

## 적외선 분광스펙트럼 및 기체크로마토그래피 분석 데이터의 다변량 통계분석을 이용한 대두 종자 지방산 함량예측

안명숙 · 지은이 · 송승엽 · 안준우 · 정원중 · 민성란 · 김석원

### Simultaneous estimation of fatty acids contents from soybean seeds using fourier transform infrared spectroscopy and gas chromatography by multivariate analysis

Myung Suk Ahn · Eun Yee Ji · Seung Yeob Song · Joon Woo Ahn · Won Joong Jeong · Sung Ran Min · Suk Weon Kim

Received: 9 March 2015 / Revised: 21 March 2015 / Accepted: 21 March 2015  
© Korean Society for Plant Biotechnology

**Abstract** The aim of this study was to investigate whether fourier transform infrared (FT-IR) spectroscopy can be applied to simultaneous determination of fatty acids contents in different soybean cultivars. Total 153 lines of soybean (*Glycine max* Merrill) were examined by FT-IR spectroscopy. Quantification of fatty acids from the soybean lines was confirmed by quantitative gas chromatography (GC) analysis. The quantitative spectral variation among different soybean lines was observed in the amide bond region (1,700 ~ 1,500  $\text{cm}^{-1}$ ), phosphodiester groups (1,500 ~ 1,300  $\text{cm}^{-1}$ ) and sugar region (1,200 ~ 1,000  $\text{cm}^{-1}$ ) of FT-IR spectra. The quantitative prediction modeling of 5 individual fatty acids contents

(palmitic acid, stearic acid, oleic acid, linoleic acid, linolenic acid) from soybean lines were established using partial least square regression algorithm from FT-IR spectra. In cross validation, there were high correlations ( $R^2 \geq 0.97$ ) between predicted content of 5 individual fatty acids by PLS regression modeling from FT-IR spectra and measured content by GC. In external validation, palmitic acid ( $R^2=0.8002$ ), oleic acid ( $R^2=0.8909$ ) and linoleic acid ( $R^2=0.815$ ) were predicted with good accuracy, while prediction for stearic acid ( $R^2=0.4598$ ), linolenic acid ( $R^2=0.6868$ ) had relatively lower accuracy. These results clearly show that FT-IR spectra combined with multivariate analysis can be used to accurately predict fatty acids contents in soybean lines. Therefore, we suggest that the PLS prediction system for fatty acid contents using FT-IR analysis could be applied as a rapid and high throughput screening tool for the breeding for modified Fatty acid composition in soybean and contribute to accelerating the conventional breeding.

M. S. Ahn · E. Y. Ji · W. J. Jeong · S. R. Min  
한국생명공학연구원 식물시스템연구센터  
(Plant systems Research Center, Korea Research Institute of  
Bioscience and Biotechnology (KRIBB), 125 Gwahak-ro,  
Yuseong-gu, Daejeon 305-806, Korea)

S. Y. Song  
제주대학교 일반대학원 생명공학과  
(Faculty of Biotechnology, Jeju National University, Jeju Special  
Self-Governing Province 690-756, Korea)

J. W. Ahn  
한국원자력연구원 첨단방사선연구소  
(Advanced Radiation Technology Institute, Korea Atomic Energy  
Research Institute (KAERI), 29 Geumgu street, Jeongeup,  
Jeollabukdo, 580-185, Korea)

S. W. Kim (✉)  
한국생명공학연구원 미생물자원센터  
(Microbiological Resource Center, Korea Research Institute of  
Bioscience and Biotechnology (KRIBB), 125 Gwahak-ro,  
Yuseong-gu, Daejeon 305-806, Korea)  
e-mail: kimsww@kribb.re.kr

**Keywords** FT-IR (Fourier transform infrared spectroscopy), GC (Gas-chromatography), PCA (principal component analysis), PLS-DA (partial least square - discriminant analysis), RMSEP (root mean square error of prediction),  $R^2$  (correlation coefficient)

#### 서론

대두는 세계 3대 작물의 하나로 단백질, 섬유소, 미량영

양성분 및 다양한 기능성 성분의 공급원이며 또한 대표적인 유지 작물 중 하나로서 대두의 기름은 인류 및 동물의 매우 중요한 영양원이다(Mateos-Aparicio *et al.* 2008). 따라서 지방산 조성을 변화시켜서 기능성을 향상시키는 것은 대두의 육종가에게는 매우 중요한 육종 목표로 자리매김을 하고 있다. 이와 같은 대두의 지방산 조성 변형 신품종을 개발하기 위한 돌연변이 육종은 오래 전부터 연구가 이루어져왔다(Rahman *et al.* 1994). 특히 리놀레익산이나 리노레닉산의 함량은 줄이고 불포화 지방산인 올레익산의 함량을 증가시키는 것은 산화방지를 통한 대두 기름의 보존기한 연장은 물론 영양학적 측면에서 기능성 향상에 매우 중요하다(Ascherio and Willett 1997). 또 다른 측면에서 포화지방산(팔미틱산이나 스테아릭산)의 함량 증가는 고흡지방을 이용하는 식품산업에서 매우 중요한 특성이다(Fehr *et al.* 1991). 그러나 이와 같은 다양한 신품종을 개발하기 위해서는 육종라인으로부터 대사체 특성 분석이 필수적으로 이루어져야 한다. 따라서 지방산의 함량 변화를 보다 빠르게 분석할 수 있고, 비용을 절감할 수 있는 효과적인 분석방법의 개발이 필수적으로 요구되고 있다.

적외선분광분석법은 다수의 시료를 단시간에 분석할 수 있는 비파괴적인 분석방법으로 농업이나 식품산업에서 시료의 특성평가를 위해 사용되고 있는 분석방법이다(Fontaine *et al.* 2001; Alander *et al.* 2013). 현재 시료의 지방산 분석에는 기체크로마토그래피 분석방법이 대표적으로 사용되고 있다. 기체크로마토그래피 분석방법은 지방산 조성이나 양적 변화를 조사하는데 있어 가장 신뢰성이 높고 정확한 분석방법이다(Patil *et al.* 2010; Quinones-Islas *et al.* 2013). 복잡한 기체크로마토그래피 분석 절차를 대체하기 위해 오래전부터 간편한 지방산 유도체화 방법이 개발되고 있지만 보다 신속한 지방산의 조성 분석을 위해서는 비파괴적 분광분석방법의 개발 또한 필요하다. 현재까지 분광학적 분석방법을 통해 참깨(Sato *et al.* 2003), 땅콩(Tilman *et al.* 2006), 유채(Wu *et al.* 2006) 및 대두(Kovalenko *et al.* 2006<sup>b</sup>) 등 다양한 유지작물에서 지방산의 조성 연구가 시도되고 있다.

최근 분광분석법과 다변량통계분석 기법을 연계하여 복잡한 데이터로부터 유용한 생물학적인 정보를 추출하는 기술들이 개발되고 있다. 이러한 기술들은 크로마토그래피를 이용한 분석방법에 비해 상대적으로 정밀도는 낮지만 매우 손쉽게, 빠른 속도로 시료 분석이 가능한 장점을 가지고 있다. 또한 이러한 접근 방법은 식품, 농산물 및 식물에서 유용한 기능성 성분들의 함량예측 수단으로 활용이 이루어지고 있다. 이러한 함량 예측 모델링에 사용되고 있는 통계적 알고리즘은 Partial least squares (PLS), artificial neural networks (ANN), 그리고 support vector machines (SVM) 분석법이 활용되고 있다. 현재까지 근적

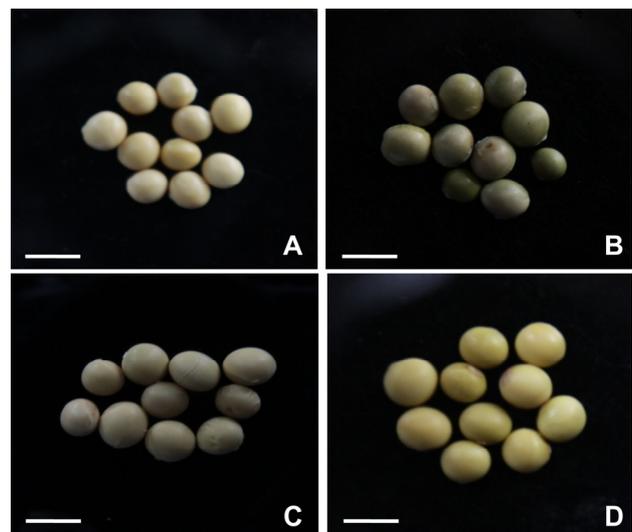
외선 및 적외선 분광분석법을 이용하여 다양한 시료에서 아미노산(Kovalenko *et al.* 2006<sup>a</sup>), 플라보노이드(Soriano *et al.* 2007), 페놀성 화합물(Fernandez and Agosin 2007; Chen *et al.* 2008; Bellincontro *et al.* 2012), 라이코펜 및 카로티노이드(Baranska *et al.* 2007) 그리고 지방산(Kovalenko *et al.* 2006<sup>b</sup>) 성분들의 함량예측이 성공적으로 이루어짐이 보고된 바 있다. 그러나 아직까지 적외선분광분석법을 이용하여 대두 종자로부터 지방산 성분의 함량 예측이 시도된 바 없다.

