ESTABLISHMENT OF CALIBRATION MODELS OF NIRS ANALYSIS FOR COTTONSEED OIL AND PROTEIN CONTENTS AND MAPPING RELATED QTLs IN UPLAND COTTON *(Gossypium hirsutum L.)*

Xinlian Shen
Peng Xu
Tingting Guo
Xianggui Zhang
Wanchao Ni
Jiangsu Academy of Agricultural Sciences
Nanjing, China

Abstract

This study was primarily conducted to investigate the feasibility of near-infrared reflectance spectroscopy (NIRS) in the determination of cottonseed oil and protein contents. A total of 112 cotton lines were used to construct the calibration models of cottonseed oil and protein contents determined by NIRS, and the coefficients of determination R^2 of these two models were 0.9228 and 0.9249, respectively. The results showed that the NIRS methods tested in this study have relatively high accuracy in prediction test for cottonseed oil and protein contents. Two F_2 populations from the crosses DPLSR3 × Yancheng115 (Pop A) and Humcopedixi14wtr5 × Yancheng90-110 (Pop B) were employed as experimental materials to identify QTLs associated with cottonseed oil and protein contents. Four significant QTLs for oil and protein content in chr11, chr13 and chr19 were identified in two populations. A stable QTL for cottonseed oil content was mapped between the SSR maker DC20120 and NAU2893 on chromosome 13 in both F_2 and F_3 generations from population A, explaining 9.21% and 12.01% of the phenotypic variation, respectively. This study not only established a fast and non-destructive methodology to facilitate the selection of cottonseed nutrient traits, but also identified some molecular markers and provided a useful base for genetic manipulation of these seed nutritional traits in breeding programs.

Introduction

Cottonseed, simultaneously produced with cotton fiber in cotton industry, is a rich source of edible oil and protein. It was experimentally proven that the oil content in cottonseed kernel could be reach at about 35% and protein content could be around 45-50%. Cottonseed oil has long been considered as a healthy vegetable oil due to its flavor stability and superior nutritive qualities. It is anticipated that a large and reliable supply of cottonseed would be required in order to develop and maintain a market for an edible products, and meet other market area such as cottonseed oil for use as a biodiesel. However, expansion of cotton product area is restricted because of the issue of food security in the wide of world, especially in China. In view of this, improvement of cottonseed nutrient content should be considered as an important breeding goal except for yield and fiber quality.

Traditional methods for measuring nutrient contents in cottonseed are costly, time-consuming and instruments-restricted, and the phenotyping obstacles often drag the legs for cotton breeding. With the development of computer science and chemometrics, the available applications of near-infrared spectroscopy (NIRS) technique become more popular and attract more attention. It has been widely used in the past several years to fast and efficient determination of starch, protein, fatty acid, amino acid, total sugar contents in several

other crops (Campbell et al., 1997; Xie et al., 2003; Delwiche, 1998; Wu et al., 2002). Because it reveals information related to the vibration behaviors of molecular bonds, NIRS therefore can provide details of the varieties of molecules present in the seed. As an indirect selection tool, NIRS provides a fast and non-destructive methodology, which requires little or no sample pretreatments, thus facilitating and speeding up the selection of cottonseed nutrient traits.

Few researches of mapping QTLs related to cottonseed nutrient traits have been reported. Song et al (2007) identified two significant QTL for cottonseed oil and protein contents using a population of 140 BC_1S_1 lines developed from a cross between "TM-1" (*G. .hirsutum*) and "Hai7214" (*G. .barbadense*) based on 2 years of phenotypic data. However, it is difficult to utilize excellent loci from sea island germplasms (*G. barbadense)* because of distorted segregation in populations between interspecific crosses. Digging alleles associated with cottonseed nutrient traits directly from upland cotton *(G. .hirsutum.)* genomes would be more effective in improvement of cottonseed nutrient content by marker-assisted selection in upland cotton.

The objectives of this study are to determine whether NIRS can substitute for conventional methods for analysis of cottonseed oil and protein contents; and to identify stable QTLs related to cottonseed oil and protein contents from upland cotton so as to facilitate application of MAS in improving seed nutrient quality.

Materials and Methods

Plant Materials

NIRS analysis: The samples were chosen from a wide range of breeding materials. Totally, 112 germplasm resources were planted for the establishment of NIRS models in Lishui experiment station, Jiangsu Academy of Agricultural Sciences (JAAS), China in 2008. Plots were 4 m with the rows spaced 80 cm apart. Seed cotton was harvested by bulking bolls from the interior middle of the plant within each plot, and their seeds were ginned and acid-delinted. Seeds were dried at 37°C for 24 hours in an oven to equalize moisture contents among samples.

