



저작자표시-동일조건변경허락 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.
- 이차적 저작물을 작성할 수 있습니다.
- 이 저작물을 영리 목적으로 이용할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



동일조건변경허락. 귀하가 이 저작물을 개작, 변형 또는 가공했을 경우에는, 이 저작물과 동일한 이용허락조건하에서만 배포할 수 있습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

박 사 학 위 논 문

FT-IR 스펙트럼 기반  
다변량통계분석을 이용한 작물의  
품종 및 형질 예측 모델링 확립과  
적용

제주대학교 대학원

생명공학과

송 승 엽

2014년 8월



FT-IR 스펙트럼 기반  
다변량통계분석을 이용한 작물의  
품종 및 형질 예측 모델링 확립과  
적용

지도교수 : 김 인 중

송 승 엽

이 논문을 이학 박사학위 논문으로 제출함

2014년 8월

송승엽의 이학 박사학위 논문을 인준함

심사위원장 \_\_\_\_\_ 印)

위 원 \_\_\_\_\_ 印)

위 원 \_\_\_\_\_ 印)

위 원 \_\_\_\_\_ 印)

위 원 \_\_\_\_\_ 印)

제주대학교 대학원

2014년 8월

FT-IR Spectrum data of Crops using  
Multivariate Analysis and Prediction of  
PLS Regression Modeling

Seung-Yeob Song

(Supervised by Professor In-Jung Kim)

Faculty of Biotechnology  
College of Applied Life Sciences  
Jeju National University

August, 2014

## 목 차

목 차 .....	i
<b>LIST OF FIGURES</b> .....	iii
<b>LIST OF TABLES</b> .....	vii
<b>1. ABSTRACT</b> .....	1
<b>2. INTRODUCTION</b> .....	5
<b>3. FT-IR 스펙트럼 기반 다변량통계분석기법에 의한 두과작물의 대사체 수준 식별체계 확립</b> .....	13
3. 1. Abstract	
3. 2. Introduction	
3. 3. Material and Methods	
3. 4. Results and Discussion	
<b>4. FT-IR 스펙트럼 데이터의 다변량통계분석을 이용한 고기능성 아프리카인 암 식별 및 기능성 성분 함량 예측 모델링</b> .....	29
4. 1. Abstract	
4. 2. Introduction	
4. 3. Material and Methods	
4. 4. Results and Discussion	
<b>5. FT-IR 스펙트럼과 HPLC 정량 분석을 이용한 감귤과실로부터 carotenoid 성분 예측</b> .....	50
5. 1. Abstract	
5. 2. Introduction	
5. 3. Material and Methods	
5. 4. Results and Discussion	
<b>6. FT-IR 스펙트럼 데이터의 다변량통계분석을 이용한 돌연변이 감귤의 당산도 예측 모델링</b> .....	74

6. 1. Abstract	
6. 2. Introduction	
6. 3. Material and Methods	
6. 4. Results and Discussion	
<b>7. Genomic DNA의 FT-IR 스펙트럼을 이용한 적배추 부모로부터 F1 식별</b>	
.....	95
7. 1. Abstract	
7. 2. Introduction	
7. 3. Material and Methods	
7. 4. Results and Discussion	
<b>8. Discussion</b> .....	114
<b>9. REFERENCES</b> .....	126
<b>ACKNOWLEDGMENT</b> .....	146

## LIST OF FIGURES

<b>Fig. 1</b> Significant events leading to the development of current FT-IR spectrometers and microspectrometers. ....	10
<b>Fig. 2</b> The seeds of leguminous plants used in the study.....	18
<b>Fig. 3</b> Representative FT-IR spectra from leguminous plants.....	22
<b>Fig. 4</b> PCA score plot of FT-IR data from leguminous plants.....	24
<b>Fig. 5</b> Loading value plot of PCA analysis from FT-IR data of leguminous plants.....	25
<b>Fig. 6</b> PLS-DA score plot of FT-IR data from leguminous plants.....	27
<b>Fig. 7</b> Hierarchical dendrogram of FT-IR data from leguminous plants.....	28
<b>Fig. 8</b> Representative FT-IR spectral from African yam tubers.....	41
<b>Fig. 9</b> PCA scores (A) and loading values plot (B) of PCA scores from FT-IR data of African yam lines.....	44
<b>Fig. 10</b> PLS-DA score plot of FT-IR data from African yam lines.....	46
<b>Fig. 11</b> Linear regression analysis between estimated and predicted values of total carotenoids (A), flavonoids (B), and phenolic contents (C) by the PLS regression model from FT-IR spectral data.....	49

<b>Fig. 12</b> Multivariate analysis of FT-IR spectroscopic data from peels of of <i>C. unshiu</i> .....	60
<b>Fig. 13</b> Multivariate analysis of FT-IR spectroscopic data from flesh of <i>C. unshiu</i> .....	63
<b>Fig. 14</b> Carotenoid biosynthetic pathway and monthly carotenoid change from peels of wild-type <i>C. unshiu</i> fruits.....	66
<b>Fig. 15</b> Carotenoid biosynthetic pathway and monthly carotenoid change from peels of wild-type <i>C. unshiu</i> fruits.....	67
<b>Fig. 16</b> Monthly variation of carotenoid content from citrus.....	69
<b>Fig. 17</b> Linear regression plots of predicted and estimated carotenoid content. ....	73
<b>Fig. 18</b> Representative FT-IR spectral from citrus fruits.....	84
<b>Fig. 19</b> PCA scores (A) and loading values plot (B) of PCA scores from FT-IR data of citrus fruits lines.....	87
<b>Fig. 20</b> PLS-DA score plot of FT-IR data from citrus fruits lines.....	89
<b>Fig. 21</b> Hierarchical dendrogram of FT-IR data from Citrus fruits.....	90
<b>Fig. 22</b> Linear regression analysis between estimated and predicted values of sugar contents (A) and acidity contents (B) by the PLS regression model from FT-IR spectral data.....	92

<b>Fig. 23</b> Multivariate analysis of FT-IR spectral data from genomic DNA of <i>B. campestris</i> subsp. <i>pekinensis</i> parental inbred lines (P1 and M) and their F1 progeny lines (F1).....	105
<b>Fig. 24</b> Multivariate analysis of FT-IR spectral data from genomic DNA of <i>B. campestris</i> subsp. <i>pekinensis</i> parental inbred lines (P2 and M) and their F1 progeny lines (F2).....	106
<b>Fig. 25</b> Multivariate analysis of FT-IR spectral data on genomic DNA from <i>B. campestris</i> subsp. <i>pekinensis</i> parental inbred lines (P1, P2, and M) and their F1 progeny lines (F1 and F2).....	109
<b>Fig. 26</b> HCA dendrograms from PCA (a) and PLS-DA (b) of <i>B. campestris</i> subsp. <i>pekinensis</i> parental inbred lines (P1, P2, and M) and their F1 progeny lines (F1 and F2).....	111
<b>Fig. 27</b> Representative FT-IR spectra from Citrus fruits.....	116
<b>Fig. 28</b> Linear regression analysis between estimated and predicted values of total carotenoids (A), flavonoids (B), and phenolics contents (C) by the PLS regression model from FT-IR spectral data.....	119
<b>Fig. 29</b> Multivariate analysis of FT-IR spectral data on genomic DNA from <i>C. hybrid</i> 'Kiyomi' cross breeding lines (CM and CO) and their progeny lines (CK).....	121
<b>Fig. 30</b> Multivariate analysis of FT-IR spectral data on genomic DNA from <i>C. hybrid</i> 'Shiranuhi' cross breeding lines (CK and CB) and their progeny lines (CR).....	123

**Fig. 31** Multivariate analysis of FT-IR spectral data on genomic DNA from *C. hybrid* 'Shiranuhi' cross breeding lines (CR and CH) and their progeny lines (CP)..... 124

## LIST OF TABLES

<b>Table 1.</b> Quantitative analysis of total carotenoids, flavonoids and phenolic contents from African yam tubers using UV-VIS spectrophotometric analysis. .....	39
<b>Table 2.</b> Summary of Carotenoid Content Prediction from FT-IR Spectroscopic Data by PLS Regression Modeling.....	72
<b>Table 3.</b> Quantitative analysis of sugar content and acidity content from <i>citrus</i> using HORIBA analysis.....	82
<b>Table 4.</b> Summary of the PLS-DA classification results from FT-IR spectral data of genomic DNA.....	113
<b>Table 5.</b> Quantitative analysis of total carotenoids, flavonoids and phenolics contents from <i>citrus</i> using UV-VIS spectrophotometric analysis.....	118

## 1. ABSTRACT

To determine whether FT-IR spectral analysis based on multivariate analysis for whole cell extracts can be used to discriminate major leguminous plant at metabolic level, seeds of six major leguminous plants were subjected to Fourier transform infrared spectroscopy (FT-IR). FT-IR spectral data from seeds were analyzed by principal component analysis (PCA), partial least square discriminant analysis (PLS-DA) and hierarchical clustering analysis (HCA). The hierarchical dendrogram based on PLS-DA separated the six leguminous plants into four major groups. The first group consisted of *Vigna radiata* var. *radiata*, *Vigna angularis* var. *angularis* and *Vigna unguiculata* subsp. *unguiculata*, whereas *Pisum sativum* var. *sativum*, *Glycine max* L and *Phaseolus vulgaris* var. *vulgaris* were clustered into a separate group respectively. These results showed that metabolic classification system were in accordance with known phylogenetic taxonomy. Thus we suggested that the hierarchical dendrogram based on PLS-DA of FT-IR spectral data from seeds represented the most probable chemotaxonomical relationship between six leguminous plants. Furthermore these metabolic discrimination systems could be applied for rapid selection and classification of useful leguminous cultivars.

We established a high throughput screening system of African yam tuber lines which contain higher contents of total carotenoids, flavonoids, and phenolic compounds using ultraviolet-visible (UV-VIS) spectroscopy and Fourier transform infrared (FT-IR) spectroscopy in combination with multivariate analysis. The total carotenoids contents from 62 African yam tubers varied from 0.01 to 0.91mg·g<sup>-1</sup> dry weight (wt). The total flavonoids and phenolic compounds also varied from 12.9 to 229mg·g<sup>-1</sup> and from 0.29 to 5.2mg·g<sup>-1</sup> dry wt. FT-IR spectra confirmed typical spectral differences between the frequency regions of 1,700–1,500, 1,500–1,300 and 1,100-950

cm<sup>-1</sup>, respectively. These spectral regions were reflecting the quantitative and qualitative variations of amide I, II from amino acids and proteins (1,700–1,500cm<sup>-1</sup>), phosphodiester groups from nucleic acid and phospholipid (1,500–1,300cm<sup>-1</sup>) and carbohydrate compounds (1,100-950cm<sup>-1</sup>). Principal component analysis (PCA) and subsequent partial least square-discriminant analysis (PLS-DA) were able to discriminate 62 African yam tuber lines into three separate clusters corresponding to their taxonomic relationship. The quantitative prediction modeling of total carotenoids, flavonoids, and phenolic compounds from African yam tuber lines were established using partial least square regression algorithm from FT-IR spectra. The regression coefficients ( $R^2$ ) between predicted values and estimated values of total carotenoids, flavonoids and phenolic compounds were 0.83, 0.86, and 0.72, respectively. These results showed that quantitative predictions of total carotenoids, flavonoids, and phenolic compounds were possible from FT-IR spectra of African yam tuber lines with higher accuracy. Therefore we suggested that quantitative prediction system established in this study could be applied as a rapid selection tool for high yielding African yam lines.

This work examined the feasibility of prediction of carotenoid content without HPLC analysis using prediction modeling of carotenoid content by multivariate analysis combined with FT-IR and HPLC data. FT-IR spectra from peels and flesh of citrus (*Citrus unshiu* Marc. cv. Miyagawa) fruit were measured at monthly intervals. Quantitative analysis of carotenoids from the same citrus fruit was confirmed by quantitative HPLC analysis. FT-IR spectroscopy showed that remarkable increase in carbohydrate region (1,000-1,200 cm<sup>-1</sup>) and decrease in amide region (1,500-1,700 cm<sup>-1</sup>). HPLC also showed that increase in  $\beta$ -cryptoxanthin and decrease in lutein content during citrus fruit development. Reliable prediction of antheraxanthin ( $R^2=0.9117$ ),  $\beta$ -carotene ( $R^2=0.8816$ ),  $\beta$ -cryptoxanthin ( $R^2=0.8856$ ), and violaxanthin ( $R^2=0.7305$ ) was obtained from partial least square (PLS) regression modeling.

Considering these results, FT-IR might be applied for metabolic evaluation of citrus fruit during ripening.

Fourier transform infrared spectroscopy (FT-IR) spectroscopy in combination with multivariate analysis was used to discriminate two different F1 hybrid lines from their parental inbred lines. Genomic DNA was isolated from leaves of three parental lines and two F1 hybrid lines of *Brassica campestris* subsp. *pekinensis*. Purified genomic DNA was analyzed by FT-IR spectroscopy in the spectral region from 4,000 to 400  $\text{cm}^{-1}$ . FT-IR spectra confirmed typical spectral differences between the frequency regions of N-H stretching (amide I) and C=O stretching vibrations (amide II) as well as  $\text{PO}_2^-$  ionized asymmetric and symmetric stretching. Principal component analysis was able to discriminate between F1 hybrid progenies depending on their parental lineages, even though they share the same maternal background. Partial least squares discriminant analysis gave a more clear discrimination between the two parental lines and their hybrid progenies. These FT-IR spectral differences might be directly related to subtle changes in the base functional group and backbone structures of genomic DNA. Considering these results, this technique could provide a solid research foundation for FT-IR spectral-based rapid diagnosis, selection, and discrimination of parental lines from their progenies. Furthermore, this technique could be applied to test purity in the hybrid seed industry.