따라서 본 연구는 총 153종의 대두 종자시료로부터 적외선 분광분석 스펙트럼 분석을 실시하고 이들 시료의 기체크로마토그래피 분석법을 이용한 각각의 지방산 함량 데이터를 PLS 회귀 예측 모델링을 통하여 정확도가 높은 지방산 성분별 함량예측 모델링을 개발하고자 한다. 이 PLS 회귀 예측 모델링은 복잡한 지방산 분석 과정을 거치지 않고 지방산 성분별 함량 예측이 가능하여 다수의 대두 자원으로부터 조성이 변환된 새로운 품종 육성에 크게 기여할 수 있을 것으로 사료된다.

## 재료 및 방법

### 식물재료 및 전세포추출물 제조

본 연구에서는 총 153종의 대두(*Glycine max* Merrill) 종자를 사용하였다. 대두 종자는 김선태교수(부산대학교)와 정순천박사(한국생명공학연구원)로부터 각각 2011년에 채종하여 건조된 종자를 제공받아 본 연구에 사용하였다. 각 대두는 계통별로 종피의 색이 황색에서 검정색까지 다양하였으며(Fig. 1), 이들 다양한 자원집단을 대상으



**Fig. 1** Plant materials used for this study. A: GB1; B: GB48; C: GB106; D: GK4. Scale bars represent 1 cm

로 하여 본 연구를 수행하였다. 각 대두 종자는 막자와 막자사발을 이용하여 고운 가루로 만들었으며 이 파우더는 실험에 사용하기 직전까지  $-70^{\circ}\text{C}$  초저온 냉동고에 보존하였다.

대두 파우더로부터 전세포추출물을 제조하기 위하여 파우더 5 mg을 Eppendorf tube (1.5 mL)로 옮겨준 다음 추출용매(20% methanol)를 100  $\mu\text{L}$  첨가하였다. 시료를 잘 섞어준 다음  $50^{\circ}\text{C}$  항온수조에서 10분간 주기적으로 용액을 혼합해 주면서 반응을 시켜주었다. 반응종료 후 13,000 rpm에서 5분간 원심분리를 수행하여 상층액을 조심스럽게 회수하였다. 회수한 상층액은 새 튜브로 옮긴 뒤 찌꺼기가 혼입되지 않도록 다시 한번 원심분리 과정을 반복하였다. 추출이 완료된 전세포추출물은 사용직전까지  $-20^{\circ}\text{C}$  냉동고에 보존하면서 실험을 수행하였다.

#### 적외선 분광 스펙트럼 조사 및 다변량 통계분석

FT-IR (Fourier transform infrared) 스펙트럼 조사는 Tensor 27 (Bruker Optics GmbH, Ettlingen, Germany)을 사용하였고, DTGS (deuterated triglycine sulfate) 검출기로 분석하였다. 추출된 각각의 시료 5  $\mu\text{L}$ 를 384-well silicon plate에 분주하여,  $37^{\circ}\text{C}$  hot plate에서 약 20분간 건조하였다. 건조된 silicon plate는 Tensor 27에 장착된 HTS-XT (Bruker Optics GmbH) 고효율 자동화 장치를 이용하여 스펙트럼을 조사하였다. 각 시료의 스펙트럼은 총  $4,000\text{--}400\text{ cm}^{-1}$  범위에서 그리고  $4\text{ cm}^{-1}$  간격으로 총 128회 반복 측정된 평균 스펙트럼을 분석에 사용되었다. 각 시료의 FT-IR 스펙트럼은 통계적 분석을 위해 각각 3반복 측정하였다. FT-IR 스펙트럼 조사 및 데이터 변환에 사용된 프로그램은 Bruker에서 제공하는 OPUS Lab (ver. 7.0, Bruker Optics Inc.)를 사용하였다. FT-IR 스펙트럼 데이터의 다변량 통계분석을 위해 먼저 FT-IR 스펙트럼 데이터의 baseline 교정, normalization 및 mean centering 등 스펙트럼의 전처리 과정을 R 프로그램(version 2.15.0)을 이용하여 수행하였다. Baseline 교정을 위해 FT-IR 스펙트럼 분석 영역의 양 끝점( $1,800\text{--}800\text{ cm}^{-1}$ )의 흡광도를 0으로 조정하였으며 실험상의 오차를 최소화하기 위하여 각 스펙트럼을 동일 면적으로 normalization 하였다. 이후 데이터의 mean centering 과정을 거쳐 2차 미분을 수행한 다음 전처리가 완료된 스펙트럼 데이터를 다변량 통계분석 분석을 위한 표준화된 데이터로 사용하였다.

다변량통계분석은 가공된 적외선분광 스펙트럼 데이터( $1,800\text{ to }800\text{ cm}^{-1}$ )를 사용하였으며 NIPALS 알고리즘(Wold 1966)을 이용하여 R 프로그램(version 2.15.0)에서 PCA (Principal component analysis)와 PLS-DA (Partial least square discriminant analysis)분석을 수행하였다. PCA 성분 추출은 아이겐벨류(eigenvalues)가 1 이상을 기준으로 총 10개를

추출하였으며 추출된 각 PCA 성분들은 상관분석을 수행하기 위한 데이터로 사용하였다. 또한 PCA 성분 추출에 중요한 역할을 하는 스펙트럼 부위를 알아보기 위해 PCA loadings을 조사하였다.

#### 지방산 추출 및 기체크로마토그래피 분석

콩으로부터 지방산 분석은 MIDI 방법에 따라 수행하였다(Sasser 1990). 총 지방산 분석을 위해 콩 건조 시료 20 mg에 2 ml의 saponification reagent (7.5 M NaOH:CH<sub>3</sub>OH, 1:1 v/v)를 넣어준 다음  $100^{\circ}\text{C}$  수조에서 30분간 반응을 시켜 지방산을 추출하였다. 추출된 지방산은 4 ml의 methylation reagent (CH<sub>3</sub>OH: 6 N HCl, 1:1 v/v)를 첨가하여  $80^{\circ}\text{C}$ 에서 10분간 에스테르화 반응을 유도하여 fatty acid methyl esters (FAMES)을 제조하였다. 에스테르화 반응(transesterification reaction) 종료 후 2.5 ml의 지방산 추출 용액(hexane:methyl tert-butyl ether, 1:1 v/v)을 첨가한 다음 10분간 반응시켰다. 4,000 rpm에서 10분간 원심분리하여 상층액을 회수한 다음 0.5 M NaOH 용액 6 ml로 세척한 다음 FAMES를 가스 크로마토그래피(YL-6100GC, Young Lin Science, Korea)로 분석하였다. 지방산 분석은 flame ionized detector (FID)를 사용하였으며 사용된 컬럼은 INNOWAX capillary column (Agilent, USA, 30 m  $\times$  0.32 mm  $\times$  0.5  $\mu\text{m}$ )을 사용하였다. 분석 온도는 초기 컬럼의 온도를  $50^{\circ}\text{C}$ 에서 1분간 유지한 다음 매분당  $15^{\circ}\text{C}$  상승하여  $200^{\circ}\text{C}$ 까지 상승시킨 다음 9분간 유지하였으며 최종적으로 분당  $2^{\circ}\text{C}$  씩 상승하여  $250^{\circ}\text{C}$ 까지 올린 다음 2분간 유지하였다. 각각의 FAMES는 FAMES 표준 용액(Supelco® 37 Component FAME Mix, Sigma, USA)의 retention times과 peak areas를 조사하여 각각 정량 및 정성 분석을 수행하였다. 각 대두 계통별 시료는 3회 반복 측정하였다.

