型，解决高油玉米的品质性状分析的问题。

## 1 实验与方法

### 1.1 实验材料

以中国农业大学高油玉米组选育的高油玉米自交系 By804(油分含量为 11.22%)为父本材料、普通玉米自交系 B73(油分含量为 3.71%)为母本材料，组配获得 F<sub>2:3</sub> 家系。分离群体和父母本于 2002 年 5 月在北京昌平试验站种植，采用完全随机区组设计，三次重复。每小区随机自交 5~8 穗。收获是自交果穗自然风干，人工脱粒，剔除病粒、破粒，选果穗中部大小均匀具有代表性子粒用于测定化学成分。最后选取 164 个样品用来测定子粒的油分、蛋白质和淀粉含量。

### 1.2 样品化学成分的测定方法

为消除水分对油分、蛋白质和淀粉含量的影响，每个样品均装在牛皮纸袋中于 45℃ 的 DHG-9146A

型电热恒温鼓风干燥箱中烘 60 h，取出后放在干燥器皿中。样品子粒油分含量采用核磁共振仪(NMR)测定，粗蛋白质含量按国标半微量凯氏法(GB2905-82)测定，粗淀粉含量按国标旋光法(GB5006-85)测定。所有样品均重复平行测定 3 次以上，结果以干基百分数表示。

### 1.3 NIRS 仪器与样品测定方法

实验仪器为德国 BRUKER 公司生产的 VEC-TOR22/N 傅里叶变换近红外光谱仪。将 200 粒完整子粒样品盛于直径为 50 mm 的旋转样品池，在波长 800~2 400 nm 范围内，扫描 64 次。每个样品重复装样 3 次，以消除样品子粒大小、均匀性不一致等因素对光谱的影响，计算其平均光谱。

### 1.4 建立 NIRS 数学模型的方法

根据近红外定标要求，按照分离群体玉米子粒的油分含量正态分布规律，拉开含量梯度设计标样。以样品子粒化学成分含量的大小顺序梯度排列，选取 164 个样品以 2:1 的比例分成校正集和验证集<sup>[14]</sup>，先用校正样品集建立校正模型，再作交叉验证，并用验证样品集进行外部验证，最后根据 R<sup>2</sup><sub>cal</sub>、R<sup>2</sup><sub>cv</sub>、R<sup>2</sup><sub>val</sub>、RMSEE、RMSECV、RMSEP 等指标确定最优模型，采用 OPUS/QUANT4.0 商用光谱定量分析软件对模型进行优化。

### 1.5 衡量校正模型优劣的参数

衡量校正模型优劣的主要参数包括定标决定系数(R<sup>2</sup><sub>cal</sub>)、交叉验证决定系数(R<sup>2</sup><sub>cv</sub>)、外部验证决定系数(R<sup>2</sup><sub>val</sub>)，定标标准误 RMSEE(root mean square error of estimation)，交叉验证标准误 RMSECV (root mean square error of cross validation)，预测标准误 RMSEP (root mean square error of prediction)。对于同一样品集所构建的近红外模型来说，各类决定系数越大，标准误差越小，则表明模型的定标及预测效果越好。

## 2 结果与分析

### 2.1 玉米子粒油分、蛋白质和淀粉含量分析结果

根据各品质性状化学测定相对误差大小，剔除部分相对误差较大和个别界外点(outlier)样品外，实际参与建模的样品数及测定结果见表 1。图 1 用来建立近红外模型的油分、蛋白质和淀粉含量定标和外部验证样品的柱状图，如图所示，供试材料子粒的油分、蛋白质和淀粉含量基本上呈正态分布，表现典型的数量性状变异。表 1 为实验样品的油分、蛋白质和淀粉含量的分析结果，包括校正集样品和验证集样品，由表 1 可见，样品中的油分、蛋白质和淀粉含