We established a high throughput screening system of Citrus fruits lines which contain higher contents of sugar content and acidity content using HORIBA and Fourier transform infrared (FT-IR) spectroscopy in combination with multivariate analysis. The sugar contents from 5 citrus fruits varied from 0.84 to 10.3°Brix. The acidity contents also varied from 0.61 to 0.99%. FT-IR spectra confirmed typical spectral differences between the frequency regions of 1,700–1,500, 1,500–1,300 and 1,100–950  $\text{cm}^{-1}$ , respectively. These spectral regions were reflecting the quantitative and qualitative variations of amide I,

II from amino acids and proteins ( $1,700\text{--}1,500\text{cm}^{-1}$ ), phosphodiester groups from nucleic acid and phospholipid ( $1,500\text{--}1,300\text{cm}^{-1}$ ) and carbohydrate compounds ( $1,100\text{--}950\text{cm}^{-1}$ ). Principal component analysis (PCA) and subsequent partial least square-discriminant analysis (PLS-DA) were able to discriminate 5 citrus fruits lines into three separate clusters corresponding to their taxonomic relationship. The quantitative prediction modeling of sugar contents and acidity contents from citrus fruits lines were established using partial least square regression algorithm from FT-IR spectra. The regression coefficients( $R^2$ ) between predicted values and estimated values of sugar contents and acidity contents were 0.99 respectively. These results showed that quantitative predictions of sugar contents and acidity contents were possible from FT-IR spectra of citrus fruits lines with higher accuracy. Therefore we suggested that quantitative prediction system established in this study could be applied as a rapid selection tool for high yielding citrus fruits lines.

## 2. INTRODUCTION

### 2. 1. 식물의 종류

두과작물은 전세계적으로 48%의 기름 생산량을 차지하고 있는 경제적으로 중요한 작물이다(Singh and Hymowiz 1988). 또한 두과작물은 단백질 40%, 탄수화물 30% 그리고 지방 20%를 함유하는 고기능성 식량작물이다(Jang et al. 2010). 두과작물은 우리나라 전역에서 재배되고 있으며, 오랜 세월 지역적 특성에 따라 다양한 품종들이 농가에서 재배되고 있다(Kim et al.1993; Lee and Kim 1993). 최근 국산 콩을 이용하는 소비자가 증가하면서 국산 콩에 대한 관심이 급증하고 있지만 우리나라의 콩 재배면적 감소에 따른 생산량 감소로 식용콩 자급률은 32.5%에 불과하여 외국산 수입 콩의 비중이 크게 증가하고 있는 실정이다(Im and Lee 2011). 이와 더불어 유전자 변형 농산물에 대한 소비자의 식품 안정성에 대한 우려도 크게 증가하고 있어서 국산 콩과 수입 콩의 원산지 식별은 물론 다양한 기능성 두과작물의 신제품 개발을 위해서도 신속한 품종 및 계통 식별방법의 개발이 요구되고 있다.

아프리칸 Yam은 주로 아프리카 지역에서 재배되고 있는 다년생작물로, 카사바, 감자, 고구마와 같이 전분 함량이 높아 중요한 식량자원으로 이용되고 있다. Yam의 구근은 전분질 15-20%, 단백질 1-1.5%, 지질 1% 및 기타 미네랄 및 비타민 등을 함유하고 있다(Adeleke, 2010; Kwon et al., 2010). 또한 Yam의 구근에는 다양한 약리적 활성을 갖는 유용 기능성물질, saponin, tannin, sitosterol, diosgenin, carotenoid, flavonoid, polyphenol, allantoin, araginine, mucin 등을 함유하고 있어 약용작물로서 활용 가치가 매우 높은 식물이다(Kim et al., 2009; Kum et al., 2006; Kwon et al., 2010). 특히 Yam의 구근에 존재하는 다양한 페놀성 화합물들은 자유 라디칼 제거, 항산화 효과 및 항염증 작용 등 다양한 약리적 기능을 가지고 있다(Yang et al., 2009; Yao et al., 2004). 따라서 Yam은 자양, 강장, 폐결핵, 고혈압, 당뇨병, 스트레스와 피로 회복에 효능이 알려져 있다(Song et al., 2010; Yoon et al., 1989). 이처럼 아프리칸 Yam은 식용 및

약용작물로서 가치가 높은 작물이지만 주요 성분분석에 대한 정확한 분석이 활발하게 이루어지지 않고 있다. Im et al. (1995)은 아프리카 양의 경우 사포닌, 알칼로이드, 페놀성 성분이 국내산 양에 비해 높아서 보다 쓴맛이 강하다는 보고가 있다. 따라서 식량 및 약리적 측면에서 우수한 아프리카 양의 농업적 활용 기반 구축을 위해 다양한 아프리카산 양의 기능성 성분 분석이 이루어져야 할 것으로 사료된다. 아울러 기능성 성분분석 데이터를 기반으로 고기능성 양 라인의 신속한 선발시스템 확립이 이루어져야 할 것으로 사료된다.

감귤은 세계에서 가장 경제적으로 중요한 과일 작물이고, 과일 중에서 복잡하고 많은 수의 carotenoid가 발견되었다(Gross 1987). 감귤은 일반적으로 유전적, 성숙 단계, 환경 조건 등 여러 가지 요인에 따라 carotenoid가 다르게 나타난다(Lee and Castle 2001; Kato et al. 2004; Fanciullino et al. 2006). 일반적인 carotenoid 성분 외에 감귤은 인간의 식단에 필요한 많은 유기 화합물을 포함하고, 건강에 가치 있는 대사물질을 함유하고 있다(Tadeo et al. 2008). Carotenoid는 식물의 chromoplasts와 다양한 광합성 생물 조류, 곰팡이, 박테리아 등에서 tetraterpenoid 유기 색소가 자연스럽게 엽록체에서 합성되어 생성된다(Moran et al. 2010). 600 개 이상의 알려진 carotenoid는 lutein과 zeaxanthin 같은 산소가 포함된 xanthophylls과  $\alpha$ -carotene,  $\beta$ -carotene, lycopene 같은 산소가 포함되지 않은 carotenoid로 분류 된다(Armstrong et al. 1996). 일반적으로 carotenoid는 식물에서 광합성의 빛에너지 흡수와 광손상으로부터 chlorophyll 보호 역할을 한다. carotenoid 색소는 과실의 특성을 결정하는데 중요하게 작용한다.1 게다가, carotenoid는 인간과 동물에 필수적인 비타민 A (mainly  $\beta$ -carotene,  $\alpha$ -carotene,  $\beta$ -cryptoxanthin)의 전구체와 항산화활성 억제와 같이 영양적인 가치가 있다 (Olson 1989).

적배추(*Brassica campestris* subsp. *pekinensis*)는 한국, 일본 그리고 중국에서 대중적으로 소비되는 채소 작물이다. 십자화과 채소는 중요한 기능성 성분인 glucosinolates, carotenoids, flavonoids 그리고 phenolic compounds를 함유하고 있다(Kim et al. 2004; Harbaum et al. 2007). 십자화과 채소에 함유되어 있는 여러 가지 성분들은 암에 위험을 감소시킨다(Giovanucci et al. 2003). 이러한 채소들의 소비는 암 또는 각종 만성 질환을 보호하는 역할을 한다(van Poppel et al.

1999; Mas et al. 2007). 인간 건강에 대한 관심의 증가로, 많은 새로운 품종과 기능성 채소의 재배가 기존 육종방법을 통해 개발되었다. 더욱이, 국제 UPOV에 따라 지적재산권 보호가 강해지면서, 새로운 품종과 재배의 특성을 구별하는 기술 개발이 필요하다.

감귤은 제주도에서 2008년 이후 6,000억원이상의 생산액을 꾸준히 나타냈고, 국내 과수 총생산량의 1/4 이상을 차지할 만큼 생산량이 높은 과실이다. 최근 중국, 일본, 미국, 유럽과의 FTA 등 수입개방화로 인해 감귤류의 수입량이 증가하면서 국내 시장에 많은 비중을 차지하고 있다. 이에 따라 국내 감귤산업의 생존이 위협받고 있는 실정이다. 또한, 2012년부터 시행된 UPOV 협약에 따라 제주도에서 재배되는 99%의 일본품종 감귤이 품종 보호되면서 로열티에 대한 문제가 발생하고 있다. 그 중 일부 품종인 궁천, 흥진, 일남 1호 등이 품종보호기간 25년이 만료되었지만, 한라봉, 천혜향, 감평, 진지향 등은 품종보호가 되고 있다(Jun and Yoon, 2013).

품종보호뿐만이 아니라 2003년 이후 급격히 증가하고 있는 오렌지 수입량이 12만 톤 이상으로 증가하고 있고, 한·미 FTA 발효에 따라 가격이 싸지면서 국내 시장을 위협하고 있다. 이것은 국내 감귤농가에 큰 피해로 다가오고 있고, 비상품 감귤의 가공적인 측면에서도 많은 피해가 발생하고 있다.

이러한 문제점들을 극복하기 위한 한 가지 방법으로 돌연변이 육종이 대두되고 있다. 이 방법은 인위적으로 유전적 변이를 유도하여 개체의 형질을 변화시키는 방법이다. 전세계적으로 80년 이상 사용되었고, 1960년대부터 국내에서도 벼, 콩, 마늘 등의 다양한 품종에 활용되고 있다. 하지만 과수에서는 주로 교배 육종이 사용되고 있어 돌연변이 육종은 거의 이루어지고 있지 않다. 또한, 감귤에서는 당산도 우수비가 다른 작물에 비해 매우 중요하게 작용하고 있고, 아직까지 당산도에 대한 예측 모델이 보고된바 없다.

## 2. 2. 분석 방법

신품종 육성에 따른 품종 구분이 대체적으로 RAPD(random amplification of polymorphic DNA), RFLP(restriction fragment length polymorphism), SSR(simple

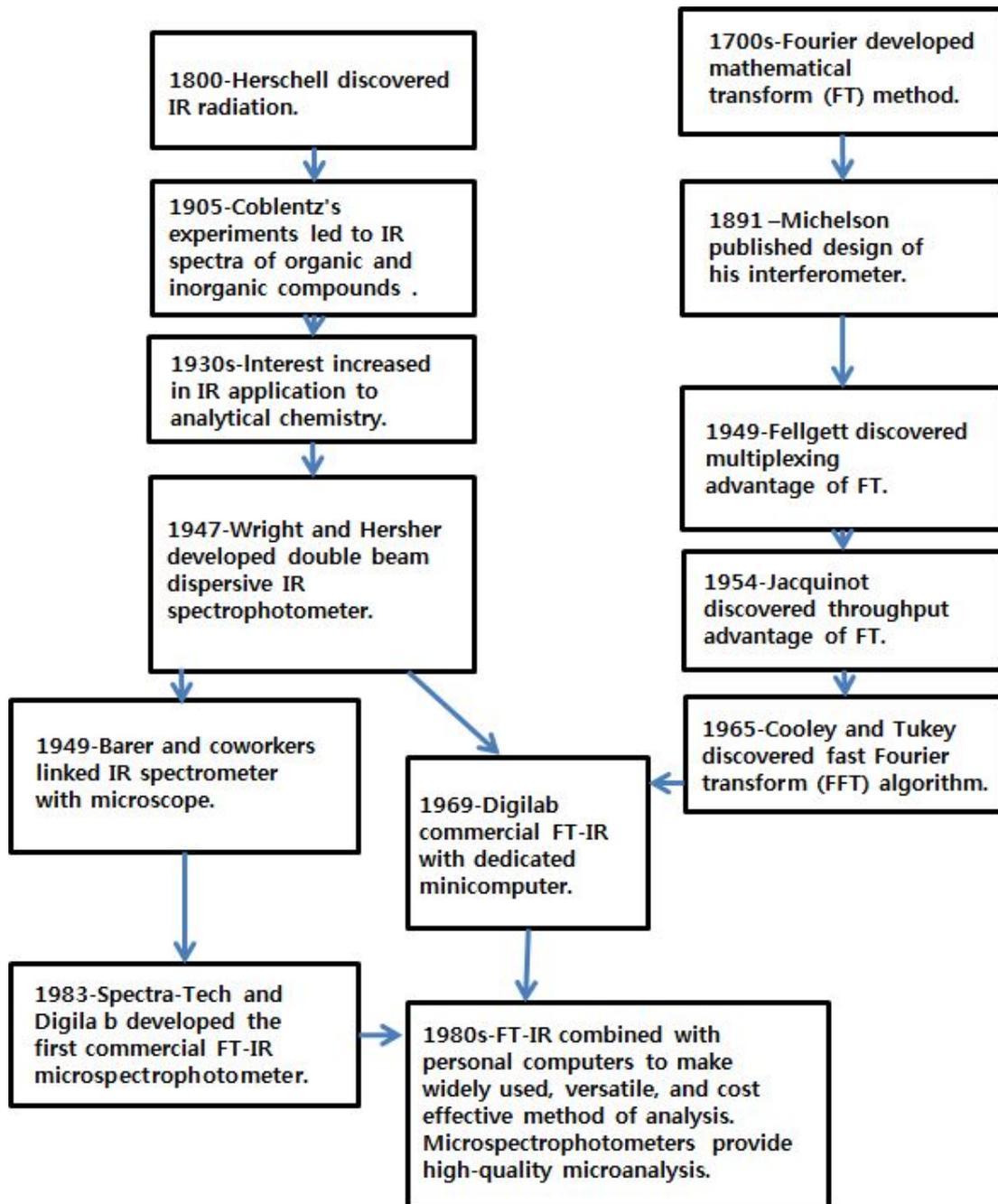
sequence repeat), AFLP(amplified fragment length polymorphism)와 같은 방법으로 분자마커를 이용한 구분 방법들이 개발되어 왔다(Sun et al., 2012; Liu et al., 2013; Nageswara-Rao et al., 2013; Chen et al., 2013). 또한 유전적 다양성 분석과 품종 구분을 위한 분자마커로 SRAP (sequence-related amplified polymorphism) 마커들이 활용되고 있다(Polat et al., 2012). 이러한 방법들은 많은 시간과 많은 분석 비용이 소모가 되는 단점을 가지고 있다.