#### 지방산 함량 예측 PLS 회귀 모델 및 정확도 검증

대두 종자시료로부터 지방산 함량 예측 모델을 개발하기 위하여 PLS 회귀 알고리즘을 사용하였다. 회귀모델 개발을 위한 두 개의 변수로 X변수는 적외선 분광스펙트럼 데이터를 사용하였으며 Y변수는 GC분석을 통해 확인된 각 시료별 지방산 함량 데이터를 사용하였다. 전체 데이터 세트는 둘로 나누어 PLS 모델 개발을 위한 트레이닝 세트(39개 계통)와 개발된 모델의 정확도 조사를 위한 검증 세트(10개 계통)로 나누어 사용하였다. 모델 개발에 필요한 최적 요인수를 결정하기 위해 leave-one-out 방법을 사용하였다(Martens and Naes 1993). 그리고 최적 모델은 RMSEP (root mean square error of prediction), 상관계수 (coefficient of determination,  $R^2$ ) 값을 이용하여 선발하였다 (Terhoeven-Urselmans *et al.* 2010; Liu *et al.* 2006). 모델의

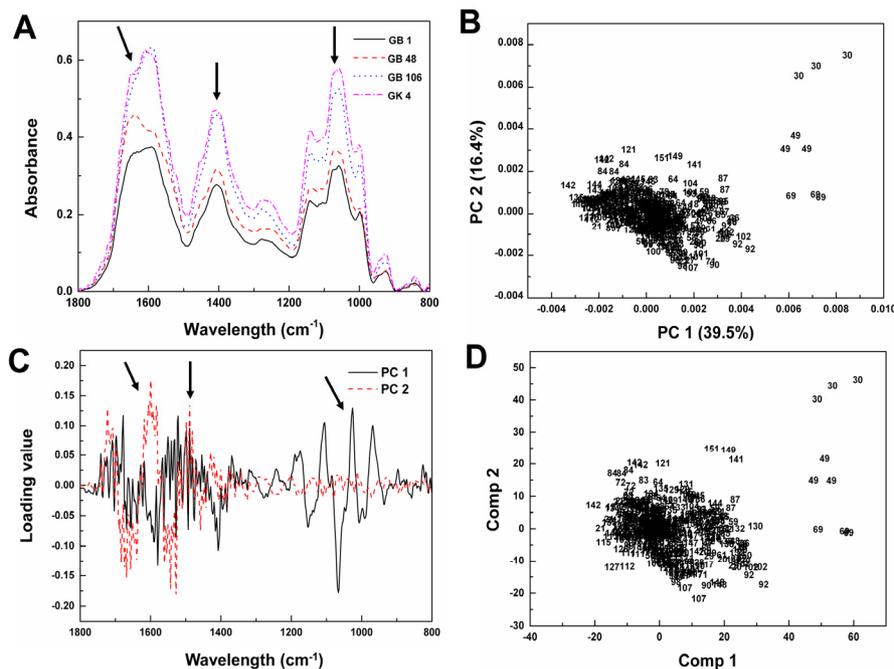
정확성 검증을 위해서는 초기 시료집단에서 사용하지 않은 시료를 이용하여 각 지방산 성분별 함량을 예측하였으며 기체크로마토그래피 분석을 통한 실측값과 회귀분석을 통하여 상관계수 값을 결정하였다. 각 시료별로 예측된 값의 통계적 유의성( $p < 0.05$ )조사는 ANOVA 분석을 통해 검증하였다.

### 결과 및 고찰

#### 대두 적외선분광스펙트럼 데이터 다변량 통계분석

대두 종자 시료로부터 전세포추출물의 적외선분광 스펙트럼 데이터 다변량통계분석을 통해 대두 계통별 수준에서 종간 유연관계 규명 및 식별체계를 확립하였다(Fig. 2). 각 대두 종자시료의 적외선분광 스펙트럼을 비교한 결과 대두 시료들은 FT-IR 스펙트럼상의 1,700~1,550, 1,550~1,300, 1,250~950  $\text{cm}^{-1}$  부위에서 대사체의 양적, 질적 패턴 변화가 크게 이루어짐을 알 수 있었다(Fig. 2A). 이와 같은 대두 종자의 적외선분광스펙트럼 패턴은 대두 건조종자에서 조사된 적외선 분광 패턴과 거의 유사하였다(Miller *et al.* 2005). 적외선 분광스펙트럼은 모든 유기물에 존재하는 화학적, functional group의 진동 패턴을 기

준으로 각 화합물의 화학적 구조를 파악하는 분석장비로서, 각각의 유기물들은 functional group의 존재 유무에 따라 특정 부위에서 특이적인 적외선 흡수 밴드를 가지게 된다. 현재까지 적외선분광스펙트럼의 각 부위별로 나타나는 밴드의 화학적 동정에 관한 많은 연구가 이루어지고 있다. 단백질이나 아미노산 등에 존재하는 amide I (C=O)과 amide II (N-H) group은 적외선 분광 스펙트럼의 1,700~1,550  $\text{cm}^{-1}$  부위에서 특이적인 밴드로 검출되며, aliphatic 및 aromatic 질소화합물의 asymmetric stretching은 주로 1,550  $\text{cm}^{-1}$  부위에서 특이적인 밴드로 검출되며, 핵산이나 인지질 등에 존재하는 phosphodiester group의 P=O asymmetric stretching 이나  $\text{CH}_3$  그룹의 C-H bending은 1,500~1,300  $\text{cm}^{-1}$  부위에서 특이적인 밴드로 검출되며, 그리고 탄수화물의 C-O, C-C, C-O-C, C-O-P 그룹은 주로 1,200~900  $\text{cm}^{-1}$  부위에서 특이적인 밴드로 검출된다(Dumas and Miller 2003; Wolkers *et al.* 2004; Marimuthu and Gurumoorthi 2013). 특히 지방산 성분들을 포함하고 있는 오일의 경우 적외선분광 스펙트럼 조사를 통하여 지방산 성분들의 특이적인 밴드를 규명하는 연구가 오래 전부터 수행되어 왔다(Lanser *et al.* 1991). 각각의 지방산 성분에 포함되어 있는 C=O stretching 1,800~1,750  $\text{cm}^{-1}$  부위에서, C=C stretching은 1,670~1,660  $\text{cm}^{-1}$  부위에서, C-H bending은 1,450~1,300  $\text{cm}^{-1}$  부위에서, 그리고 C-O stretching과 C-H bending은 1,300~1,260  $\text{cm}^{-1}$



**Fig. 2** Multivariate analysis of FT-IR spectral data from 20% methanol extract from seeds of soybean lines. A: Representative FT-IR spectra of soybean lines. Each line represents a soybean line and arrows represent FT-IR spectral range, which shows significant variation between samples. B: Two-dimensional PCA score plot of FT-IR spectral data from 153 soybean lines. The first two principal components are displayed, accounting for 39.5% and 16.4% of the total variation (55.9% of total variation). C: PC loading plot based on PCA data of soybean lines. Arrows represent significant FT-IR spectral ranges for determining PC 1 and PC 2 of the PCA score plot (B). D: PLS-DA score plot based on PCA data of 153 soybean lines