量的变幅较大，其中油分含量最大值为 11.12%，接近于高油亲本 By804 的油分含量 11.22%，最小值为 3.65%，接近于普通亲本 B73 的油分含量 3.71%。蛋白质含量的变幅在 8.55%~18.14%，其普通亲本 B73 和高油亲本 By804 的蛋白质含量分别为 10.67% 和 16.46%。淀粉含量的变幅在 55.01%~71.00%，其普通亲本 B73 和高油亲本 By804 的淀粉含量分别为 69.03% 和 58.14%。因而，所选样品的各成分含量能够很好的涵盖其双亲的各成分含量，可以用来建立高油玉米子粒品质性状的近红外光谱校正模型。

表 1 样品子粒的油分、蛋白质和淀粉含量分布  
Table 1 Variations of kernel oil, protein, and starch contents in samples

| 性 状<br>Trait | 油分含量<br>Oil content | 蛋白质含量<br>Protein content | 淀粉含量<br>Starch content |
|--------------|---------------------|--------------------------|------------------------|
| 总样品数         | 162                 | 164                      | 161                    |
| 变幅(%)        | 3.65 ~ 11.12        | 8.55 ~ 18.14             | 55.01 ~ 71.00          |
| 平均值(%)       | 6.95                | 11.77                    | 65.01                  |
| 标准差          | 1.30                | 1.85                     | 2.93                   |
| B73(%)       | 3.71                | 10.67                    | 69.06                  |
| By804(%)     | 11.22               | 16.46                    | 58.14                  |