유용 기능성 물질들의 정량 및 정성분석은 주로 gas chromatography(GC)나 high performance liquid chromatography(HPLC) 및 질량분석기와 연계를 통해 정확한 성분의 정량 및 정성 분석이 이루어진다(Aktumsek et al., 2013; Kimura et al., 2002). 그러나 이러한 분획 컬럼을 통한 기능성 성분 정량 분석의 경우 분석 결과는 매우 정확하지만, 상대적으로 시료의 준비과정이 복잡하며, 추출 및 시료 분석 비용이 많이 소요되고, 다수의 시료를 분석할 경우 많은 시간이 요구되는 한계점을 가지고 있다(Yuan et al., 2013). 이러한 정량 및 정성분석의 복잡성 문제를 해결하기 위한 하나의 대안으로 ultraviolet-visible (UV-VIS) spectrophotometer를 이용한 기능성 성분들의 신속한 함량 조사가 가능함이 보고되고 있다(Kofalvi and Nassuth, 1995; Stadnik and Buchenauer, 2000; Zhishen et al., 1999).

농작물의 생산성을 높이기 위해, 잡종 강세는 F1 종자를 생산하는데 이용되고 있다. 그러나, 잡종 강세는 신품종 육성에서 품종에 대한 분쟁을 완전히 피할 수 없다. F1 잡종 종자의 상태는 종자 생산의 표준화 부족, 종자 생산 분야의 불안정, 종자의 기계적 혼합물 등의 여러 가지 요인으로 인해 종종 낮게 나타난다. 재배 실험은 정확하고 신뢰할 수 있지만, 전체 과정은 너무 길고 비용이 많이 발생한다. 분자 마커의 동정은 매우 정확 하지만, 이들은 표적 유전자 서열에 대한 정확한 정보가 필요로 한다. 따라서 정확하고 빠른 비파괴 방법을 개발할 필요가 있다. FT-ICR MS(Fourier transform ion cyclotron resonance mass spectrometry)은 대사체 분석에 대한 강력한 방법이고, 유전자의 변화에 결과를 확인할 수 있는 기술이다(Mungur et al. 2005). FT-ICR MS의 단점은 대사체 분석 실험에 일반적으로 사용되기 힘든 높은 비용에 있다(Lei et al. 2011). 그것에 비해 FT-IR은 화학, 농업, 제약, 식품 산업 등 다양한 연구 분야에서 널리 사용되고 있다.

## 2. 3. Fourier transform infrared(FT-IR) spectroscopy와 partial least squares (PLS) regression 기법

최근 오믹스 연구기법의 발달에 힘입어, 대사체 수준에서 시료간의 양적, 질적 패턴 차이를 구명하는 대사체학 연구가 활발하게 이루어지고 있다. 대사체 연구는 주로 Fourier transform infrared(FT-IR) spectroscopy, proton nuclear magnetic resonance(HNMR) spectroscopy 및 mass spectrometry(MS)등의 분석 데이터를 이용하여 진행되고 있으며 각종 시료의 식별 및 대사체 마커 개발에 활용이 되고 있다(Krishnan et al., 2005). 특히, FT-IR 분석 방법은 스펙트럼이 가지고 있는 다양한 정보를 다변량통계분석을 이용하여 신속하고 정확하게 구분할 수 있다(Gallardo-Velázquez et al., 2009). 이 장치는 1800년 William Herschel에 의해 햇빛의 가열 효과 측정에서 발견되었지만, IR 분광을 빛의 파장과 상호작용하는 IR 영역에 대한 측정이 적합한 검출기를 구축하는데 어려움이 있었다. 이에 1903년, William W. Coblentz은 무기 화합물과 유기 화합물을 IR 스펙트럼으로 정확한 측정을 시작하였다. 1930년대에 들어서 IR 분광법이 나타났고, 이에 첫 번째 타입의 IR 분광계가 등장하였다. 합성 고무 산업이 발달하면서 IR 분광계에 대한 필요성이 나타났고, 빠르게 개발이 진행되었다. 이 개발을 통해 IR 스펙트럼의 대한 모든 기능에 깊은 이론 연구가 진행될 수 있는 동기가 부여되었다. 1950년대, IR 분광법은 학술 및 산업 실험실의 주요 분석 방법으로 설립되었습니다. 1970년대에 사용되는 대부분의 IR 계측은 prism이나 grating monochromators이 기반이 되었다. Fourier transform infrared (FT-IR) spectrometers의 도입은 IR 기술의 혁신으로 작용하였고, 프랑스 수학자 Fourier에 의해 스펙트럼의 정보가 수학적 절차에 의해 변환될 수 있음을 증명하였다. 하지만, Fourier의 스펙트럼 계산은 수치상 복잡성이 존재해 1949년까지도 사용되기 어려웠다. 1970년에 천문학자 Fellgett이 천문학에 FT-IR을 사용하면서 재조명 받았다. 이후 컴퓨터가 발달하면서 IR 스펙트럼의 해상도가 증가하고, 고속데이터 수집과 에너지 처리 등이 개선되면서 FT-IR에 사용이 증가하였다(Colthup et al. 1990; Griffiths et al. 1986; Silverstein et al. 1981; Smith. 1979)(Fig. 1).



**Fig. 1** Significant events leading to the development of current FT-IR spectrometers and microspectrometers.

분석기술과 함께 동반되는 기술이 Chemometrics이다. Chemometrics란 화학적인 데이터들이 나타내는 정보를 추출하는 방법이다. 이것은 다변량통계,

응용 수학, 컴퓨터 과학과 같은 코어 데이터 분석 분야에 사용되는 방법을 사용하여 화학, 생화학, 약품, 생물학, 화학 기술 등의 문제들을 해결할 수 있다. Chemometrics는 특히 화학, 생명 과학 실험에서 서술과 예측 문제를 해결하기 위해 적용된다. 서술은 화학 시스템의 특성에 관계와 구조를 분석하는 목적으로 모델링되고, 예측은 화학 시스템의 특성에 대한 새로운 속성이나 행동을 예측하는 모델링이 된다. 두 경우 모두 데이터 세트가 관찰의 변수만큼 많은 변수를 나타낼 수 있고, 종종 크고 복잡하게 나타날 수 있다. 대표적으로 principal components analysis (PCA)와 partial least-squares (PLS)를 이용하여 데이터를 분석하고 상관성을 표현한다. PCA는 스펙트럼으로 나타나는 데이터를 수치화 시켜 평면상에 샘플들 간에 유효성을 판단할 수 있는 기술이고, PLS는 regression, clustering 그리고 pattern recognition을 통해서 다양한 표현과 유효관계를 판단할 수 있게 하는 기술이다. 이 기술들은 실험적으로 나타날 수 있는 오류나 패턴 변화 등의 모든 변수를 수학적 통계의 적용으로 조절해주는 효과가 있다.

따라서 대사체학은 대사성분의 작용기작 구명 등 기초 학문적 활용은 물론 농산물의 품종, 원산지, 재배연령, 진위 식별, 식품이나 의약품 산업의 표준화 및 안정성 관련 다양한 산업적 활용이 이루어지고 있다. 대사체 분석데이터로부터 보다 유용한 생물학적 정보를 추출하기 위하여 다양한 다변량통계분석 기법이 필수적으로 동반되고 있다. 이 중에서 대표적인 다변량통계분석 기법 중 하나인 partial least squares (PLS) regression은 시료로부터 정확한 정량분석 데이터와 동일 시료의 스펙트럼 데이터의 상관분석을 통해 다양한 성분들의 함량 예측 모델링 방법으로 활용이 이루어지고 있다(Bastiena et al., 2005; Höskuldsson, 1988; Mevik and Wehrens, 2007; Wold et al., 2001). 최근에는 PLS regression 기법을 이용하여 동백나무속 식물(*Camellia oleifera*)에서 지방산 함량 예측 모델링(Yuan et al., 2013), 사과, 오렌지 그리고 복숭아 주스에 함유되어 있는 carotenoid 성분 함량 예측 모델링(Leopold et al., 2011), 버섯에 존재하는 단당류와 다당류 함량 예측 모델링(Chen et al., 2012), 육류의 부패 및 손상에 대한 예측 모델링(Argyri et al., 2013), 아프리카칸 암의 기능성 성분에 대한 예측(Song et al., 2014)이 가능한 것으로 보고되고 있으며 예측정확도 또한 매우

높아 산업적 활용 측면에서 실용화 가능성이 매우 높아지고 있는 실정이다.

따라서 본 연구는 재배되고 있는 작물들의 품종, 기능성 성분, 교배 육종 그리고 당산도에 대한 대사체 정보를 이용하여 품종을 식별하고, 기능성 성분 함량과 당산도 함량을 예측하고 더 나아가 교배 육종에 대한 정보를 분석하는데 있다.

### **3. FT-IR 스펙트럼 기반 다변량통계분석기법에 의한 두과작물의**

#### **대사체 수준 식별체계 확립**

**(Establishment of rapid discrimination system of leguminous plants at metabolic level using FT-IR spectroscopy based multivariate analysis)**

### 3. 1. Abstract

To determine whether FT-IR spectral analysis based on multivariate analysis for whole cell extracts can be used to discriminate major leguminous plant at metabolic level, seeds of six major leguminous plants were subjected to Fourier transform infrared spectroscopy (FT-IR). FT-IR spectral data from seeds were analyzed by principal component analysis (PCA), partial least square discriminant analysis (PLS-DA) and hierarchical clustering analysis (HCA). The hierarchical dendrogram based on PLS-DA separated the six leguminous plants into four major groups. The first group consisted of *Vigna radiata* var. *radiata*, *Vigna angularis* var. *angularis* and *Vigna unguiculata* subsp. *unguiculata*, whereas *Pisum sativum* var. *sativum*, *Glycine max* L and *Phaseolus vulgaris* var. *vulgaris* were clustered into a separate group respectively. These results showed that metabolic classification system were in accordance with known phylogenic taxonomy. Thus we suggested that the hierarchical dendrogram based on PLS-DA of FT-IR spectral data from seeds represented the most probable chemotaxonomical relationship between six leguminous plants. Furthermore these metabolic discrimination systems could be applied for rapid selection and classification of useful leguminous cultivars.

### 3. 2. Introduction

두과작물은 전세계적으로 48%의 기름 생산량을 차지하고 있는 경제적으로 중요한 작물이다(Singh and Hymowiz 1988). 또한 두과작물은 단백질 40%, 탄수화물 30% 그리고 지방 20%를 함유하는 고기능성 식량작물이다(Jang et al. 2010). 두과작물은 우리나라 전역에서 재배되고 있으며, 오랜 세월 지역적 특성에 따라 다양한 품종들이 농가에서 재배되고 있다(Kim et al.1993; Lee and Kim 1993). 최근 국산 콩을 이용하는 소비자가 증가하면서 국산 콩에 대한 관심이 급증하고 있지만 우리나라의 콩 재배면적 감소에 따른 생산량 감소로 식용콩 자급률은 32.5%에 불과하여 외국산 수입 콩의 비중이 크게 증가하고 있는 실정이다(Im and Lee 2011). 이와 더불어 유전자 변형 농산물에 대한 소비자의 식품 안전성에 대한 우려도 크게 증가하고 있어서 국산 콩과 수입 콩의 원산지 식별은 물론 다양한 기능성 두과작물의 신제품 개발을 위해서도 신속한 품종 및 계통 식별방법의 개발이 요구되고 있다.

콩의 품종 및 변종의 식별 방법으로 RFLP, RAPD, SSR, AFLP와 같은 분자마커를 이용한 구분 방법들이 개발되어왔다(Keim and Shoemaker 1988; Thompson and Nelson 1998; Kim et al. 2006; Choi et al. 2000). 또한 유전적 다양성 분석과 품종 구분을 위한 분자마커로 SNP 마커들이 활용되고 있다(Yoon et al. 2007). 그러나 콩에 함유되어 있는 불포화지방산, 식이섬유, 올리고당, 이소플라본, 피틴산, 사포닌, 피니톨 및 티아민과 같은 다양하게 대사물질을 대사체 마커로 이용하여 품종 선별이 이루어지고 있으나 아직 대사체 수준에서 식별 기술 개발에 대한 연구는 진행된 바가 없다(Jang et al. 2010; Hong et al. 2010). 최근 기기분석의 발달에 힘입어 대사체 분석기술에 대한 관심이 고조되고 있으며 급진적인 기술발전이 이루어지고 있다. 특히 대사체지문분석기술(metabolite fingerprinting)은 시료의 전세포추출물 대사체 정보로부터 다양한 다변량통계분석기법을 이용하여 대사체 수준에서 시료간의 양적, 질적 패턴 차이를 규명하는 연구 분야로 주로 FT-IR (Fourier transform infrared spectroscopy), HNMR (proton nuclear magnetic resonance spectroscopy) 및 MS (mass spectrometry) 등의 분석 데이터를 이용하여 시료의 식별에 활용이 되고 있다 (Krishnan et al.