부위에서 특이적인 밴드 패턴을 나타내는 것으로 알려져 있다(Yang *et al.* 2005; Schulz and Baranska 2007). 본 연구에서는 총 153종의 대두 종자 시료가 분석되었으며 이들 시료들은 상기의 부위에서 계통별로 특이적인 적외선 분광스펙트럼을 보이는 것을 확인할 수 있었다. 더 나아가 이들 지방산 특이적인 적외선 흡광 부위에서 흡광도는 지방산 함량이 높은 계통의 경우 증가하지만 지방산 함량이 낮은 계통 시료들의 경우 반대로 흡광도가 낮아짐을 관찰할 수 있었다. 이는 지방산 함량 차이와 적외선 분광스펙트럼의 흡광도 차이가 양의 상관관계가 있음을 나타내는 결과라 사료된다. Sato *et al.* (1998)은 이미 적외선 분광스펙트럼의 다변량통계분석을 이용하여 대두의 계통식별이 가능함을 보고한 바 있다. 이러한 결과들을 고려해보면 지방산의 양적, 또는 질적인 변화가 대두의 계통 식별에서 특이적인 대사체 마커로 활용이 가능함을 나타내는 결과라 사료된다. 또한 본 연구에서 제시한 것처럼 적외선스펙트럼의 질적, 양적 차이는 대두 종자에 함유되어 있는 아미노산이나 단백질, 지방산, 그리고 탄수화물계통의 화합물들의 질적, 양적 차이를 직접적으로 반영하고 있음을 의미한다. 따라서 FT-IR 스펙트럼 분석은 대두 계통별로 주요 대사체의 질적, 양적 변화를 규명하고 이들 화합물의 함량이 높은 계통의 신속한 선발 수단으로 활용이 가능할 것으로 기대된다.

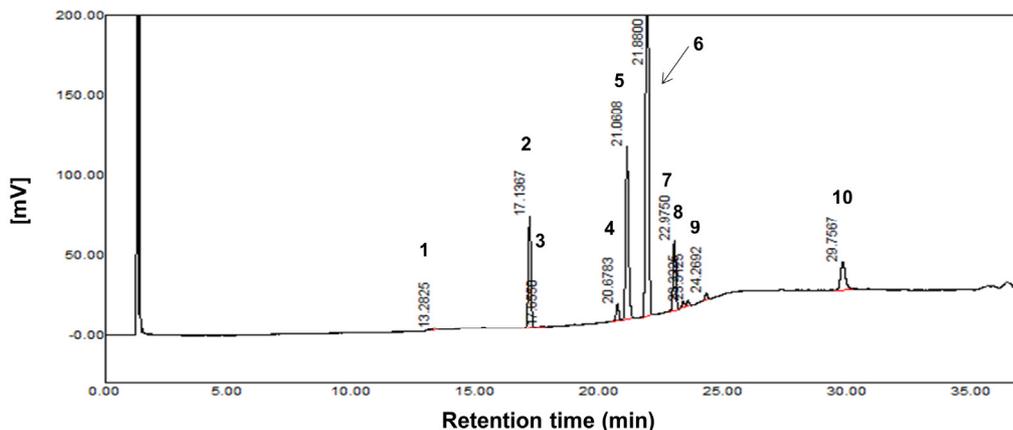
대두 적외선 분광스펙트럼 데이터의 PCA 분석 결과 PC 1과 2 score는 전체 변이량의 각각 35.9%, 16.4%의 설명력을 갖고 있으며 이는 전체 변이량의 약 55.9%를 반영하고 있음을 알 수 있었다(Fig. 2B). 또한 동일 계통에 속하는 반복 조사구들이 다른 계통들 시료와는 별도로 PCA plot 상의 특이적인 위치에서 비교적 인접하여 분포하고 있음을 알 수 있었다. 이 결과는 적외선 분광스펙트

럼 조사가 비교적 높은 재현성을 가짐을 의미하는 것으로 대두의 계통 식별 수단으로 활용이 가능함을 나타내는 결과라 사료된다.

PCA score plot 상에서 대두 시료들의 집단 구분 및 클러스터링에 중요한 역할을 하는 적외선 분광스펙트럼 부위를 조사하기 위하여 PC 1과 2를 결정하는데 중요한 FT-IR 스펙트럼 부위를 조사하였다(Fig. 2C). PC loading value 분석결과 대두 계통식별에 중요한 적외선 분광스펙트럼 부위는 주로 1,700~1,500과 1,500~1,300  $\text{cm}^{-1}$  및 1,100~950  $\text{cm}^{-1}$  부위임을 알 수 있었다(Fig. 2C). 이들 적외선 분광 스펙트럼 부위는 대두 시료의 적외선 분광스펙트럼 비교에서 보였던 것처럼 계통별로 차이가 크게 일어나는 부위(Fig. 2A)와 일치하는 것으로 amide I, II, 지방산과 carbohydrates 계열의 화합물들의 질적, 양적 차이가 대두 종자의 대사체 수준에서 계통별 차이를 결정하는 중요한 화학적 물질임을 의미하는 결과라 사료된다. PLS-DA 분석의 경우 PCA 분석 보다 더욱 확실하게 대두 시료들의 그룹 식별이 가능하였다(Fig. 2D). 각각의 대두 시료들의 그룹 경계는 PCA에 비해 더욱 작아졌으며 동일 종에 속하는 각 시료들의 반복구들도 각 그룹 내에 위치하여 PCA보다 식별능력이 향상 되었음을 알 수 있었다. 따라서 PCA와 마찬가지로 PLS-DA 또한 대사체 수준에서 대두의 계통 식별 수단으로 활용이 가능할 것으로 기대된다.

#### 대두 계통별 지방산 변화

지방산 함량예측 모델링 개발을 위해 39개 대두 계통별 종자로부터 기체크로마토그래피 분석을 통한 지방산의 정량 분석을 수행하였다. 대두 시료의 대표적인 지방산



**Fig. 3** Representative GC chromatograms of fatty acids from soybean seed. Numbers represent identified individual fatty acids for qualitative and quantitative analysis. 1: C14:0 (myristic acid); 2: C16:0 (palmitic acid); 3: C16:1 (palmitoleic acid); 4: C18:0 (stearic acid); 5: C18:1n9c (oleic acid); 6: C18:2n6c (linoleic acid); 7: C18:3n3; 8: U8 (α-linolenic acid); 9: C20:1n9 (cis-11 eicosenoic acid); 10: C23:0 (tricosanoic acid)

분석 크로마토그램은 Fig. 3에 제시하였으며 또한 29개 대두 계통별 시료의 지방산 정량 분석 결과를 도표 1에 제시하였다(Table 1). 총 29종의 대두 계통별 시료로부터 지방산 함량을 조사한 결과 총 지방산의 함량은 건조 시료 0.1 g 당 185.57  $\mu\text{g}$ 에서 325.9  $\mu\text{g}$ 으로 계통간에 차이가 있었음을 알 수 있었으며 평균 함량은 244.48  $\mu\text{g}$ 이었다. 대두 종자에 존재하는 주요 지방산 성분들은 팔미틱산, 스테아릭산, 올레익산, 리노레익산 그리고 리노레닉산임을 알 수 있었다. 본 연구의 대두 지방산 분석 결과는 대두에 존재하는 지방산 성분들의 함량이 리노레익산(55%),

올레익산(20%), 팔미틱산(13%), 리노레닉산(8%) 그리고 스테아릭산(4%)으로 존재한다는 보고(Pham *et al.* 2010)와 거의 일치하였다.