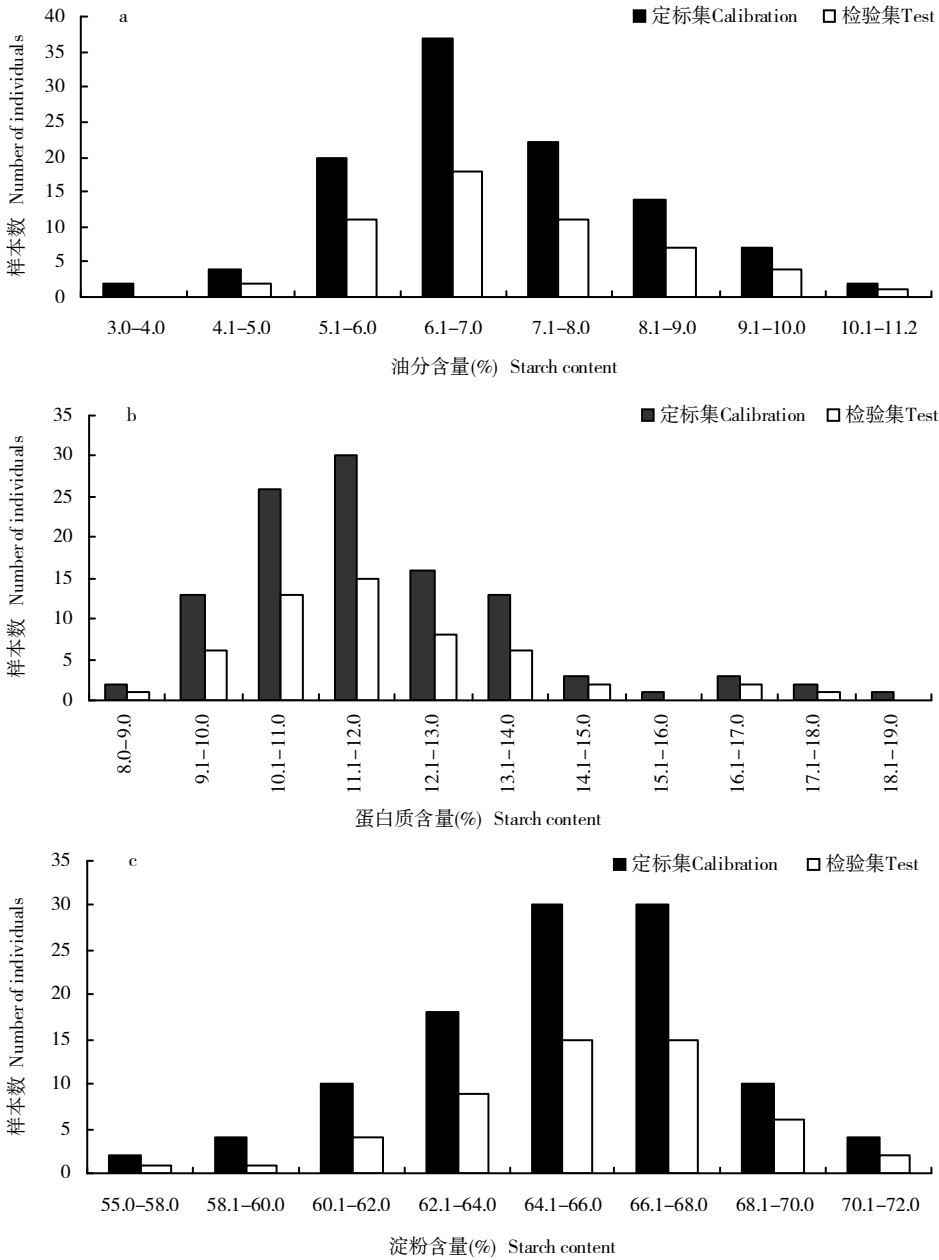


图 1 油分(a)、蛋白质(b)和淀粉(c)含量的定标集和验证集频率分布

Fig.1 Frequency distributions of oil(a), protein(b), and starch(c) contents for calibration and test

2.2 玉米子粒油分、蛋白质和淀粉含量 NIRS 数学模型的建立与优化

根据已经测定的各成分含量及采集的相应各样品的近红外光谱,利用 OPUS/QUANT4.0 软件,建立了高油玉米子粒油分、蛋白质和淀粉含量的校正模型。3 成分的最佳光谱处理方法、谱区范围、所选择光谱数据点和主成分维数见表 2。3 个成分含量的

校正模型的校正决定系数 ( $R^2_{cal}$ ) 分别为 0.950、0.973、0.976;校正误差(RMSEE)为 0.305%、0.321%、0.492%;交叉验证决定系数 ( $R^2_{cv}$ ) 分别为 0.918、0.939、0.942;交叉验证误差 (RMSECV) 分别为 0.379%、0.464%、0.721%,接近化学分析的精确度。结果表明,所建模型的各项决定系数和标准误差可以满足玉米子粒样品分析对准确度的要求。

表 2 玉米子粒油分、蛋白质和淀粉含量的定标、交叉验证和外部验证结果

Table 2 Results of calibration, validation, and test for kernel oil, protein, and starch contents

| 统计参数                  | 油分含量                                | 蛋白质含量                               | 淀粉含量                                |
|-----------------------|-------------------------------------|-------------------------------------|-------------------------------------|
| Statistical parameter | Oil content                         | Protein content                     | Starch content                      |
| 定标光谱                  | 108                                 | 110                                 | 108                                 |
| 外部验证光谱                | 54                                  | 54                                  | 53                                  |
| 光谱预处理                 | FD+VN                               | FD+MSC                              | SLS                                 |
| 选择光谱数据点               | 1 624                               | 1 624                               | 1 431                               |
| 谱区范围                  | 11998.9 ~ 6097.8<br>4601.3 ~ 4246.5 | 11998.9 ~ 5449.8<br>4601.3 ~ 4246.5 | 11998.9 ~ 6097.8<br>5453.7 ~ 4246.5 |
| 主成分维数                 | 5                                   | 9                                   | 9                                   |
| 校正决定系数 $R^2_{cal}$    | 0.950                               | 0.973                               | 0.976                               |
| 交叉验证决定系数 $R^2_{cv}$   | 0.918                               | 0.939                               | 0.942                               |
| 外部验证决定系数 $R^2_{val}$  | 0.948                               | 0.931                               | 0.946                               |
| 校正误差                  | 0.305                               | 0.321                               | 0.492                               |
| 交叉验证误差                | 0.379                               | 0.464                               | 0.721                               |
| 外部验证误差                | 0.307                               | 0.429                               | 0.645                               |

注:FD+VN:一阶导数+矢量归一化;FD+MSC:一阶导数+多元散射校正;SLS:直线相减法;RMSEE:均方根误差;RMSECV:交叉验证均方差;RMSEP:预测均方根误差。

Note: FD+VN: first derivative + vector normalization; FD+MSC: first derivative + multiplication scattering correction; SLS: straight line subtraction; RMSEE: root mean square error of estimation; RMSECV: root mean square error of cross validation; RMSEP: root mean square error of prediction.