2005). 특히 FT-IR 스펙트럼 데이터의 다변량 통계분석 기술은 유연관계가 높은 미생물종의 식별 (Goodacre et al. 1998; Timmins et al. 1998; Wenning et al. 2002), 식물세포벽 돌연변이주 선발 (Stewart et al. 1997; Chen et al. 1998), 고등식물의 종 구분 (Kim et al. 2004) 및 품종 식별 (Kim et al. 2007; 2009)에 활용이 가능함이 보고된 바 있다.

따라서 본 연구에서는 FT-IR 스펙트럼의 다변량통계분석 기법을 이용하여 국내에서 재배되고 있는 주요 두과작물들의 대사체 수준에서 신속한 구분 및 식별 체계를 확립함으로써 재배환경에 따른 대사물질의 질적, 양적 변화를 규명하고 더 나아가 기능성 성분의 함량이 우수한 계통 선발을 위한 신속 품종식별 기술 발판을 구축하고자 한다.

### 3. 3. Material and Methods

#### 3. 3. 1. 식물재료

본 연구에 사용된 두과작물은 농촌진흥청 국립식량과학원에서 분양 받은 6종의 두과작물 건조종자를 사용하였다. 실험에 사용된 두과작물은 대두(*Glycine max* L; Deawon, Eunha, Sohuk), 완두(*Pisum sativum* var. *sativum*; Oi, Cheongmi, Sacheol), 강낭콩(*Phaseolus vulgaris* var. *vulgaris*; Kangnang-1ho, 3ho), 팥(*Vigna angularis* var. *angularis*; Yenkum, Chilbo, Chungju), 녹두(*Vigna radiata* var. *radiata*; Eooul, Jangan, Dasun) 그리고 동부(*Vigna unguiculata* subsp. *unguiculata*; Kochang, Hwaseong, Kimje)이며 강낭콩 2종을 제외하고는 각 종별로 3개의 품종을 대사체 분석에 사용하였다(Fig. 2). 각각의 종자시료는 막자와 막자사발을 이용하여 분말형태로 분쇄하였다. 분쇄된 콩 종자시료는 -70°C 초저온냉동고에 보관하여 사용하였다.



**Fig. 2** The seeds of leguminous plants used in the study. (A) *Glycine max* L. (B) *Pisum sativum* var. *sativum*. (C) *Phaseolus vulgaris* var. *vulgaris*. (D) *Vigna radiata* var. *radiate*. (E) *Vigna angularis* var. *angularis*. (F) *Vigna unguiculata* subsp. *Unguiculata*. Scale bars represent 1 cm.

### 3. 3. 2. 두과작물 전세포추출물 제조

대두, 완두, 강낭콩, 팥, 녹두, 동부의 각 종자 분말 20 mg을 1.5 ml Eppendorf tube에 넣고 20%(v/v) methanol 용액을 200  $\mu$ L씩 첨가하여 잘 섞어 주었다. 각 tube를 50°C water bath에서 20분간 반응시킨 다음 13,000rpm에서 15분간 원심분리한 후 상층액을 새로운 tube에 옮겼다. 새로운 tube에 옮긴 시료는 다시 한 번 13,000rpm에서 5분간 원심분리하고 찌꺼기가 들어가지 않도록 상층액을 조심스럽게 새로운 tube에 옮겼다. 추출된 추출물은 -20°C에서 보관하여 FT-IR 스펙트럼 분석에 사용하였다.

### 3. 2. 3. FT-IR 스펙트럼과 데이터 전처리 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 시료를 5  $\mu$ l씩 384-well ZnSe plate에 분주하여, 37°C hot plate에서 약 20분간 건조하였다. 건조된 ZnSe plate는 Tensor 27에 장착된 HTS-XT (Bruker Optics GmbH) 고효율 자동화 장치를 이용하여 스펙트럼을 측정하였다. 각 시료의 스펙트럼은 총 4000 ~ 400  $\text{cm}^{-1}$ 범위에서 그리고 4  $\text{cm}^{-1}$ 간격으로 총 128회 반복 측정된 평균 스펙트럼을 분석에 사용되었다. 각 시료는 각각 3반복으로 FT-IR 스펙트럼을 반복 측정하였다. FT-IR 스펙트럼 조사 및 데이터 변환에 사용된 프로그램은 Bruker 에서 제공하는 OPUS Lab (ver. 6.5, 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 2.13.2)을 이용하여 수행하였다. Baseline 교정을 위해 FT-IR 스펙트럼 분석 영역의 양 끝점 (800-1800 $\text{cm}^{-1}$ )의 흡광도를 0으로 조정하였으며 실험상의 오차를 최소화하기 위하여 각 스펙트럼을 동일 면적으로 normalization 하였다. 이후 데이터의 mean centering 과정을 거쳐 2차 미분한 다음 다변량 통계분석 분석을 위한 표준화된

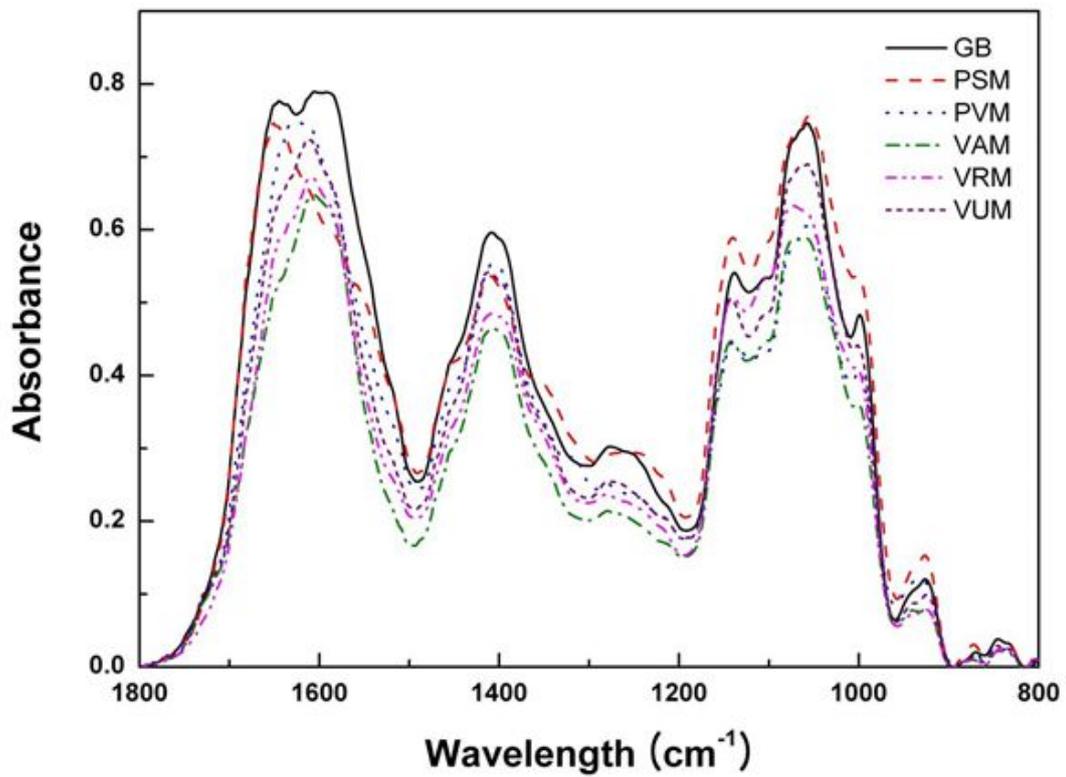
데이터로 사용하였다.

가공된 FT-IR 스펙트럼 데이터는 NIPALS 알고리즘 (Wold 1966)을 이용하여 R 프로그램(version 2.13.2)을 사용하여 PCA (Principal component analysis)와 PLS-DA (Partial least squares discriminant)분석을 수행하였다 (Fiehn et al. 2000; Trygg et al. 2007). PCA 및 PLS-DA 분석을 통해 얻어진 score를 이용하여 HCA (hierarchical clustering analysis) 분석하였으며 유사도 지수로 UPGMA (unweighted pair group method with arithmetic mean analysis)를 사용한 euclidean distance를 측정하여 각 시료의 유연관계를 덴드로그램으로 나타냈다.

### 3. 4. Results and Discussion

#### 3. 4. 1. 두과작물의 FT-IR 스펙트럼 비교 분석

두과작물로부터 전세포추출물의 FT-IR 스펙트럼 데이터의 다변량통계분석을 통해 주요 두과작물의 대사체 수준에서 유연관계 규명 및 식별체계를 확립하였다. 대두, 완두, 강낭콩, 팥, 녹두, 동부 등 주요 두과작물은 FT-IR 스펙트럼상의 1500 - 1700, 1300 - 1500, 950 - 1100  $\text{cm}^{-1}$ 부위에서 대사체의 양적, 질적 패턴 변화가 크게 이루어짐을 알 수 있었다(Fig. 3). FT-IR 스펙트럼의 1500 - 1700, 1300 - 1500 그리고 950 - 1100  $\text{cm}^{-1}$ 부위는 각각 Amide I 과 II을 포함하는 아미노산 및 단백질계열의 화합물, phosphodiester group을 포함한 핵산 및 인지질 그리고 단당류나 복합 다당류를 포함하는 carbohydrates 계열의 화합물들의 질적, 양적 정보를 반영하는 부위이다(Parker 1983; Dumas and Miller 2003; Wolkers et al. 2004; Yee et al. 2004; D'Souza et al. 2008; Lopez-Sanchez et al. 2010). 즉 FT-IR 스펙트럼 상의 질적, 양적 차이는 각 두과작물별로 종자내에 함유되어 있는 아미노산이나 단백질, 지방산, 그리고 탄수화물계통의 화합물들의 질적, 양적 차이가 현저하게 나타남을 의미한다. 따라서 FT-IR 스펙트럼 분석은 두과작물의 주요 대사체의 질적, 양적 변화를 예측하고 함량이 높은 계통의 선발 수단으로 신속한 활용이 가능할 것으로 기대된다.

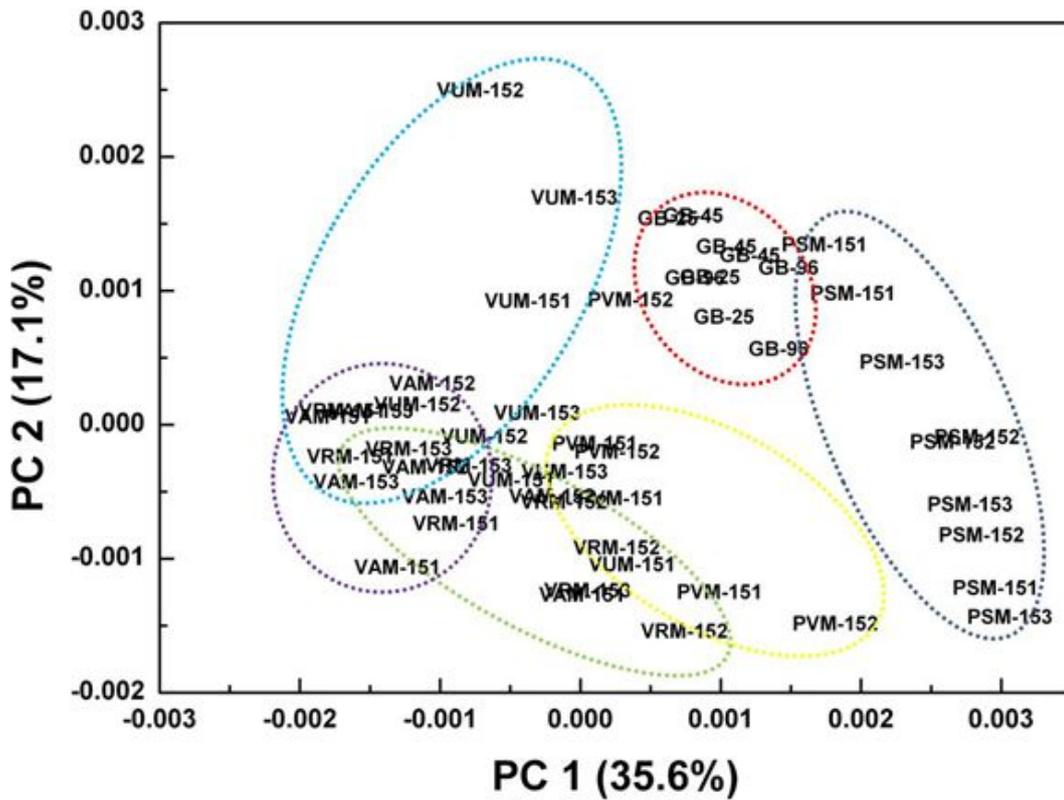


**Fig. 3** Representative FT-IR spectra from leguminous plants. (A) GB : *Glycine max* L. (B) PSM : *Pisum sativum* var. *sativum*. (C) PVM : *Phaseolus vulgaris* var. *vulgaris*. (D) VAM : *Vigna angularis* var. *angularis*. (E) VRM : *Vigna radiata* var. *radiate*. (F) VUM : *Vigna unguiculata* subsp. *Unguiculata*.