대두에서 지방산의 조성 및 함량은 유전자형에 의해 영향을 받음은 물론 환경요인 특히 종자발달 초기과정에서 온도와 광주조건에 의해 크게 영향을 받는 것으로 알려져 있다(Lee *et al.* 2009). 본 연구에서 대두 계통인 GK15가 지방산 함량이 가장 높은 계통으로 밝혀졌으며, 그 함량은 가장 적은 계통인 GB45 계통에 비해 약 1.75배 높음을 알 수 있었다. 이 결과는 본 연구에 사용된 총 153개 대두

**Table 1** Summary of fatty acid contents from 29 soybean lines using GC

Soybean line	Contents of fatty acids ( $\mu\text{g}/0.1$ g of dry weight)					
	Palmitic acid	Stearic acid	Oleic acid	Linoleic acid	Linolenic acid	Total fatty acids
GB1	32.98 $\pm$ 3.30 <sup>a</sup>	4.76 $\pm$ 0.30	59.50 $\pm$ 3.88	133.18 $\pm$ 16.8	18.41 $\pm$ 2.73	248.84
GB4	29.49 $\pm$ 1.27	4.57 $\pm$ 0.74	73.62 $\pm$ 6.43	128.17 $\pm$ 2.34	17.77 $\pm$ 0.34	253.62
GB7	30.58 $\pm$ 4.95	6.24 $\pm$ 0.19	61.25 $\pm$ 6.28	124.89 $\pm$ 22.46	18.88 $\pm$ 3.45	241.85
GB15	23.21 $\pm$ 3.74	3.67 $\pm$ 0.25	40.30 $\pm$ 4.15	106.39 $\pm$ 14.68	16.55 $\pm$ 1.99	190.12
GB20	31.47 $\pm$ 0.79	5.31 $\pm$ 0.35	75.54 $\pm$ 3.84	124.40 $\pm$ 4.15	19.13 $\pm$ 0.8	255.85
GB27	29.26 $\pm$ 0.67	5.01 $\pm$ 0.52	64.32 $\pm$ 0.8	132.37 $\pm$ 4.07	20.03 $\pm$ 1.32	251.0
GB32	31.08 $\pm$ 4.08	5.25 $\pm$ 0.03	51.2 $\pm$ 12.94	117.88 $\pm$ 19.67	15.76 $\pm$ 2.34	221.16
GB38	27.40 $\pm$ 3.95	4.03 $\pm$ 0.64	39.18 $\pm$ 6.35	120.38 $\pm$ 16.6	22.95 $\pm$ 3.27	213.94
GB39	25.19 $\pm$ 1.93	3.29 $\pm$ 0.61	67.47 $\pm$ 10.81	93.97 $\pm$ 2.36	15.43 $\pm$ 0.82	205.35
GB44	28.40 $\pm$ 8.64	4.33 $\pm$ 1.43	34.00 $\pm$ 9.78	123.42 $\pm$ 34.21	22.19 $\pm$ 5.77	212.34
GB45	22.60 $\pm$ 0.88	3.26 $\pm$ 0.06	35.11 $\pm$ 0.23	107.15 $\pm$ 6.45	17.43 $\pm$ 1.99	185.57
GB48	25.71 $\pm$ 6.54	4.79 $\pm$ 1.24	46.49 $\pm$ 12.14	119.00 $\pm$ 27.34	15.81 $\pm$ 3.13	211.8
GB49	30.44 $\pm$ 2.63	5.46 $\pm$ 1.34	55.11 $\pm$ 8.87	130.47 $\pm$ 7.45	17.18 $\pm$ 0.05	238.65
GB55	24.86 $\pm$ 3.28	4.08 $\pm$ 0.79	40.70 $\pm$ 4.78	111.76 $\pm$ 14.46	14.40 $\pm$ 1.35	195.81
GB70	24.32 $\pm$ 6.76	4.32 $\pm$ 1.84	43.13 $\pm$ 3.39	101.13 $\pm$ 28.14	16.26 $\pm$ 3.53	189.16
GB75	33.41 $\pm$ 0.55	5.78 $\pm$ 0.69	57.67 $\pm$ 12.31	140.09 $\pm$ 6.64	19.98 $\pm$ 4.12	256.92
GB80	31.09 $\pm$ 1.43	5.28 $\pm$ 0.26	66.91 $\pm$ 6.42	132.04 $\pm$ 4.17	18.39 $\pm$ 0.56	253.7
GB81	37.78 $\pm$ 1.43	5.78 $\pm$ 0.90	81.52 $\pm$ 8.14	155.26 $\pm$ 1.3	22.82 $\pm$ 1.24	303.15
GB86	32.23 $\pm$ 9.8	5.17 $\pm$ 2.10	54.07 $\pm$ 15.9	150.67 $\pm$ 35.84	23.09 $\pm$ 5.07	265.23
GB88	33.18 $\pm$ 6.17	5.08 $\pm$ 1.45	67.46 $\pm$ 22.55	141.94 $\pm$ 11.95	19.55 $\pm$ 2.98	267.2
GB91	39.32 $\pm$ 5.33	5.74 $\pm$ 1.52	71.87 $\pm$ 2.17	155.26 $\pm$ 16.93	23.66 $\pm$ 3.44	295.85
GB94	34.04 $\pm$ 1.25	4.58 $\pm$ 0.72	67.22 $\pm$ 20.81	123.74 $\pm$ 6.13	19.92 $\pm$ 2.82	249.51
GB106	38.95 $\pm$ 1.88	5.82 $\pm$ 1.26	85.6 $\pm$ 10.17	155.74 $\pm$ 2.22	20.96 $\pm$ 0.74	307.07
GB107	35.32 $\pm$ 1.91	5.20 $\pm$ 0.87	55.45 $\pm$ 7.16	150.04 $\pm$ 6.52	23.64 $\pm$ 0.07	269.65
GK9	35.52 $\pm$ 1.6	6.25 $\pm$ 0.49	41.71 $\pm$ 2.42	175.37 $\pm$ 3.48	30.74 $\pm$ 0.27	289.6
GK11	27.41 $\pm$ 6.97	3.80 $\pm$ 1.41	33.53 $\pm$ 9.64	140.64 $\pm$ 29.88	25.65 $\pm$ 3.59	231.04
GK15	41.46 $\pm$ 3.8	6.80 $\pm$ 1.19	63.13 $\pm$ 8.81	193.11 $\pm$ 13.73	21.40 $\pm$ 0.08	325.9
GK27	25.89 $\pm$ 3.44	3.38 $\pm$ 0.82	22.5 $\pm$ 4.25	111.20 $\pm$ 10.85	29.36 $\pm$ 0.72	192.33
GK38	40.37 $\pm$ 4.19	4.57 $\pm$ 0.64	61.56 $\pm$ 7.59	134.40 $\pm$ 11.05	26.92 $\pm$ 1.71	267.82
Average	31.14	4.88	55.76	132.21	20.49	244.48
SD	6.23	1.15	17.06	25.50	4.60	45.05

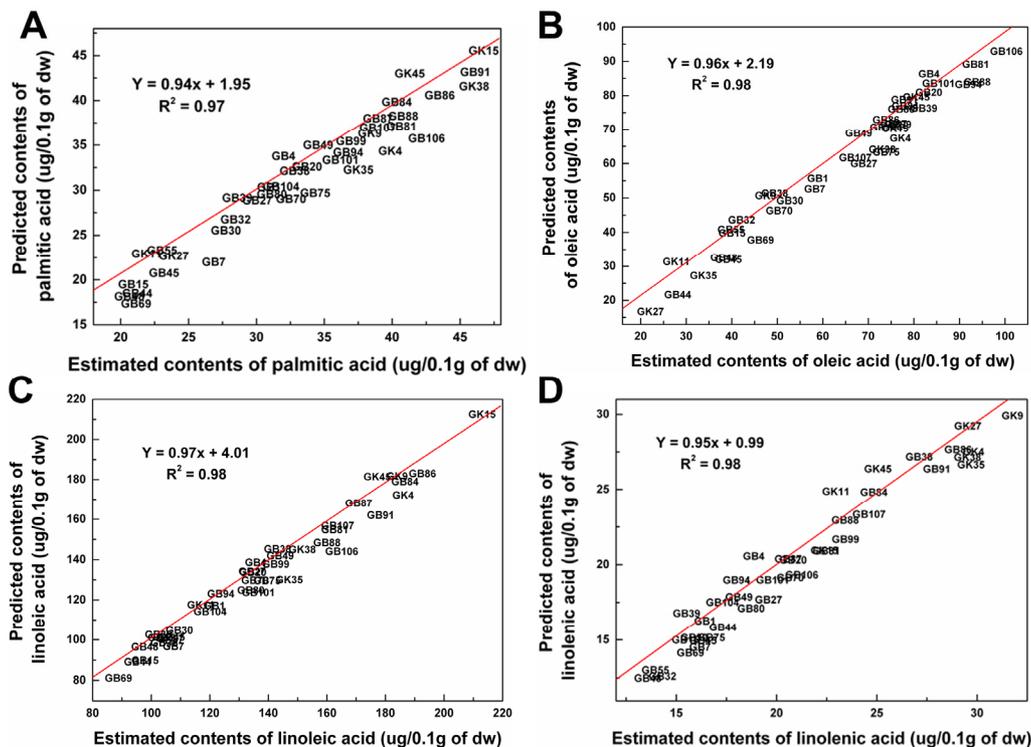
<sup>a</sup>Data are the means  $\pm$  standard deviation