QTL mapping: Two high oil and low protein content parents (DPLSR3 and Humcopedixi14wtr5), and two low oil and high protein content parents (Yancheng115 and Yancheng90-110), were used as parents to construct two populations for molecular tagging of QTLs for cottonseed nutrition quality. Two F_2 mapping populations with 151 individuals were derived from the crosses of DPLSR3×Yancheng115 (Pop A) and Humcopedixi14wtr5×Yancheng90-110 (Pop B), respectively. In 2009, all F_2 plants were grown in Lishui experiment station, JAAS, China, and were self-pollinated to produce F_2 seeds for F_3 generations. All F_3 lines as well as their parents were planted in 2010.

NIRS Analysis

Cottonseed oil and protein contents for 112 upland cotton germplasms were determined by chemical analysis. Cottonseed oil content was analyzed by Soxhlet extractor method referred to AOAC (1984). Cottonseed protein content was estimated through total nitrogen content using Microkjeldahl determination method referred to Humphries (1956). Seed protein content was calculated by multiplying the nitrogen content with a factor 6.25 and the result was expressed as percentage. NIR spectrometer (DA7200 Perten instrument) was employed in this research with the spectral range from 930nm to 1650nm. The whole measurement generally consists of the following three steps (Cheewapramon, 2007).

[1] Collect the spectral data of samples in calibration set by NIR spectrometer at room temperature and humidity;

[2] Pre-process spectral data before modeling, and then build calibration models using calibration set with known analyzed concentration of samples. This step was accomplished by the technician from the Perten company;

[3] Evaluate calibration model using validation set.

QTL Mapping

MapMaker/EXP 3.0b (Lander et al. 1987) was employed to construct the linkage map. The parameters included the kosambi function (Kosambi, 1944). The linkage criteria were a LOD score greater than 3.0 and the maximum genetic distance of 50cM. Composite interval mapping (CIM) (Zeng, 1994) analysis was performed using QTL cartographer V2.5.

Results

Establishment of Calibration Models of NIRS Predicting Cottonseed Oil and Protein Contents

A total of 112 cotton samples were analyzed on oil and protein contents by referred method. The mean value of two repetitions was taken as the chemical value of this sample. According to the distribution of total cottonseed oil and protein contents, 87 samples were chose as calibration set to construct the model of NIRS predicting cottonseed oil content and 80 samples for model of cottonseed protein content; the remaining samples made up the validation set. The oil content in calibration set ranged between 25.6-40%, and the protein between 35.8-50.8%, covering almost available variability for oil and protein content in this species.

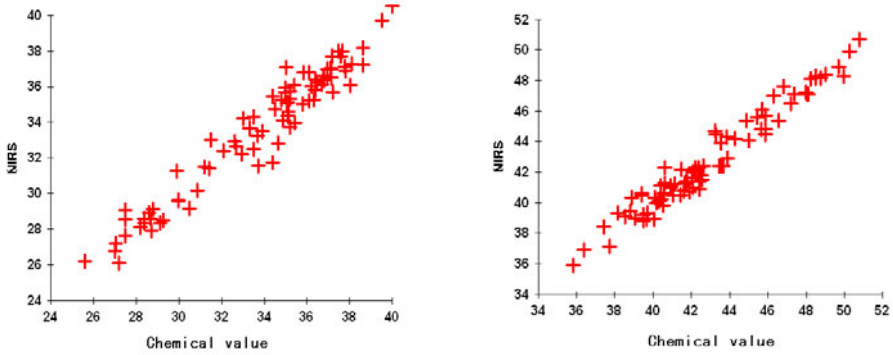

Figure 1. Linear relationship between chemical and NIRS values of oil and protein contents in calibration set.

Chemometrics methods were used in calibration set to construct the model by comparing the NIRS spectrum data and chemical values. Quantitative analysis showed that NIRS predicted values of cottonseed oil content were consistent significantly with chemical analysis results which was supported by R^2=0.944 and SEC=0.0346; and for protein content, the coefficient of determination R^2 was 0.9543 and SEC was 0.0332 (Figure 1).

The validation set was further used to validate the model constructed above. The coefficients of determination R^2 were 0.9228 for oil content and 0.9249 for protein content, and the values of SEP were 0.0351 for oil content and 0.0356 for protein content, respectively. These results showed that the NIRS methods established in this study had relatively high accuracy in prediction test for cottonseed oil and protein contents.

QTL Mapping of Cottonseed Oil and Protein Contents

Construction of intraspecific linkage maps: A total of 1221 SSR markers covering the whole cotton genome were selected to screen polymorphism between parents. In Pop A, 52 loci showed polymorphic, among which 27 loci were mapped into 9 linkage groups. They were assigned to 8 different chromosomes based on a referenced cotton genetic map (Guo et al 2007). The remaining 1 linkage group could not be associated with any chromosome. The genetic maps spanned 364.2cM with an average distance of 12.7cM between adjacent markers (Fig.2). In Pop B, 42 loci showed polymorphic, among which 18 loci were assigned to 8 linkage groups, spanning 192.8cM with an average genetic distance of 10.7cM between adjacent markers (Fig.3).