2.3 校正模型对玉米子粒油分、蛋白质和淀粉含量预测效果分析

对所建立的校正模型的实际预测效果进行外部验证的结果列于表 2。可以看出,油分、蛋白质和淀粉含量的外部验证的决定系数( $R^2_{val}$ )分别为 0.948、0.931、0.946,外部验证误差(RMSEP)分别为 0.307%、0.429%、0.645%,外部验证的决定系数较高而误差

较低,说明所建立的校正模型的质量较高。相关分析的结果表明(图 2),油分、蛋白质和淀粉含量的预测值与化学值间的相关系数分别为 0.965、0.941、0.936。结果表明预测值与化学值比较接近,两者具有很好的线性关系,具有类似的准确性和测定精度,可用来对高油玉米群体的其他样品进行实际预测。

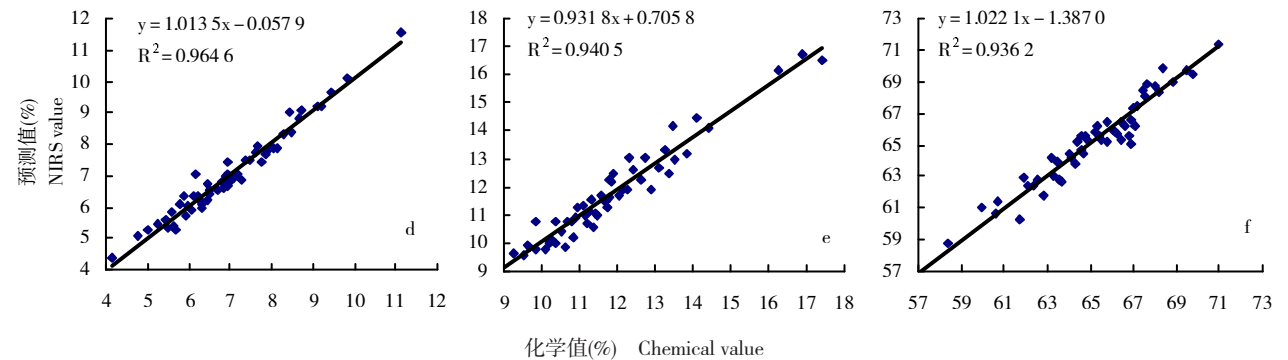


图 2 油分(d)、蛋白质(e)和淀粉(f)含量的化学值与预测值的相关系数

Fig.2 Correlations between measured and predicted values for oil(d), protein(e), and starch(f) contents



### 3 结 论

现代近红外光谱分析技术是一种无损分析技术,具有分析速度快、效率高、成本低的优点,最重要的是样品不需预先处理,测定简单,对玉米育种和生产来说,不用破坏完整子粒,可以快速大批量地同时测定样品的多种品质性状,而且样品经品质分析后仍可以继续繁殖,满足实验和育种的需要。近红外光谱特征反映了样品的结构和组成特征,而作物的品质特性决定于子粒结构和组成成分。由于作物的品质性状是可遗传的,因此玉米杂交后代的近红外光谱特征可以作为选择的重要指标。本实验所用的高油玉米分离群体的子粒油分含量没有发现明显的超亲分离<sup>[15]</sup>,因而根据子粒油分的浓度梯度选择的标样,其光谱的范围可以涵盖整个分离群体的光谱。本实验所建立的模型提供了同时分析高油玉米子粒油分、蛋白质和淀粉含量的方法,对于同一样品可以同时分析其油分、蛋白质和淀粉的含量。