계통별 시료에서 총 지방산의 함량 차이가 2배 이상 벗어나지 않음을 의미하는 결과라 사료된다. 그러나 각 계통별로 개별 지방산 성분들의 함량을 비교한 결과 계통별로 차이가 크게 이루어짐을 알 수 있었다. 지방산 함량이 가장 높은 리놀레익산과 올레익산의 경우 계통별 함량변이는 각각 93.97에서 193.11  $\mu\text{g}$  그리고 22.5에서 85.6  $\mu\text{g}$ 으로 나타났다(Table 1). 이 결과는 대두 계통별로 전체 지방산의 함량차이보다는 개별 지방산의 조성차이가 각각의 계통별로 더욱 크게 이루어짐을 나타내는 결과라 사료된다.

대두 적외선분광스펙트럼으로부터 PLS 회귀모델을 이용한 지방산 함량의 동시 예측 시스템 확립

대두에 존재하는 주요 지방산(팔미틱산, 스테아릭산, 올레익산, 리놀레익산 그리고 리놀레닉산)의 함량 예측 PLS 모델을 확립하였다(Fig. 4). 본 연구에서는 PLS 회귀모델의 정확도 향상을 위하여 서로 다른 지역에서 재배된 대두 계통 시료를 사용하였다. RMSEP (root-mean-square error of prediction) 값이 최소화 되고 그리고  $R^2$  (correlation coefficient of determination) 값이 최대화 될 수 있도록 calibration models 개발에 사용되는 적외선 분광스펙트럼 레인지와 PLS factors의 수를 최적화하여 PLS regression 모델을 개발하였다. 각 지방산 성분별 PLS regression 모델의 교차검증에서 팔

미틱산(RMSEP = 1.885,  $R^2 = 0.97$ ) (Fig. 4A), 스테아릭산(RMSEP = 0.383,  $R^2 = 0.93$ ), 올레익산(RMSEP = 3.917,  $R^2 = 0.98$ ) (Fig. 4B), 리놀레익산(RMSEP = 5.458,  $R^2 = 0.98$ ) (Fig. 4C) 그리고 리놀레닉산(RMSEP = 1.142,  $R^2 = 0.98$ ) (Fig. 4D)이 모두 높은 상관계수 값을 가져 calibration models의 예측 정확도가 높음을 알 수 있었다. 실측 검증을 위해 확립된 최적의 PLS regression 모델을 이용하여 10종의 대두 계통으로부터 각각의 지방산 함량 값을 예측하였으며 기체크로마토그래피 분석에 의해 얻어진 실측값과 상관분석을 실시하였다(Table 2). 또한 예측값의 통계적 유의성 검증결과 스테아릭산( $p > 0.01$ )을 제외하고 모든 지방산 성분들의 예측값이 통계적 유의성을 가지고 있음을 알 수 있었다. 실측 검증을 통한 상관계수를 조사한 결과, 팔미틱산( $R^2 = 0.8002$ ) (Fig. 5A), 올레익산( $R^2 = 0.8909$ ) (Fig. 5B) 그리고 리놀레익산( $R^2 = 0.815$ ) (Fig. 5C)은 회귀 분석 상관계수가 0.8 이상으로 예측 정확도가 비교적 높음을 알 수 있었다. 그러나 스테아릭산( $R^2 = 0.4598$ ) 과 리놀레닉산( $R^2 = 0.6867$ ) (Fig. 5D)의 경우 상관계수가 0.7 이하로 상대적으로 예측 정확도가 낮음을 알 수 있었다. 이 결과를 보면 콩 종자 내에 비교적 함량이 높았던 팔미틱산, 올레익산 그리고 리놀레익산은 모델의 함량 예측 정확도가 높게 나타나지만 반대로 지방산 함량이 낮은 스테아릭산과 리놀레닉산은 모델의 예측 정확도가 상대적으로 낮아짐을 알 수 있었다.

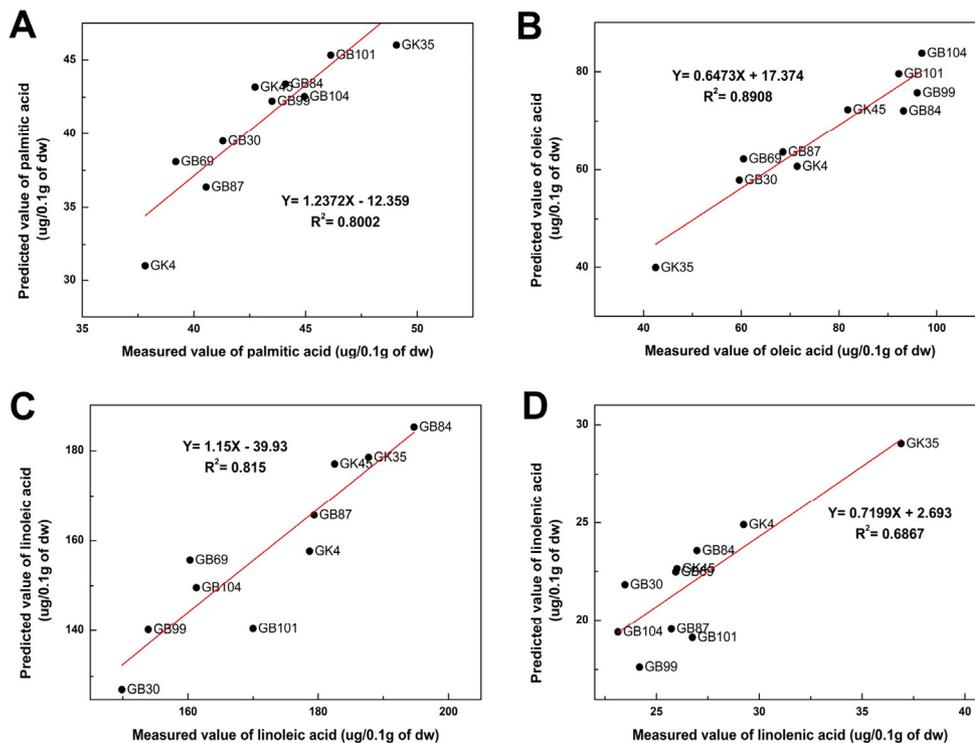


**Fig. 4** Scatter plots showing relationship between predicted values of fatty acids contents by PLS regression modeling and measured values by GC. A: palmitic acid; B: oleic acid; C: linoleic acid; D: linolenic acid

본 연구는 대두의 종자로부터 적외선 분광분석 스펙트럼으로부터 PLS calibrations 모델을 이용하여 지방산의 함량을 높은 정확도로 예측할 수 있음을 입증한 최초의 성공적인 보고이다. 그러므로 다른 선행연구결과들과 정확도의 직접적인 비교분석이 불가능하지만 최근 다양한 PLS calibrations 모델들에 관한 보고들과 간접비교는 가능하다. Scibisz *et al.* (2011)은 적외선 분광분석 스펙트럼을 이용하여 토마토의 당성분, 시트릭산, 그리고 solid matter의 함량예측 PLS calibrations 모델을 개발하여 상관계수가 0.9 이상으로 정확하게 함량예측이 가능함을 보고한 바 있다. 또한 Kovalenko *et al.* (2006<sup>b</sup>)은 근적외선 분광분석 스펙트럼을 이용하여 대두 식물체에서 지방산 성분 함량예측 모델링을 보고한 바 있으며 이 경우 상관계수가 지방산 성분별로 0.91~0.94 사이로 정확하게 예측이 가능함을 보고한 바 있다. 아울러 상관계수가 0.7 이상이 되면 예측능력이 매우 우수한 모델로 시료의 지방산 함량 스크리닝 수단으로 사용이 가능함을 제시한 바 있다. Roberts *et al.* (2006)도 근적외선 분광분석 스펙트럼을 이용하여 대두의 자엽으로부터 지방산 함량 예측 모델링을 개발하여 상관계수가 0.8 이상 가능함을 보고한 바 있다. Patil *et al.* (2010)은 최근 근적외선 분광분석 스펙트럼을 이용하여 대두 종자로부터 지방산 함량 예측 모델링을 개발하였으며 이들은 올레익산, 리노레익산 그리고 팔미틱산은 예측 정확도가 높지만 스테아릭산과 리