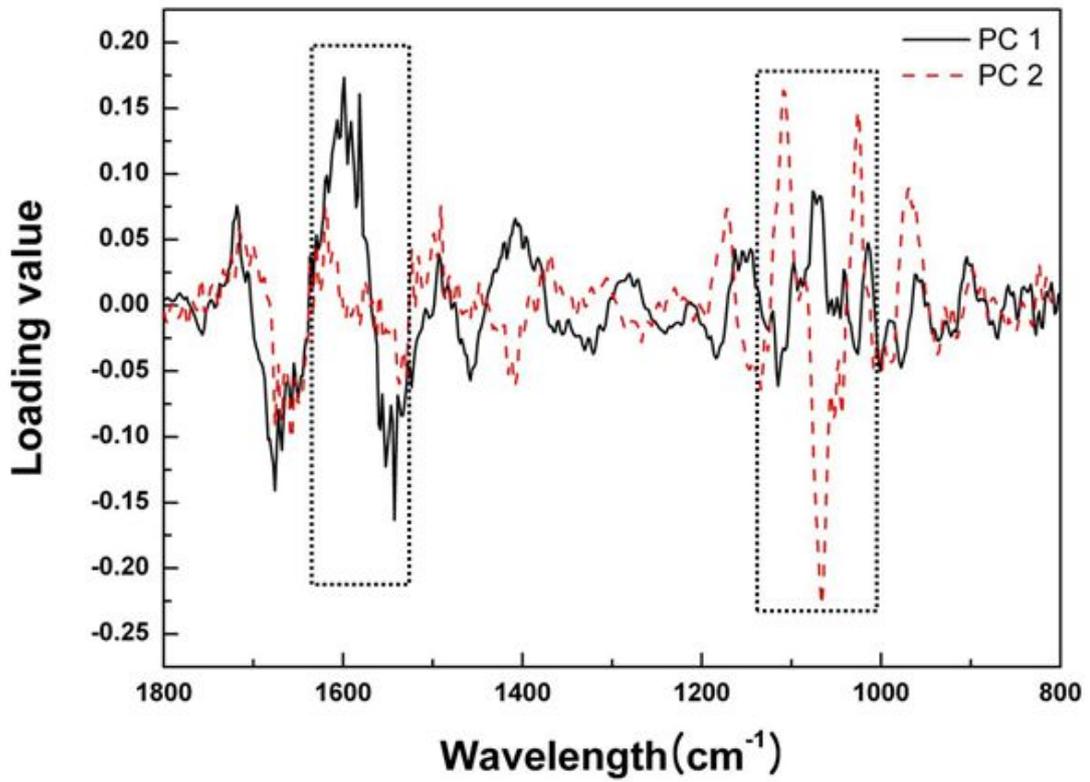
### 3. 4. 2. 두과작물의 FT-IR 스펙트럼 다변량 통계분석 및 대사체 수준 유연관계 비교

6 종의 두과작물 FT-IR 스펙트럼 데이터의 PCA 분석 결과 PC 1과 2 score 는 각각 35.6%, 17.1%의 설명력을 갖고 있으며 전체 변이량의 약 52.7%를 반영하고 있음을 알 수 있었다(Fig. 4). 비록 전체 변이량 전부를 설명할 수는 없었지만 각 두과작물 종의 영역내에 각각의 반복구는 물론 각 종에 속하는 3개 품종 시료들이 분포하는 것으로 보아 대사체 수준에서 각 두과작물의 종 식별이 이루어지고 있음을 알 수 있었다. 대두콩, 강낭콩 및 완두콩의 경우 각 종의 영역이 중첩되지 않고 별도의 그룹을 형성하였으나 반면에 팥, 녹두, 동부는 그룹의 중첩이 이루어지면서 각 종별로 완전한 식별이 불가능하였다. 이는 팥, 녹두, 동부의 경우 대사체 수준에서 유연관계가 콩보다는 매우 높음을 의미하는 것이다. 따라서 이 결과는 팥, 녹두, 동부가 분류체계상 동일한 속에 포함되는 계통분류학적 유연관계가 높은 식물종들이라는 사실과 일치하는 것으로 대사체 수준에서 유연관계가 계통분류학적 유연관계를 반영하고 있으며 대사체 수준에서 마커 탐색을 통하여 두과작물의 식별체계 확립이 가능할 것으로 기대된다.

각 두과작물의 대사체 수준 식별에 있어 중요한 FT-IR 스펙트럼 부위를 조사하였다(Fig. 5). FT-IR loading value를 보면 PC 1과 2의 분류하는 기준을 확연히 볼 수 있다. 두과작물의 식별에 있어 중요한 FT-IR 스펙트럼 부위는 1650 - 1550 과 1150 - 1000  $\text{cm}^{-1}$ 부위임을 알 수 있었다(Fig. 5). 이 부위는 FT-IR 스펙트럼상에서 차이를 보였던 부위 (Fig. 3)와 거의 일치하는 것으로 amide I, II와 carbohydrates 계열의 화합물들의 질적, 양적 차이가 두과작물의 대사체 수준 식별에 중요한 역할을 하고 있음을 알 수 있었다(Fig. 5).



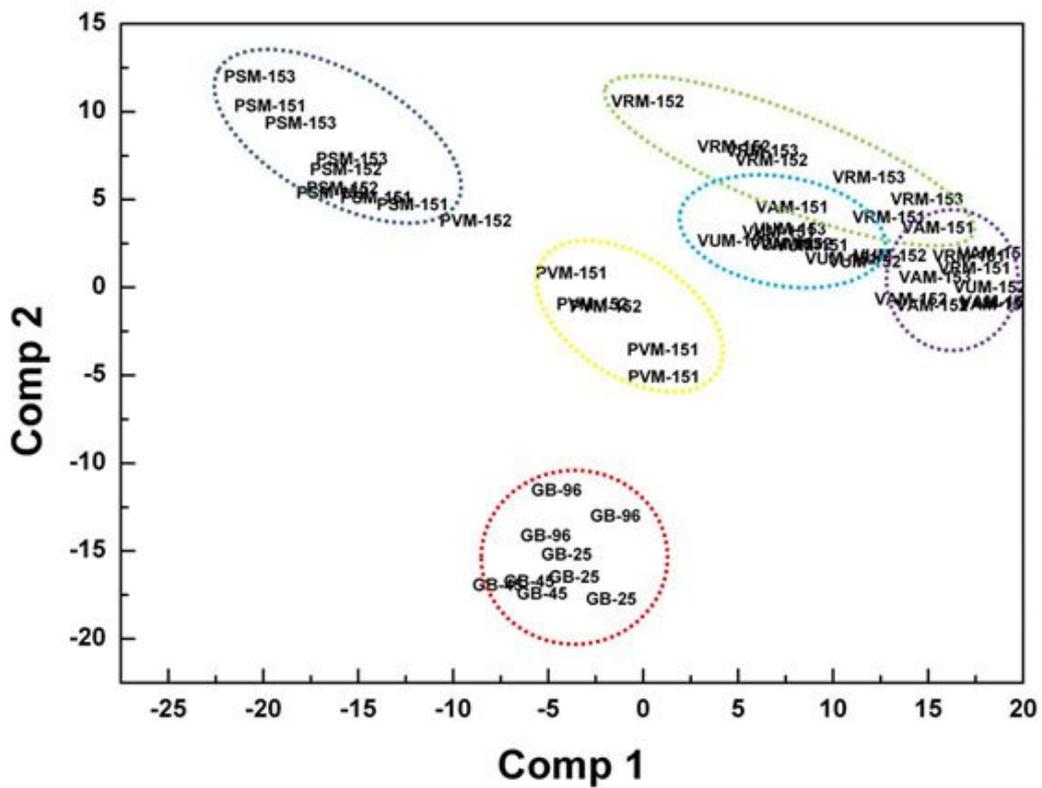
**Fig. 4** PCA score plot of FT-IR data from leguminous plants. (A) GB : *Glycine max* L. (B) PSM : *Pisum sativum* var. *sativum*. (C) PVM : *Phaseolus vulgaris* var. *vulgaris*. (D) VAM : *Vigna angularis* var. *angularis*. (E) VRM : *Vigna radiata* var. *radiata*. (F) VUM : *Vigna unguiculata* subsp. *Unguiculata*. Dot circles in figure represent boundary of each leguminous plant.



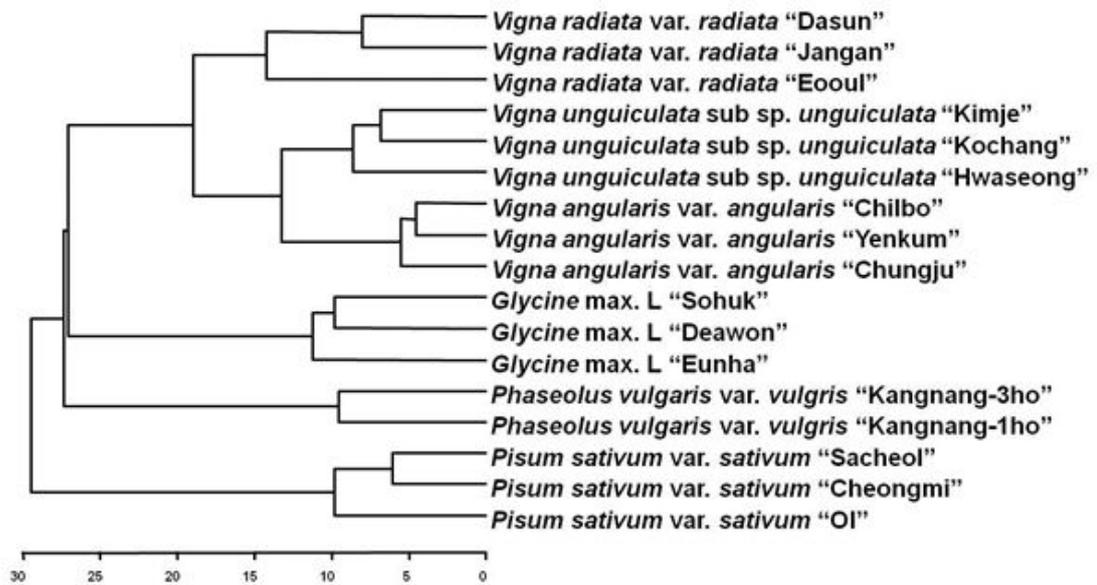
**Fig. 5** Loading value plot of PCA analysis from FT-IR data of leguminous plants. Dot rectangles represent significant FT-IR spectral region for metabolic discrimination of each leguminous plant.

PLS-DA 분석의 경우 PCA 분석 보다 두과작물의 종간 식별이 뚜렷하게 나타났다(Fig. 6). PLS-DA score plot을 보면 대두콩, 완두, 강낭콩은 각각 별도의 그룹을 형성하지만 녹두, 팥, 동부는 하나의 그룹을 형성하였다. 즉 녹두, 동부, 팥이 콩에 비하여 상대적으로 대사체 수준에서 유연관계가 높음을 알 수 있었다(Fig. 6). 이처럼, FT-IR 스펙트럼 데이터 이용한 PLS-DA 분석으로 두과작물 간의 종 구분이 가능함을 알 수 있었다. HCA dendrogram 분석 결과를 보면 총 17개의 두과작물 시료들이 4개의 주요 그룹으로 구분이 이루어짐을 알 수 있었다(Fig. 7). 녹두, 동부, 팥이 하나의 그룹을 형성하고 있고, 대두콩, 강낭콩, 완두콩이 각각 그룹을 형성하고 있었다. 이상의 연구결과로 볼 때 두과작물의 FT-IR 스펙트럼 다변량통계분석을 통해 대사체 수준에서 두과작물의 식물 종 구분이 가능함을 알 수 있었다. SSR 마커를 이용한 야생콩의 유전적 다양성과 유연관계를 분석한 결과를 보면, 한국, 일본, 중국에서 수집된 야생콩의 유연관계는 중국 야생콩이 단독적으로 그룹을 형성하고, 한국, 일본 야생콩이 그룹 간에 경계가 중첩되는 패턴을 볼 수 있었고, 중국 야생콩보다 일본 야생콩이 유전적으로 유사도가 높다는 것이 보고된 바 있다(Kim et al. 2010). 그리고 최근 콩에 존재하는 mannose와 galactose 함량비가 지리적 경향, 환경적 요인에 변화될 수 있으며 육종을 통해 고함유 계통의 선발이 가능함이 보고된 바 있다(Kim 2011). 또한, FT-IR 스펙트럼 데이터를 이용하여 바위술을 PLS 분석을 통해서 종간 유연관계를 조사한 보고에 따르면 개체 간에 대사물질에 대한 차이를 알 수 있었고, 분류학적 유연관계가 거의 일치하고, 예측의 정확도가 92.9% 높게 나타났다고 보고된 바 있다(Kim et al. 2011).

따라서 본 연구에서 확립된 대사체 수준에서 두과작물의 식별 기술은 각 두과작물로부터 유용 기능성 물질의 함량이 높은 품종, 계통의 신속한 선발 수단으로 활용이 가능할 것으로 기대되며 육종을 통한 기능성 품종개발 가속화에 기여할 수 있을 것으로 예상된다. 아울러 본 기술은 타작물의 기능성 품종 육성에도 효과적인 수단으로 활용이 가능할 것으로 기대된다.



**Fig. 6** PLS-DA score plot of FT-IR data from leguminous plants. (A) GB : *Glycine max* L. (B) PSM : *Pisum sativum* var. *sativum*. (C) PVM : *Phaseolus vulgaris* var. *vulgaris*. (D) VAM : *Vigna angularis* var. *angularis*. (E) VRM : *Vigna radiata* var. *radiata*. (F) VUM : *Vigna unguiculata* subsp. *unguiculata*. Dot circles in figure represent boundary of each leguminous plant.