近红外数学模型的优劣与定标过程中标准样本的多少也有关系。数量太少不足以反映被测样本群体常态分布规律,数量太多,工作量大。徐坤等<sup>[16]</sup>研究认为定标样本数在 70 以上,定标结果基本变异不大,预测结果符合化学分析,定标决定系数可达 0.93~0.99,预测标准误差较小。本实验的油分、蛋白质和淀粉三成分的定标样本数分别为 108、110、108,定标的决定系数均在 0.95 以上,保证了分析结果的准确性。

魏良明以我国常用的自交系、杂交种样品为材料,建立了玉米完整子粒粗蛋白质、粗淀粉和油分的校正模型,其外部验证决定系数分别为 0.96、0.87、0.95,预测标准误分别为 0.31、0.88、0.33。本实验的蛋白质和油分含量的近红外模型的外部验证决定系数及预测标准误与其相当,淀粉含量的近红外模型的外部验证决定系数(0.946)高于其淀粉含量的近红外模型的外部验证决定系数,而预测标准误(0.645)则低于其外部验证决定系数的预测标准误。本实验所建立油分和蛋白质含量的近红外模型质量与其相

当,而淀粉含量的近红外模型具有较高的质量。

#### 参考文献:

- [1] 宋同明. 我国高油玉米育种及其发展趋势[J]. 中国农业科技导报, 2001, 3(3): 40-43.
- [2] Boyer C D, Hannah L C. Kernel mutants of corn: Specialty Corns[M]. CRC Press Inc, Boca Raton, Florida, Hallauer AR (eds), 2001.
- [3] Lambert R J. High-Oil Corn Hybrids: Specialty Corns[M]. CRC Press Inc, Boca Raton, Florida, Hallauer AR (eds), 2001.
- [4] Benitez J A, Gernat A G, Murillo J G, Araba M. The use of high oil corn in broiler diets[J]. Poultry Sci., 1999, 78: 861-865.
- [5] O'Quinn P R, Nelssen J L, Goodband R D, et al. Nutritional value of a genetically improved high-lysine, high-oil corn for young pigs[J]. Animal Sci., 2000, 78: 2144-2149.
- [6] Dudley J W, Lambert R J. 100 generations of selection for oil and protein in corn[J]. Plant Breeding Review, 2004, 24: 97-110.
- [7] Song T M, Chen S J. Long term selection for oil concentration in five maize populations[J]. Maydica, 2004, 49(1): 9-14.
- [8] Alexander D E, Silvela S L, Collins F I, Rodgers R C. Analysis of oil content of maize by wideline NMR[J]. Amer oil chem Soc., 1967, 44: 555-558.
- [9] Bauman L F, Conway T F, Watson S A. Heritability of variations in oil content of individual corn kernels[J]. Science, 1963, 139: 498.
- [10] 宋同明. 脉冲核磁共振仪(Pulsed NMR)对作物种子含油量的快速测定[J]. 作物学报, 1989, 15(2): 160-166.
- [11] Ciureczak E W. Use of near infrared spectroscopy in cereal products[J]. Food Testing and Analysis, 1995, 5: 35-39.
- [12] Orman B A, Schumann J R A. Comparison of near-infrared spectroscopy calibration methods for the prediction of protein, oil, and starch in maize grain[J]. Journal of Agricultural Food Chemistry, 1991, 39: 883-888.
- [13] 魏良明, 姜海鹰, 等. 玉米杂交种品质性状的近红外光谱分析技术研究[J]. 光谱与光谱分析, 2005, 25(9): 1404-1407.
- [14] Campbell M, Brumm T J, Glover D V. Whole grain amylose analysis in maize using near-infrared transmittance spectroscopy[J]. Cereal Chemistry, 1997, 74: 300-303.
- [15] Song X F, Song T M, Dai J R, Rocheford T R, Li J S. QTL mapping of kernel oil concentration with high-oil maize by SSR markers[J]. Maydica, 2004, 49(1): 41-48.
- [16] 徐 坤, 刘鹏起, 等. 近红外光谱分析技术及应用[J]. 莱阳农学院学报, 2001, 18(3): 237-240.

(责任编辑: 李万良)