노레익산의 경우 예측 정확도가 상대적으로 낮음을 보고한 바 있다(Patil *et al.* 2010). 이들의 보고는 지방산 함량의 절대량이 낮을수록 예측 정확도가 낮아진다는 본 연구의 결과와 매우 일치하는 결과라 사료된다. 예측모델을 확립함에 있어 데이터의 변이 폭이 예측모델의 정확성에 크게 영향을 준다(Velasco *et al.* 1999). 따라서 본 연구에서 스테아릭산과 리노레익산의 예측 정확도가 상대적으로 낮은 것은 이들의 절대적인 함량이 낮음에 기인하고 있을 것으로 추정된다. 또한 각각의 지방산 성분들의 경우 공통적인 functional group들을 가지고 있어서 적외선분광분석 스펙트럼의 특이적인 밴드 패턴 또한 유사할 것이다. 즉 동일구조를 가지는 화합물들의 유사한 적외선 분광 스펙트럼 패턴 또한 상대적으로 소량으로 존재하는 물질들의 함량 예측 정확도를 감소시키는 요인으로 작용하였을 것으로 추측된다. 따라서 스테아릭산과 리노레익산의 예측 정확도를 높이기 위해서는 초기 집단의 변이폭을 증가시킬 수 있도록 초기 시료 수가 증가되거나 또는 PLS regression 외에 다른 알고리즘의 도입을 통한 새로운 모델링이 가능할 것으로 기대된다(Wu *et al.* 2006; Patil *et al.* 2010).

본 연구에서 적외선 분광분석법을 이용하여 확립된 올레익산, 리노레익산 그리고 팔미틱산의 상관계수는 근적외선분광분석법을 통한 예측모델들(Roberts *et al.* 2006; Kovalenko *et al.* 2006<sup>b</sup>; Patil *et al.* 2010)과 거의 비슷한 수



**Fig. 5** Linear regression plots for external validation of fatty acid predictions. Plots are shown for A: palmitic acid; B: oleic acid; C: linoleic acid; D: linolenic acid. Predicted values from FT-IR spectral data by PLS regression and measured values for fatty acids from averages of GC analysis were used in linear regression analysis

**Table 2** Prediction of 5 individual fatty acid contents from soybean lines using PLS regression

Soybean line	Predicted values of fatty acid contents ( $\mu\text{g}/0.1 \text{ g}$ of dry weight)				
	Palmitic acid	Stearic acid	Oleic acid	Linoleic acid	Linolenic acid
GB30	39.49	5.09	57.87	127.17	21.82
GB69	38.09	6.38	62.16	155.69	22.48
GB84	43.39	7.95	72.04	185.17	23.56
GB87	36.37	7.84	63.56	165.65	19.53
GB99	42.23	5.73	75.72	140.16	17.62
GB101	45.33	6.05	79.59	140.37	19.10
GB104	42.54	6.11	83.74	149.58	19.38
GK4	30.99	5.69	60.65	157.66	24.90
GK35	46.02	7.53	39.94	178.51	29.05
GK45	43.18	6.53	72.26	177.07	22.64
Average	40.8	6.5	66.8	157.7	22.0
SD	4.60	0.97	12.73	18.96	3.36
Ratio <sup>a</sup>	1.05	1.03	1.14	1.09	1.22
R <sup>2</sup> <sup>b</sup>	0.8002	0.4598	0.8909	0.8150	0.6868
P value <sup>c</sup>	0.0003	0.0186	0.0000	0.0002	0.0019

<sup>a</sup>Ratio represent the value of measured / predicted contents

<sup>b</sup>R<sup>2</sup>: correlation coefficient of linear regression

<sup>c</sup>P value was evaluated by ANOVA

준으로 정확하게 함량예측이 가능할 것으로 예상된다. 새로운 지방산 조성 변환 신품종 대두 종자를 신속하게 육성하기 위한 핵심 기술요소중의 하나는 바로 간편하고, 빠르고 정확하게 지방산의 함량 조사할 수 있는 방법의 도입이다. 본 연구에서 확립된 것처럼 적외선분광분석법은 간편하고, 재현성이 우수한 방법으로 콩 종자에서 지방산 함량을 동시에 예측 가능하며 더 나아가 각 대사체 수준에서의 차이로 대두 계통 식별이 가능함을 알 수 있었다. 따라서 본 기술의 도입은 기체크로마토그래피에 의한 복잡한 지방산 성분 정성 및 정량 분석을 대체할 수 있는 신속수단으로 활용이 가능할 것으로 기대된다.

결론적으로 본 연구에서 확립된 지방산 성분 동시예측 모델링 기술은 대두는 물론 지방산의 조성변화가 요구되는 다양한 유지작물에서도 지방산 성분 예측 모델로 사용이 가능할 것으로 예측되며 이는 궁극적으로 대두의 계통육성에 요구되는 특성검증 시간을 획기적으로 단축시켜 신품종 육종의 효율을 크게 증대시킬 수 있는 수단으로 활용이 가능할 것으로 기대된다. 아울러 본 기술은 식물 종자는 물론 농산품이나 식품, 의약품의 지방산 관련 특성 검증 수단으로 활용이 가능할 것으로 기대된다.

## 사 사

본 연구는 농업진흥청 차세대바이오그린21사업(PJ008329)

과 한국한의학연구원의 ‘생명공학기술 기반 한약자원 생산 다양화 기술 개발’(K14418) 과제의 연구지원에 의하여 이루어진 결과물입니다.

## 적 요

본 연구의 목적은 적외선 분광스펙트럼 데이터를 이용하여 대두 종자내의 지방산 함량을 동시에 예측할 수 있는지 여부를 조사하기 위한 것이다. 총 153종의 대두(*Glycine max* Merrill) 종자로부터 적외선 분광스펙트럼 및 지방산의 함량을 기체크로마토그래피 분석을 통하여 확인하였다. 적외선 분광스펙트럼 조사결과 대두는 단백질이나 아미노산의 amide bond region ( $1,700 \sim 1,500 \text{ cm}^{-1}$ ), 핵산이나 인지질의 phosphodiester groups ( $1,500 \sim 1,300 \text{ cm}^{-1}$ ) 그리고 탄수화물 등 다당류의 sugar region ( $1,200 \sim 1,000 \text{ cm}^{-1}$ )에서 계통별로 큰 차이가 이루어짐을 알 수 있었다. 총 29라인의 대두 계통별 시료로부터 지방산 함량을 조사한 결과 총 지방산의 함량은 건조 시료 0.1 g 당  $185.57 \mu\text{g}$ 에서  $325.9 \mu\text{g}$ 으로 계통간에 차이가 있었음을 알 수 있었으며 평균 함량은  $244.48 \mu\text{g}$ 이었다. PLS regression 분석을 이용하여 총 5개 지방산(팔미틱산, 스테아릭산, 올레익산, 리노레익산 그리고 리노레닉산) 함량 예측 calibration models의 실측 검증 결과, 팔미틱산( $R^2 = 0.8002$ ), 올레익산( $R^2 = 0.8909$ ) 그리고 리노레익산( $R^2 = 0.815$ )은 회귀분

석 상관계수가 0.8 이상으로 정확도 높음을 알 수 있었다. 그러나 스테아릭산( $R^2 = 0.4598$ )과 리노레닉산( $R^2 = 0.6868$ )의 경우 상관계수가 0.7 이하로 상대적으로 예측 정확도가 낮음을 알 수 있었다. 본 연구에서 확립된 기술은 지방산의 조성 변환을 통하여 새로운 대두 품종 개발을 위한 계통선발 과정에서 매우 효율적인 수단으로 활용이 가능할 것으로 사료된다. 더 나아가 본 기술은 대두는 물론 대두 유래 농산물이나 식품의 품질 검증 수단으로 활용이 가능할 것으로 기대된다.