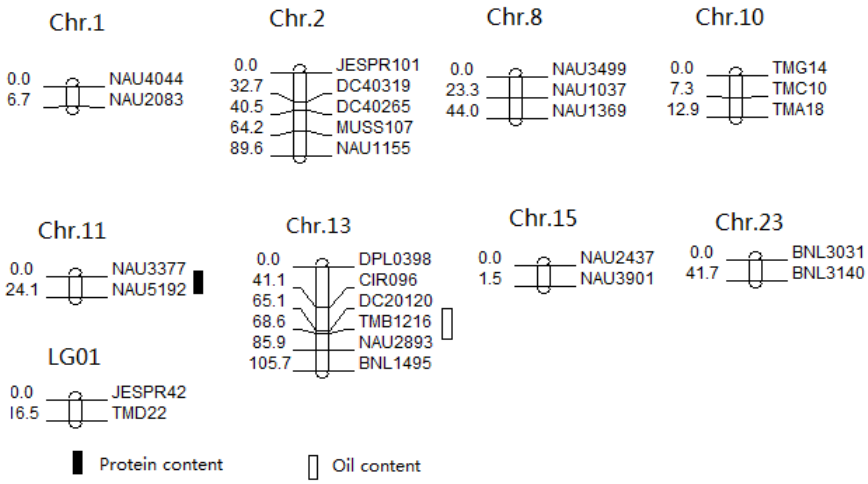

Figure 2. Genetic map of PopA and the detected QTLs for nutrient quality.

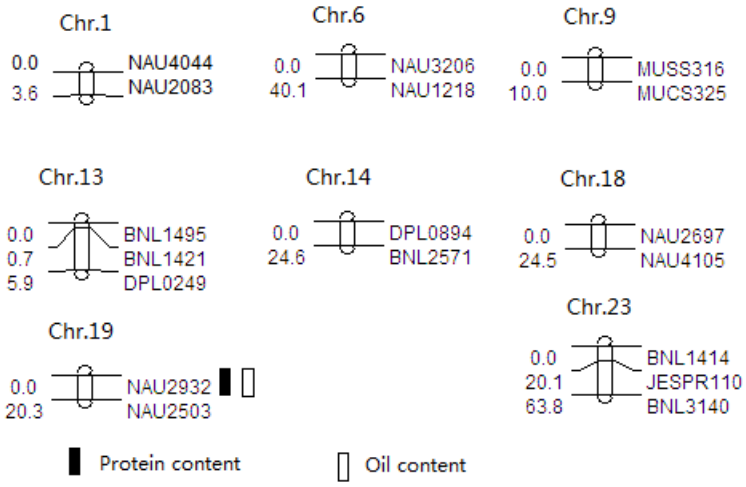

Figure 3. Genetic map of PopB and the detected QTLs for nutrient quality.

QTL mapping for cottonseed oil and protein contents: Based on the linkage maps constructed with simple sequence repeat (SSR) makers from these two populations, composite interval mapping was employed to identify quantitative traits loci (QTL) associated with cottonseed oil and protein contents by using the software Cartographer (V2.5). The list of the QTLs identified in Pop A and Pop B were presented in Table 1. Their most likely positions on the linkage map were shown in Fig2, Fig3. Altogether, 4 QTLs significant for cottonseed nutrition quality were detected. Among them, two QTLs were for oil content, two for protein content.

In Pop A, a QTL for cottonseed oil content, *qOC-13-1* was identified between the SSR maker DC20120 and NAU2893 on Chr.13 in both F_2 and F_3 generations, explaining 9.21% and 12.01% of the phenotypic variations, respectively. The molecular marker identified here could be a means for genetic manipulation of cottonseed oil content in breeding programs. Another QTL for protein content was detected between NAU3377 and NAU5192 on Chr.11 only in F3 generation, explaining 9.85% of phenotype variation.

In Pop B, a significant QTL *qOC-19-1* for oil content was mapped between the marker NAU2932 and NAU2503, explaining 24.67% of the phenotypic variation in F_3 generation. The genotype of Humcopedixi14wtr5 was in the direction of increasing oil content. Interestingly, another significant QTL *qPC-19-1* for protein content was also mapped in this maker interval, explaining 10.1% of the phenotypic variation in F_3 generation. Yancheng90-110 was genotyped as increased protein content.

Table 1. Related parameters of cottonseed oil and protein contents QTL in two populations.