**Fig. 7** Hierarchical dendrogram of FT-IR data from leguminous plants.

**4. FT-IR 스펙트럼 데이터의 다변량통계분석을 이용한  
고기능성 아프리카 감 식별 및 기능성 성분 함량 예측 모델링  
(Discrimination of High Functional African Yams using FT-IR Fingerprinting  
Combined by Multivariate Analysis and Quantitative Prediction of Functional  
Compounds by PLS Regression Modeling)**

#### 4. 1. Abstract

We established a high throughput screening system of African yam tuber lines which contain higher contents of total carotenoids, flavonoids, and phenolic compounds using ultraviolet-visible (UV-VIS) spectroscopy and Fourier transform infrared (FT-IR) spectroscopy in combination with multivariate analysis. The total carotenoids contents from 62 African yam tubers varied from 0.01 to 0.91mg·g<sup>-1</sup> dry weight (wt). The total flavonoids and phenolic compounds also varied from 12.9 to 229mg·g<sup>-1</sup> and from 0.29 to 5.2mg·g<sup>-1</sup> dry wt. FT-IR spectra confirmed typical spectral differences between the frequency regions of 1,700–1,500, 1,500–1,300 and 1,100-950 cm<sup>-1</sup>, respectively. These spectral regions were reflecting the quantitative and qualitative variations of amide I, II from amino acids and proteins (1,700–1,500cm<sup>-1</sup>), phosphodiester groups from nucleic acid and phospholipid (1,500–1,300cm<sup>-1</sup>) and carbohydrate compounds (1,100-950cm<sup>-1</sup>). Principal component analysis (PCA) and subsequent partial least square-discriminant analysis (PLS-DA) were able to discriminate 62 African yam tuber lines into three separate clusters corresponding to their taxonomic relationship. The quantitative prediction modeling of total carotenoids, flavonoids, and phenolic compounds from African yam tuber lines were established using partial least square regression algorithm from FT-IR spectra. The regression coefficients( $R^2$ ) between predicted values and estimated values of total carotenoids, flavonoids and phenolic compounds were 0.83, 0.86, and 0.72, respectively. These results showed that quantitative predictions of total carotenoids, flavonoids, and phenolic compounds were possible from FT-IR spectra of African yam tuber lines with higher accuracy. Therefore we suggested that quantitative prediction system established in this study could be applied as a rapid selection tool for high yielding African yam lines.

## 4. 2. Introduction

아프리칸 양은 주로 아프리카 지역에서 재배되고 있는 다년생작물로, 카사바, 감자, 고구마와 같이 전분 함량이 높아 중요한 식량자원으로 이용되고 있다. 양의 구근은 전분질 15-20%, 단백질 1-1.5%, 지질 1% 및 기타 미네랄 및 비타민 등을 함유하고 있다(Adeleke, 2010; Kwon et al., 2010). 또한 양의 구근에는 다양한 약리적 활성을 갖는 유용 기능성물질, saponin, tannin, sitosterol, diosgenin, carotenoid, flavonoid, polyphenol, allantoin, araginine, mucin 등을 함유하고 있어 약용작물로서 활용 가치가 매우 높은 식물이다(Kim et al., 2009; Kum et al., 2006; Kwon et al., 2010). 특히 양의 구근에 존재하는 다양한 페놀성 화합물들은 자유 라디칼 제거, 항산화 효과 및 항염증 작용 등 다양한 약리적 기능을 가지고 있다(Yang et al., 2009; Yao et al., 2004). 따라서 양은 자양, 강장, 폐결핵, 고혈압, 당뇨병, 스트레스와 피로 회복에 효능이 알려져 있다(Song et al., 2010; Yoon et al., 1989). 이처럼 아프리카 양은 식용 및 약용작물로서 가치가 높은 작물이지만 주요 성분분석에 대한 정확한 분석이 활발하게 이루어지지 않고 있다. Im et al. (1995)은 아프리카 양의 경우 사포닌, 알칼로이드, 페놀성 성분이 국내산 양에 비해 높아서 보다 쓴맛이 강하다는 보고가 있다. 따라서 식량 및 약리적 측면에서 우수한 아프리카 양의 농업적 활용 기반 구축을 위해 다양한 아프리카산 양의 기능성 성분 분석이 이루어져야 할 것으로 사료된다. 아울러 기능성 성분분석 데이터를 기반으로 고기능성 양 라인의 신속한 선발시스템 확립이 이루어져야 할 것으로 사료된다.

유용 기능성 물질들의 정량 및 정성분석은 주로 gas chromatography(GC)나 high performance liquid chromatography(HPLC) 및 질량분석기와 연계를 통해 정확한 성분의 정량 및 정성 분석이 이루어진다(Aktumsek et al., 2013; Kimura et al., 2002). 그러나 이러한 분획 컬럼을 통한 기능성 성분 정량 분석의 경우 분석 결과는 매우 정확하지만, 상대적으로 시료의 준비과정이 복잡하며, 추출 및 시료 분석 비용이 많이 소요되고, 다수의 시료를 분석할 경우 많은 시간이 요구되는 한계점을 가지고 있다(Yuan et al., 2013). 이러한 정량 및 정성분석의 복잡성 문제를 해결하기 위한 하나의 대안으로 ultraviolet-visible (UV-VIS)

spectrophotometer를 이용한 기능성 성분들의 신속한 함량 조사가 가능함이 보고되고 있다(Kofalvi and Nassuth, 1995; Stadnik and Buchenauer, 2000; Zhishen et al., 1999).

최근 오믹스 연구기법의 발달에 힘입어, 대사체 수준에서 시료간의 양적, 질적 패턴 차이를 구명하는 대사체학 연구가 활발하게 이루어지고 있다. 대사체 연구는 주로 Fourier transform infrared(FT-IR) spectroscopy, proton nuclear magnetic resonance(HNMR) spectroscopy 및 mass spectrometry(MS)등의 분석 데이터를 이용하여 진행되고 있으며 각종 시료의 식별 및 대사체 마커 개발에 활용이 되고 있다(Krishnan et al., 2005). 특히, FT-IR 분석 방법은 스펙트럼이 가지고 있는 다양한 정보를 다변량통계분석을 이용하여 신속하고 정확하게 구분할 수 있다(Gallardo-Velázquez et al., 2009). 따라서 대사체학은 대사성분의 작용기작 구명 등 기초학문적 활용은 물론 농산물의 품종, 원산지, 재배연령, 진위 식별, 식품이나 의약품 산업의 표준화 및 안정성 관련 다양한 산업적 활용이 이루어지고 있다. 대사체 분석데이터로부터 보다 유용한 생물학적 정보를 추출하기 위하여 다양한 다변량통계분석 기법이 필수적으로 동반되고 있다. 이 중에서 대표적인 다변량통계분석기법중 하나인 partial least squares (PLS) regression은 시료로부터 정확한 정량분석 데이터와 동일 시료의 스펙트럼 데이터의 상관분석을 통해 다양한 성분들의 함량 예측 모델링 방법으로 활용이 이루어지고 있다(Bastiena et al., 2005; Höskuldsson, 1988; Mevik and Wehrens, 2007; Wold et al., 2001). 최근에는 PLS regression 기법을 이용하여 동백나무속 식물(*Camellia oleifera*)에서 지방산 함량 예측 모델링(Yuan et al., 2013), 사과, 오렌지 그리고 복숭아 주스에 함유되어 있는 carotenoid 성분 함량 예측 모델링(Leopold et al., 2011), 버섯에 존재하는 단당류와 다당류 함량 예측 모델링(Chen et al., 2012), 육류의 부패 및 손상에 대한 예측 모델링(Argyri et al., 2013) 이 가능한 것으로 보고되고 있으며 예측정확도 또한 매우 높아 산업적 활용 측면에서 실용화 가능성이 매우 높아지고 있는 실정이다.

따라서 본 연구에서는 FT-IR 스펙트럼 데이터의 다변량통계분석 기법을 이용하여 아프리카 안의 신속한 구분 및 식별체계를 확립하고 더 나아가 아프리카 안의 구근에 존재하는 total carotenoids, flavonoids, phenolic

compounds의 신속한 함량예측 모델링을 개발하여 기능성 성분이 우수한 아프리카 안의 신속한 선별체계를 확립하고자 하였다.

### 4. 3. Material and Methods

#### 4. 3. 1. 식물재료

본 연구에 사용된 아프리카 감(Dioscorea)은 탄자니아 넬슨만델라 아프리카과학기술원 생명공학대학으로부터 제공받은 아프리카 감 구근을 사용하였다. 실험에 사용된 각각의 감 라인은 *Dioscorea alata* 15라인, *D. bulbifera* 6라인, *D. cayensis* 5라인, *D. dumetorum* 13라인, *D. rotunalata* 12라인, *D. mangenotiana* 1라인, *D. prahensilis* 2라인 그리고 분류학적 동정이 이루어지지 않은 케냐의 감 시료 8라인 등 총 62라인을 사용하였다(Table 1). 건조된 아프리카 감 구근은 막자와 막자사발을 이용하여 미세한 분말로 분쇄하였으며 -70°C 초저온냉동고에 보관하면서 실험에 사용하였다.

#### 4. 3. 2. UV-VIS spectrophotometer를 이용한 chlorophyll a+b/carotenoids 측정

아프리카 감으로부터 total carotenoids 함량은 Yang et al. (1998)의 방법을 이용하여 UV-VIS spectrophotometer (DU<sup>®</sup>730, BECKMANCOULTER, Brea, CA, America)를 통해 조사하였다. 아프리카 감 분말 시료 20mg을 80% acetone 1mL이 들어있는 tube에 넣어주고 용매와 잘 섞이도록 흔들어 주었다. 용매와 잘 섞인 시료를 상온에서 13,000rpm으로 5분간 원심분리 하였다. 원심분리가 끝난 후 시료의 상층액을 새로운 tube에 옮겨 주었다. 새로운 tube에 옮겨진 추출물을 흡광도가 0.3-0.85이 되도록 희석하였고, 350-750nm 측정범위에서 >2nm 간격으로 UV-VIS spectrophotometer를 이용하여 흡광도를 측정하였다. 측정된 흡광도로부터 Lichtenthaler and Buschmann (2001)의 방법을 이용하여 아프리카 감 시료로부터 total carotenoids 함량을 조사하였고 함량계산식은 아래와 같다.

$$C_a \text{ (mg}\cdot\text{L}^{-1}\text{)}=12.25\cdot A_{663}-2.79\cdot A_{647}$$

$$C_b \text{ (mg}\cdot\text{L}^{-1}\text{)}=21.50\cdot A_{647}-5.1\cdot A_{663}$$

$$C_{(x+c)} \text{ (mg}\cdot\text{L}^{-1}\text{)}=(1000\cdot A_{470}-1.82\cdot C_a-85.02\cdot C_b)/198$$

함량계산식에 사용된  $C_a$ 는 chlorophyll a,  $C_b$ 는 chlorophyll b를 나타낸다. 또한,  $C_{(x+c)}$ 에서  $x+c$ 는 xanthophylls와 carotenes의 합을 total carotenoids 함량을 나타낸다. 측정된 값은 ANOVA를 이용하여 유의성을 판별하였다.

#### 4. 3. 3. UV-VIS spectrophotometer을 이용한 total flavonoids 측정

아프리카 안으로 부터 total flavonoids 함량은 Zhishen et al. (1999)의 방법을 이용하여 UV-VIS spectrophotometer를 통해 조사하였다. 아프리카 안 시료 10mg을 absolute ethanol 1mL 들어가 있는 tube에 넣어주었다. 용매와 잘 섞이도록 흔들어준 후 1시간 동안 상온에서 반응 시켰다. 반응이 끝난 시료를 상온에서 13,000rpm으로 5분간 원심분리하고, 원심분리가 끝난 시료의 상정액 100 $\mu$ l를 새로운 tube에 옮겼다. 옮겨진 시료 100 $\mu$ l에 2차 증류수 400 $\mu$ l를 첨가하여 혼합한 다음 5%  $\text{NaNO}_2$  30 $\mu$ l를 첨가하여 상온에서 10분간 반응 시켰다. 10분간 반응이 끝난 후 10%  $\text{AlCl}_3$ 용액 30 $\mu$ l를 첨가하여 상온에서 다시 1분간 반응 시켰다. 반응이 끝난 후 1M NaOH 200 $\mu$ l를 첨가하여 반응액의 최종 부피가 1mL가 되도록 2차 증류수 240 $\mu$ l를 첨가하였다. 준비된 시료는 UV-VIS spectrophotometer를 이용하여 510nm 파장에서 흡광도를 측정하였다. Reference compounds로는 catechin(100mg·L<sup>-1</sup>)을 사용하여 조사하여 total flavonoids 함량을 측정하였다. 측정된 값은 ANOVA를 이용하여 유의성을 판별하였다.

#### 4. 3. 4. UV-VIS spectrophotometer을 이용한 total phenolics 함량 측정

아프리카 안으로 부터 total phenolics compounds 함량은 Wu et al. (2007)의 방법을 이용하여 UV-VIS spectrophotometer를 통해 조사하였다. 아프리카 안 시료 10mg을 80% ethanol 50 $\mu$ l가 들어가 있는 tube에 넣어주었다. 용매와 잘 섞이도록 흔들어준 후 95°C에서 5분간 가열하여 추출하였다. 추출된 시료는 상온에서 12,000rpm으로 15분간 원심분리하고, 원심분리가 끝난 시료의 상정액 20 $\mu$ l를 새로운 tube에 옮겼다. 옮겨진 추출물 20 $\mu$ l에 2N Folin-Ciocalteu

reagent 50 $\mu$ l와 20% Na<sub>2</sub>CO<sub>3</sub> 100 $\mu$ l를 첨가하고 최종 부피가 1mL가 되게 2차 증류수 830 $\mu$ l를 첨가하여 준비하였다. Control로 사용될 용액은 2차 증류수 850mL에 2N Folin-Ciocalteu reagent 50 $\mu$ l와 20% Na<sub>2</sub>CO<sub>3</sub> 100 $\mu$ l를 첨가하고 최종 부피가 1mL가 되게 준비하였다. 표준물질은 20 $\mu$ l ethanol이 포함된 chlorogenic acid에 2N Folin-Ciocalteu reagent 50 $\mu$ l와 20% Na<sub>2</sub>CO<sub>3</sub> 100 $\mu$ l를 첨가하고 최종 부피가 1mL가 되도록 2차 증류수 830 $\mu$ l를 첨가하여 부피를 조정하였다. 준비된 시료는 상온에서 20분간 반응 시킨 후에 UV-VIS spectrophotometer를 이용하여 725nm 파장에서 흡광도를 조사하였다. 측정된 값은 ANOVA를 이용하여 유의성을 판별하였다.