## References

- Alander JT, Bochko V, Martinkauppi B, Saranwong S, Mantere T (2013) A Review of Optical Nondestructive Visual and Near-Infrared Methods for Food Quality and Safety. *Int J Spectrosc* 2013 : 341-402
- Ascherio A, Willett WC (1997) Health effects of trans fatty acids. *Am J Clin Nutr* 66(Suppl 4) : 1006-1010
- Baranska M, Schütze W, Schulz H (2007) Determination of lycopene and beta-carotene content in tomato fruits and related products: Comparison of FT-Raman, ATR-IR, and NIR spectroscopy. *Anal Chem* 78 : 8456-8461
- Bellincontro A, Taticchi A, Servili M, Esposito S, Farinelli D, Mencarelli F (2012) Feasible application of a portable NIR-AOTF tool for on-field prediction of phenolic compounds during the ripening of olives for oil production. *J Agric Food Chem* 60 : 2665-2673
- Chen Q, Zhao J, Liu M, Cai J, Liu J (2008) Determination of total polyphenols content in green tea using FT-NIR spectroscopy and different PLS algorithms. *J Pharm Biomed Anal* 46 : 568-573
- Dumas P, Miller LM (2003) The use of synchrotron infrared microspectroscopy in biological and biomedical investigations. *Vib Spectrosc* 32 : 3-21.
- Fehr WR, Welke GA, Hammond EG, Duvick DN, Cianzo SR (1991) Inheritance of elevated palmitic acid content in soybean seed oil. *Crop Sci* 31 : 1522-1524
- Fernandez K, Agosin E (2007) Quantitative analysis of red wine tannins using Fourier-transform mid-infrared spectrometry. *J Agric Food Chem* 55 : 7274-7300
- Fontaine J, Hörr J, Schirmer B (2001) Near-infrared reflectance spectroscopy enables the fast and accurate prediction of the essential amino acid contents in soy, rapeseed meal, sunflower meal, peas, fishmeal, meat meal products, and poultry meal. *J Agric Food Chem* 49 : 57-66
- Kovalenko IV, Rippke GR, Hurburgh CR (2006a) Determination of amino acid composition of soybeans (*Glycine max*) by near-infrared spectroscopy. *J Agric Food Chem* 54 : 3485-3491
- Kovalenko IV, Rippke GR, Hurburgh CR (2006b) Measurement of soybean fatty acids by near-infrared spectroscopy; linear and nonlinear calibration methods. *J Am Oil Chem Soc* 83 : 421-427
- Lanser AC, List GR, Holloway RK, Mounts TL (1991) FTIR estimation of free fatty acid content in crude oils extracted from damaged soybeans. *J Am Oil Chem Soc* 68 : 448-449
- Lee JD, Woolard M, Slepser DA, Smith JR, Pantalone VR, Nyinyi CN, Cardinal AC, Shannon JG (2009) Environmental effects on oleic acid in soybean seed oil of plant introductions with elevated oleic concentration. *Crop Sci* 49 : 1762-1768
- Liu Y, Ying Y, Yu H, Fu X (2006) Comparison of the HPLC method and FT-NIR analysis for quantification of glucose, fructose, and sucrose in intact apple fruits. *J Agric Food Chem* 54 : 2810-2815
- Marimuthu M, Gurumoorthi P (2013) Phytochemical screening and FT-IR studies on wild and common south indian legumes. *Asian J Pharm Clin Res* 6(supple 2) : 141-144
- Martens H, Naes T (1993) *Multivariate Calibration*; John Wiley and Sons: Chichester, U.K.
- Mateos-Aparicio I, Redondo Cuenca A, Villanueva-Suarez MJ, Zapata-Revilla MA (2008) Soybean, a promising health source. *Nutr Hosp* 23 : 305-312
- Miller SS, Pietrzak LN, Wetzel DL (2005) Preparation of soybean seed samples for FT-IR microspectroscopy. *Biotechnic Histochemistry* 80(3-4) : 117-121
- Patil AG, Oak MD, Taware SP, Tamhankar SA, Rao VS (2010) Nondestructive estimation of fatty acid composition in soybean (*Glycine max* (L.) Merrill) seeds using near-infrared transmittance spectroscopy. *Food Chem* 12 : 1210-1217
- Pham AT, Lee JD, Shannon JG, Bilyeu KD (2010) Mutant alleles of FAD2-1A and FAD2-1B combine to produce soybeans with the high oleic acid seed oil trait. *BMC Plant Biol* 10 : 195
- Quinones-Islas N, Meza-Márquez OG, Osorio-Revilla G, Gallardo-Velazquez T (2013) Detection of adulterants in avocado oil by Mid-FTIR spectroscopy and multivariate analysis. *Food Res Int* 51 : 148-154
- Rahman SM, Takagi Y, Kubota K, Miyamoto K, Kawakita T (1994) High oleic mutant in soybean induced x-ray irradiation. *Biosci Biotechnol Biochem* 58 : 1070-1072
- Roberts CA, Ren C, Beuselinck PR, Benedict HR, Bilyeu K (2006) Fatty acid profiling of soybean cotyledons by near-infrared spectroscopy. *Appl Spectrosc* 60 : 1328-1333
- Sasser M (1990) Identification of Bacteria by gas chromatography of cellular fatty acids. *MIDI Technical Note 101*, MIDI, Newark
- Sato T, Uezono I, Morishita T, Tetsuka T (1998) Nondestructive estimation of fatty acid composition in seeds of *Brassica napus* L. by near-infrared spectroscopy. *J Am Oil Chem Soc* 75 : 1877-1881
- Sato T, Maw A, Katsuta M (2003) NIR reflectance spectroscopic analysis of the FA composition in sesame (*Sesamum indicum* L.) seeds. *J Am Oil Chem Soc* 80 : 1157-1161
- Schulz H, Baranska M (2007) Identification and quantification of valuable plant substances by IR and Raman spectroscopy. *Vib Spectrosc* 43 : 13-25
- Scibisz I, Reich M, Bureau S, Gouble B, Causse M, Bertrand D, Renard CMGC (2011) Mid-infrared spectroscopy as a tool for rapid determination of internal quality parameters in tomato.

- Food Chem 125 : 1390-1397
- Soriano A, Perez-Juan PM, Vicario A, Gonzalez JM, Perez-Coello MS (2007) Determination of anthocyanins in red wine using a newly developed method based on Fourier transform infrared spectroscopy. Food Chem 104 : 1295-1303
- Terhoeven-Urselmans T, Vagen T-G, Spaargaren O, Shepherd KD (2010) Prediction of soil fertility properties from a globally distributed soil mid-infrared spectral library. Soil Sci Soc Am J 74 : 1792-1799
- Tilman BL, Gorbet DW, Person G (2006) Prediction oleic and linoleic acid content of single peanut seed using near-infrared reflectance spectroscopy. Crop Sci 46 : 2121-2126
- Velasco L, Goffman D, Becker HC (1999) Development of calibration equations to predict oil content and fatty acid composition in Brassicaceae germplasm by near-infrared reflectance spectroscopy. J Am Oil Chem Soc 76 : 25-30
- Wold H (1966) Nonlinear Estimation by Iterative least squares procedures. In: David F, (eds), Research Papers in Statistics, Wiley, New York, pp. 411-444
- Wolkers WF, Oliver AE, Tablin F, Crowe JH (2004) A fourier transform infrared spectroscopy study of sugar glasses. Carb Res 339 : 1077-1085
- Wu JG, Shi CH, Zhang HZ (2006) Study on developing calibration model of fat acid composition in intact rapeseed by near infrared reflectance spectroscopy. Spectrosc Spect Anal 26 : 259-262
- Yang H, Irudayaraj J, Paradkar MM (2005) Discriminant analysis of edible oils and fats by FTIR, FT-NIR and FT-Raman spectroscopy. Food Chem 93 : 25-32