QTL	Pop	Generation	Chr.	Marker interval	Range (cM)	*LOD*	PV	Additive	Dominance
qOC-13-1	A	F_2	Chr.13	DC20120-NAU2893	67.8-79.6	3.1	9.21%	-1.8534	0.1905
	A	F_3	Chr.13	DC20120-NAU2893	60.1-75.7	3.9	12.01%	-1.7787	0.1280
qPC-11-1	A	F_3	Chr.11	NAU3377-NAU5192	16.8-22.5	4	9.85%	1.039	0.4934
qOC-19-1	B	F_3	Chr.19	NAU2932-NAU2503	0-12.9	4.2	24.67%	1.7424	1.8780
qPC19-1	B	F_3	Chr.19	NAU2932-NAU2503	0-12.4	4.2	10.1%	-1.0442	-1.172

Discussion

As one of the various technologies for the evaluation of crop nutrient quality, NIRS technique has its advantage in quality analysis, such as money- and time- saving and non-destructive. The spectral measurement for one individual sample could be finished in one minute, and multiple samples could be analyzed by one spectral measurement and built models, which is particularly convenient for multi-indexes analysis. Compared with traditional methods, NIRS has low cost because of no demand for other chemical reagents except for the electrical consumption. Further, NIRS will not bring any side effects to the seeds under testing which made itself easy for the evaluation of crop nutrient quality.

A NIRS predication equation should be constructed from samples with all available variability. In this study, the oil and protein content of cottonseed had a wide distribution from 25.6% to 40% for oil and from 35.8% to 50.8% for protein as determined by chemical methods. Therefore, even we didn't adopt a large population for developing a NIRS predication equation (only 87 samples for oil content and 80 samples for protein content). The results from validation set data still showed considerably consistent with that from chemical analysis.

Cotton is an important source of edible oil and protein meal. But little attention has been paid to breeding for high nutrition quality varieties. Song et al. (2007) firstly reported study on QTL mapping for cottonseed quality

with an interspecies population from *G. barbadense* and *G. hirsutum* cross. This is the first report on QTL mapping of cottonseed quality traits in upland cotton. Four significant QTLs for oil and protein content in chr11, chr13 and chr19 were identified in two populations. Comparing to Song's results based on a dense genetic map from a cross of *G. barbadense* and *G. hirsutum*, there is no common QTL either between *G. barbadense* and *G. hirsutum* cultivars. All these results were recognized that cottonseed nutritional quality has complex genetic base.

Acknowledgement

Partial funding in support of this project by the JAAS CX (08)124 and JAAS CX(11)1021 was greatly appreciated. Senior research fellow from Perten company are gratefully acknowledged for conducting the analysis of cottonseed protein and oil concentration in this study.

References

AOAC. 1984. Official methods of analysis of the association of official analytical chemists (14th edn.), pp. 441-442.

Campbell M.R., T.J. Brumm and D.V. Glover. 1997. Whole grain amylose analysis in maize using Near-infrared transmittance spectroscopy. Cereal Chemistry, 74(3):300-303.

Cheewapramon P. 2007, Use of near-infrared spectroscopy for qualitative and quantitative analyses of grains and cereal products. Dissertation in food science and technology for degree of doctor of philosophy, university of Nebraska. - Lincoln, USA, 11-14.

Delwiche S.R. 1998. Protein content of single kernels of wheat by near-infrared reflectance spectroscopy. Journal of Cereal Science, 27:241-254.

Guo W.Z., C.P. Cai, C.B. Wang., Z.G. Han., X.L. Song, K. Wang, X.W. Niu, C. Wang, K.Y. Lu, B. Shi, T.Z. Zhang. 2007. A micro-satellite-based, gene-rich linkage map reveals genome structure, function and evolution in *Gossipium*. Genetics, 176:527-541.

Humphries E.C. 1956. Mineral components and ash analysis. In: Modern methods of plant analysis. Springes-verlag, Berlin, 468-502.

Kosambi D.D. 1944. The estimation of map distances from recombination values. Ann Eugen, 12:172-175.

Lander E.S., P. Green, J. Abrahamson, A. Barlow, M.J. Daly, S.E. Lincoln. 1987. MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics, 1:174-181.

Song X. L., T. Z. Zhang. 2007. Identification of quantitative traits loci controlling seed physical and nutrient traits in cotton. Seed Science Research, 17:243-251.

Wu J.G., C.H. Shi, H.Z. Zhang, L.J. Fan. 2002. Study on analysis of oil content in intact rapeseed with few sample size by near infrared reflectance spectroscopy. *Acta Agronomica Sinica,* 28(3):421-425.

Xie F., F.E. Dowell, X.S. Sun. 2003. Comparison of Near-infrared reflectance spectroscopy and texture analyzer for measuring wheat bread changes in storage. Cereal Chemistry, 80(1):25-29.

Zeng Z.B. Precision mapping of quantitative trait loci. Genetics, 1994 136:1457-1468.