#### 4. 3. 5. FT-IR 스펙트럼 조사 및 스펙트럼 데이터 다변량 통계분석

아프리카인 Yam 62개의 구근 분말 20mg을 tube에 넣고 20% methanol 용액을 200 $\mu$ l씩 첨가하여 잘 섞어주었다. 각 tube를 50°C 수조에서 20분간 반응시킨 다음 13,000rpm에서 15분간 원심분리한 후 상징액을 새로운 tube에 옮겼다. 회수한 상징액은 원심분리과정을 반복하여 찌꺼기가 들어가지 않도록 조심스럽게 새로운 tube에 상징액을 옮겼다. 회수된 상징액은 -20°C에서 보관한 다음 FT-IR 스펙트럼 조사에 사용하였다.

FT-IR(Fourier transform infrared) 스펙트럼 조사는 Tensor 27(Bruker Optics GmbH, Ettlingen, Germany)를 사용하였고, DTGS(deuterated triglycine sulfate) 검출기로 분석하였다. 추출된 각각의 시료 5 $\mu$ l를 384-well ZnSe plate에 분주하여, 37°C hot plate에서 약 20분간 건조하였다. 건조된 ZnSe plate는 Tensor 27에 장착된 HTS-XT(Bruker Optics GmbH) 고효율 자동화 장치를 이용하여 스펙트럼을 조사하였다. 각 시료의 스펙트럼은 총 4,000-400cm<sup>-1</sup> 범위에서 그리고 4cm<sup>-1</sup> 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. 6.5, 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, Auckland, New Zealand)을 이용하여 수행하였다. Baseline 교정을 위해 FT-IR 스펙트럼 분석 영역의 양 끝점( $1,800-800\text{cm}^{-1}$ )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. 가공된 FT-IR 스펙트럼 데이터는 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) 분석(Fiehn et al., 2000; Trygg et al., 2007)을 수행하였다.

#### **4. 3. 6. 아프리카 안 구근으로부터 total carotenoids, flavonoids 및 phenolics 함량 예측 PLS 모델링**

아프리카 안 구근의 FT-IR 스펙트럼 데이터로부터 total carotenoids, flavonoids, phenolics 함량 예측 모델을 개발하였다. 총 62개 아프리카 안 시료의 FT-IR 스펙트럼 데이터와 동일 시료로부터 얻어진 total carotenoids, flavonoids, phenolics 함량 실측 정량 데이터를 이용하여 X 변수에는 FT-IR 스펙트럼 데이터를 사용하였고, 3개의 Y 변수에는 UV-VIS spectrophotometer를 통해 얻어진 total carotenoids, flavonoids, phenolics 함량 정량 데이터를 각각 사용하였다. PLSR(partial least square regression) 분석은 R 프로그램(version 2.15.0)을 이용하였다. 예측 모델링의 정확도를 향상시키기 위하여 1개의 X 변수에 대해 3개의 Y 변수를 각각 교차검정을 실시하였다. 확립된 예측 모델링을 이용하여 각각의 아프리카 안 시료의 total carotenoids, flavonoids 및 phenolics 함량 예측을 수행하였다. 함량 예측 모델링의 정확도를 조사하기 위하여 아프리카 안 시료로부터 total carotenoids, flavonoids 및 phenolics 함량 실측치와 예측치에 대한 선형회귀분석을 수행하여 상관계수를 조사하였다.

## 4. 4. Results and Discussion

### 4. 4. 1. 아프리카 얀 구근으로부터 total carotenoids, flavonoids 및 phenolics 함량 조사

UV-VIS spectrophotometer를 이용하여 아프리카 얀 구근에 존재하는 total carotenoids, flavonoids 및 phenolics 함량을 조사하였다(Table 1). 여러 아프리카 얀 종들에 대한 각각의 성분에 대한 유의한 차이를 보이는지를 ANOVA 프로그램을 통해 평가한 결과 유의성( $P < 0.05$ )이 있는 차이를 보였다. Total carotenoids 함량은 *D. bulbifera*가  $0.61\text{mg}\cdot\text{g}^{-1}$  dry weight (wt)로 아프리카 얀 종 가운데서 가장 높았으며 다음으로 *D. prahensilis* 및 *D. dumetorum*이 각각  $0.59$  및  $0.53\text{mg}\cdot\text{g}^{-1}$  dry wt로 나타나 total carotenoids 함량이 높은 종임을 알 수 있었다. 반면에 *D. alata*와 *D. cayensis*는 다른 아프리카 얀 종들에 비해 total carotenoids 함량이 낮은 종임을 알 수 있었다(Table 1). Total flavonoids 함량의 경우, *D. dumetorum*이  $171\text{mg}\cdot\text{g}^{-1}$  dry wt 으로 가장 함량이 높은 종으로 나타났으며, *D. bulbifera* 가 다음으로 total flavonoids 함량이 높은 종임을 알 수 있었다. 아직까지 정확한 종의 동정이 이루어지지 않은 시료인 *Dioscorea* spp 의 경우 비교적 total flavonoids 함량이 높은 종임을 알 수 있었다(Table 1). *D. alata*와 *D. cayensis*는 carotenoids의 함량분석 결과와 마찬가지로 total flavonoids 함량이 가장 낮은 종으로 나타났다. Total phenolic 함량의 경우 *D. dumetorum*이  $3.56\text{mg}\cdot\text{g}^{-1}$  dry wt로 가장 높은 종이며, carotenoids와 flavonoids 함량이 낮았던 *D. alata*가 다른 아프리카 얀 종들에 비해 함량이 높음을 알 수 있었다(Table 1). 아프리카 얀 시료의 기능성 성분 정량분석 결과를 보면, total flavonoids와 phenolics 함량에 비해 total carotenoids 함량이 낮게 나타났으며, total flavonoids와 phenolics 함량은 각각의 아프리카 얀 종들 사이에 함량 차이가 매우 크게 나타남을 알 수 있었다(Table 1). 이 결과는 Yang et al. (2009)이 얀에 존재하는 total carotenoids, flavonoids, phenolics 함량 등 기능성 성분들의 경우 종에 따라서 질적 차이는 물론 양적 차이가 매우 크다는 보고와 일치한다.

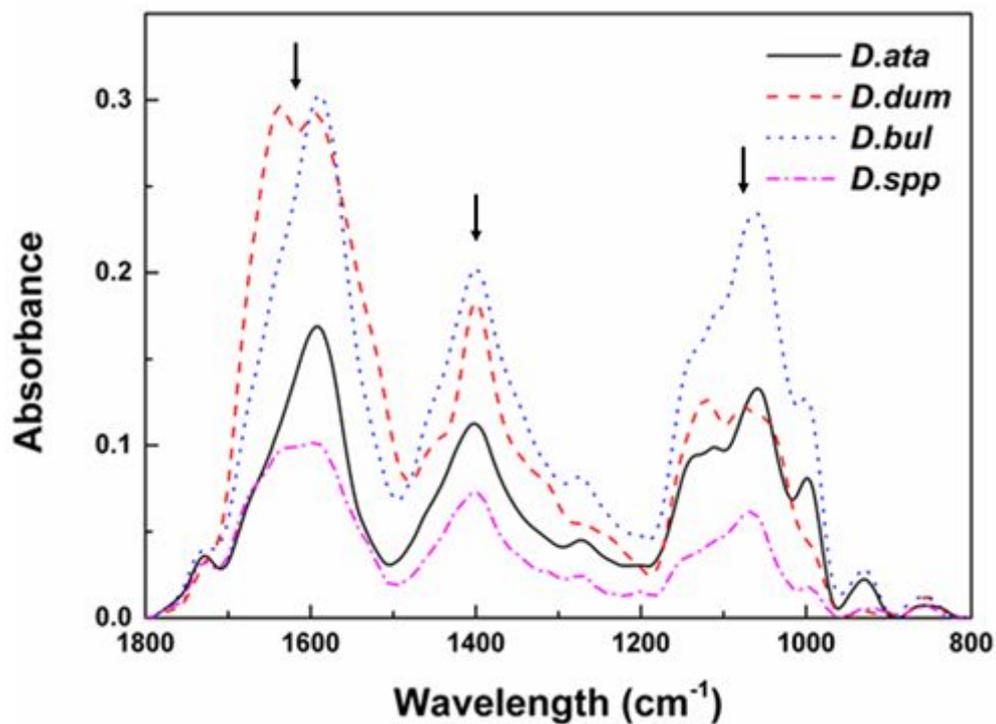
**Table 1.** Quantitative analysis of total carotenoids, flavonoids and phenolic contents from African yam tubers using UV-VIS spectrophotometric analysis. All samples were run in three replicates.

Species	Line Numbers	Compound contents		
		Total carotenoids (mg·g <sup>-1</sup> dry wt)	Total flavonoids (mg·g <sup>-1</sup> dry wt)	Total phenolic compounds (mg·g <sup>-1</sup> dry wt)
<i>D. alata</i>	15	0.09±0.05	43.87±19.53	2.01±1.44
<i>D. rotundata</i>	12	0.26±0.16	58.60±20.85	1.20±0.51
<i>D. dumetorum</i>	13	0.53±0.17	170.97±45.26	3.56±1.32
<i>D. cayenesis</i>	5	0.20±0.01	45.81±25.54	1.86±1.67
<i>D. mangenotiana</i>	1	0.73	41.94	1.83
<i>D. prahensilis</i>	2	0.60±0.14	54.84±4.56	1.87±0.78
<i>D. bulbifera</i>	6	0.61±0.22	85.48±36.14	1.73±0.99
<i>Dioscorea</i> spp	8	0.15±0.05	76.61±35.03	1.58±0.77

Data represent the mean±SD of measurements.

#### 4. 4. 2. 아프리카 안 시료의 FT-IR 스펙트럼 비교분석

아프리카 안 시료로부터 전세포 추출물의 FT-IR 스펙트럼 데이터 다변량통계분석을 통해 아프리카 안의 대사체 수준에서 중간 유연관계 구멍 및 식별체계를 확립하였다. 각 아프리카 안 시료의 FT-IR 스펙트럼을 비교한 결과 아프리카 안 시료들은 FT-IR 스펙트럼상의 1,700-1,500, 1,500-1,300, 1,100-950 $\text{cm}^{-1}$ 부위에서 대사체의 양적, 질적 패턴 변화가 크게 이루어짐을 알 수 있었다(Fig. 8). FT-IR 스펙트럼의 1,700-1,500, 1,500-1,300 그리고 1,100-950 $\text{cm}^{-1}$ 부위는 각각 아미노산 및 단백질의 amide bond I 과 II, 핵산 및 인지질로부터 phosphodiester bond 및 인을 포함하는 유기산, 그리고 단당류나 복합 다당류를 포함하는 carbohydrates 계열의 화합물들의 질적, 양적 정보를 반영한다(D'Souza et al., 2008; Dumas and Miller, 2003; Lopez-Sanchez et al., 2010; Parker, 1983; Wolkers et al., 2004; Yee et al., 2004). 즉 FT-IR 스펙트럼상의 질적, 양적 차이는 아프리카 안에 함유되어 있는 아미노산이나 단백질, 지방산, 그리고 탄수화물계통의 화합물들의 질적, 양적 차이가 현저하게 나타남을 의미한다. 따라서 FT-IR 스펙트럼 분석은 아프리카 안 시료들 사이에서 주요 대사체의 질적, 양적 변화를 조사하고 이들 화합물의 함량이 높은 계통이나 라인의 신속한 선별 수단으로 활용이 가능할 것으로 기대된다.



**Fig. 8** Representative FT-IR spectral from African yam tubers. FT-IR spectral ranges showed quantitative information of protein/amide I, II ( $1500\text{-}1700\text{cm}^{-1}$ ), phosphodiester group ( $1300\text{-}1500\text{cm}^{-1}$ ), and sugar compound ( $950\text{-}1100\text{cm}^{-1}$ ). Solid or dotted lines and abbreviations represent each African yam species. *D.ata* : *Dioscorea alata*; *D.dum* : *Dioscorea dumetorum*; *D.bul* : *Dioscorea bulbifera*; *D.spp* : *Dioscorea* spp. Arrows indicate the FT-IR regions showing significant spectral variations between African yam samples.

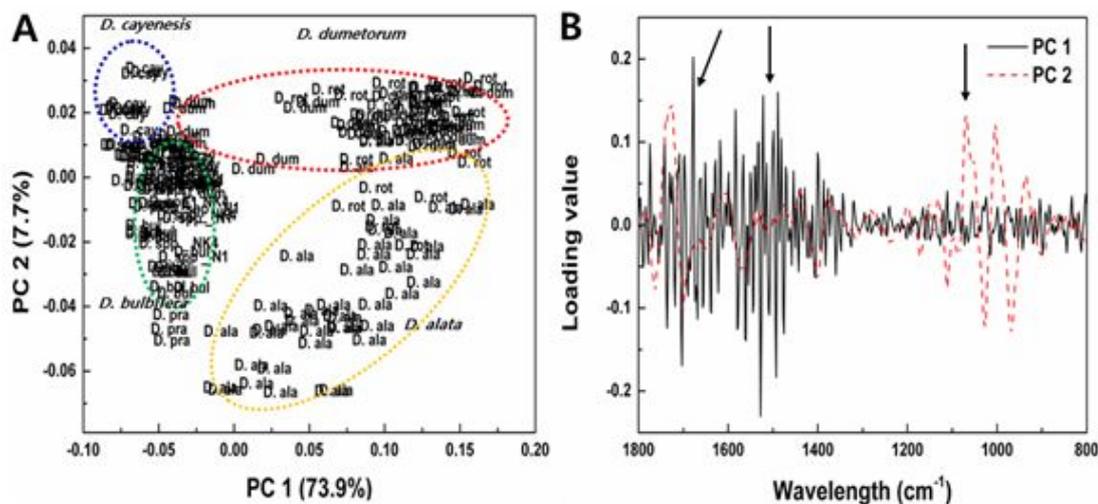
#### 4. 4. 3. 아프리카 안 시료의 FT-IR 스펙트럼 데이터 다변량통계분석

FT-IR 스펙트럼 데이터의 PCA 분석 결과 PC 1과 2 score는 전체 변이량의 각각 75.8%, 7.3%의 설명력을 갖고 있으며 이는 전체 변이량의 약 83.1%를 반영하고 있음을 알 수 있었다(Fig. 9A). 이처럼 설명력이 높은 PC1과 PC2를 기준으로 아프리카 안 시료들의 PCA score plot을 관찰한 결과, 아프리카 안 시료들은 PC 1을 기준으로 크게 좌우로 구분이 이루어지는 것을 관찰할 수 있었다. PCA score plot 상의 좌측에는 *D. cayenesis*, *D. bulbifera*, *D. mangenotiana*, *D. prahensilis*와 *Dioscorea* spp. Kenyan 종들이 주로 위치하였으며, 우측에는 *D. alata*, *D. dumetorum*과 *D. rotundata* 종들이 위치하였다. 이들 아프리카 안 시료들을 PC 2를 기준으로 보면 상하로 집단의 분포가 이루어짐을 알 수 있었다. 위쪽에는 *D. dumetorum*, *D. rotundata*에 속하는 시료들이 주로 분포하였으며, 반대로 아래쪽에는 주로 *D. alata* 시료들이 분포함을 알 수 있었다(Fig. 9A).

아프리카 안 시료의 PCA score plot상에서 공간적인 분포위치가 주는 생물학적인 정보를 파악하기 위하여 먼저 아프리카 안 시료들의 기능성 성분들 함량 차이 결과(Table 1.)와 PCA score plot상의 이들 시료들의 위치와 상관관계를 조사하였다. Total carotenoids, flavonoids이 비교적 낮았던 시료인 *D. alata*의 경우 재미있게도 PCA score plot상에서 우측 하단에 별도의 그룹을 형성하였으며, 반대로 total carotenoids, flavonoids 및 phenolic 함량이 가장 높았던 시료인 *D. dumetorum*의 경우 PCA score plot 상에서 우측 상단에 별도의 그룹을 형성하며 분포하는 것을 관찰할 수 있었다. 이상의 결과를 미루어볼 때 PCA score plot 상에서 공간적인 분포위치와 이들 시료의 기능성 성분들 total carotenoids, flavonoids 및 phenolic의 함량과 밀접한 상관관계가 있음을 알 수 있었다.

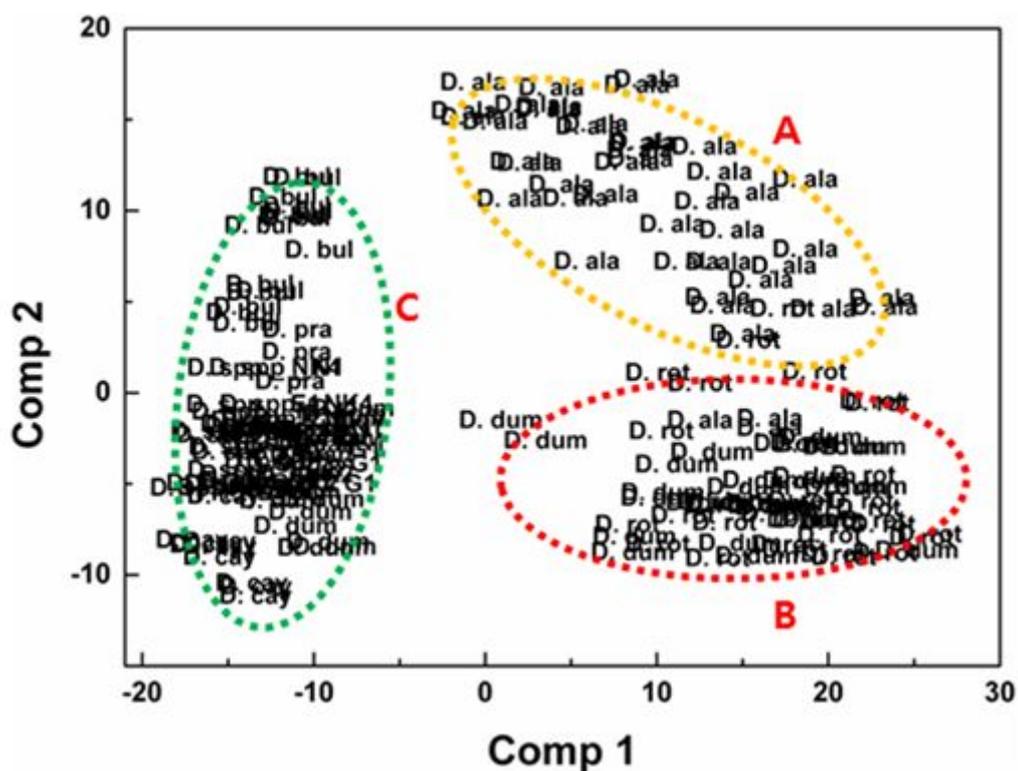
PCA score plot 상에서 기능성 성분들의 함량에 따른 아프리카 안 시료들의 집단 구분 및 클러스터링에 중요한 역할을 하는 FT-IR 스펙트럼 부위를 조사하기 위하여 PC 1과 2를 결정하는데 중요한 FT-IR 스펙트럼 부위를 조사하였다(Fig. 9B). Loading value 분석결과 아프리카 안 시료의 좌우 구분에 중요한 PC 1을

결정하는 중요한 FT-IR 스펙트럼 부위는 주로 1,700-1,500과 1,500-1,300 $\text{cm}^{-1}$  region임을 알 수 있었고, 양 시료의 상하 구분에 중요한 PC 2의 경우 주로 1,100-950 $\text{cm}^{-1}$  region임을 알 수 있었다(Fig. 9B). 이들 FT-IR 스펙트럼 부위는 아프리카 양 시료의 FT-IR 스펙트럼 비교에서 보였던 차이가 큰 부위(Fig. 8)와 일치하는 것으로 amide I, II, 지방산과 carbohydrates 계열의 화합물들의 질적, 양적 차이가 아프리카 양의 대사체 수준 식별에 중요한 역할을 하고 있음을 의미하는 결과라 사료된다(Fig. 9B). 특히 기능성 성분들의 함량 차이에 따른 아프리카 양 시료의 구분에 있어서 중요한 역할을 하는 PC 2의 경우 carbohydrates 계열의 화합물들의 질적, 양적 차이가 크게 영향을 줄 수 있었다. 이 결과로 미루어볼 때 이들 당 계열의 일차대사산물들의 양적 변화가 이차대사산물의 질적, 양적 변화와 양의 상관관계에 놓여 있음을 의미하는 결과라 사료된다.



**Fig. 9** PCA scores (A) and loading values plot (B) of PCA scores from FT-IR data of African yam lines. Dotted eclipses represent the clustering boundary of yam tuber samples with high (*D. dumetorium*), medium(*D. bulbifera*) and low(*D. alata*) contents of total phenolic compound and flavonoids. Abbreviations in the PCA score plots represent each African yam samples: Solid line and dotted line represent PC 1 and PC 2 score, respectively (A); Arrows indicate the FT-IR regions playing important roles in African yam sample clustering (B) *D. ata* : *Dioscorea alata*; *D. bul* : *Dioscorea bulbifera*; *D. cay* : *Dioscorea cayenensis*; *D. dum* : *Dioscorea dumetorium*; *D. man* : *Dioscorea mangelotiana*; *D. pra* : *Dioscorea prahensis*; *D. rot* : *Dioscorea rotundata*; *D. spp* : *Dioscorea* spp. Kenyan.

PLS-DA 분석의 경우 PCA 분석 보다 더욱 확실하게 아프리카 안 시료들의 그룹 식별이 가능하였다(Fig. 10). 각각의 아프리카 안 시료들의 그룹 경계는 PCA에 비해 더욱 작아졌으며 동일종에 속하는 각 시료들의 반복구들도 각 그룹 내에 위치하여 PCA보다 아프리카 안의 식별능력이 향상되었음을 알 수 있었다. Total carotenoids, flavonoids 함량이 비교적 낮았던 시료인 *D. alata*의 경우 PCA score plot상에서 우측 상단에 별도의 그룹을 형성하였으며, 반대로 total carotenoids, flavonoids 및 phenolic 함량이 가장 높았던 시료인 *D. dumetorum*의 경우 PLS-DA score plot 상에서 *D. rotundata*와 함께 우측 하단에 별도의 그룹을 형성하며 분포하는 것을 관찰할 수 있었다. 기능성 성분들의 함량이 중간인 그룹은 주로 좌측에 분포하였다(Fig. 10). 이상의 결과를 미루어볼 때 PCA와 마찬가지로 PLS-DA 역시 각 아프리카 안 시료들의 score plot 상에서 공간적인 분포위치와 이들 시료의 기능성 성분들 total carotenoids, flavonoids 및 phenolic의 함량과 밀접한 상관관계가 있음을 알 수 있었다.



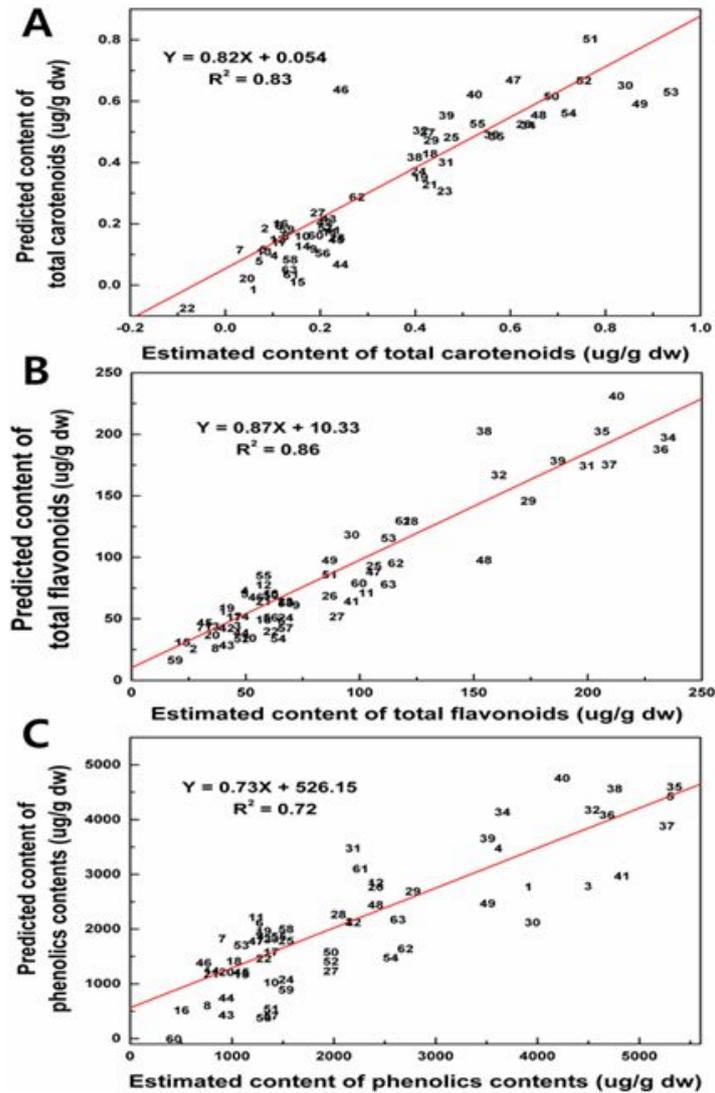
**Fig. 10** PLS-DA score plot of FT-IR data from African yam lines. Dotted eclipses and capitals represent the clustering boundary of yam tuber samples with high (B), medium (C) and low (A) contents of total phenolic compound and flavonoids. Abbreviations in the PLS-DA score plots represent each African yam samples. *D. ata* : *Dioscorea alata*; *D. bul* : *Dioscorea bulbifera*; *D. cay* : *Dioscorea cayenensis*; *D. dum* : *Dioscorea dumetorum*; *D. man* : *Dioscorea mangelotiana*; *D. pra* : *Dioscorea prahensilis*; *D. rot* : *Dioscorea rotundata*; *D. spp* : *Dioscorea* spp. Kenyan.

#### 4. 4. 4. 아프리카 양의 FT-IR 스펙트럼으로부터 total carotenoids, flavonoids 및 phenolic 함량 예측 PLS regression 모델링

FT-IR 스펙트럼 데이터와 UV-VIS spectrophotometer를 이용하여 측정된 total carotenoids, flavonoids 그리고 phenolic 함량 데이터를 이용하여 각 성분별 함량 예측 PLS 모델링을 개발하였다(Fig. 11). Total carotenoids의 경우, PLS 모델링을 통해 FT-IR 스펙트럼 데이터로부터 예측된 함량 값과 동일 시료의 실측 함량 값을 이용하여 회귀분석을 수행한 결과 상관계수가  $R^2 = 0.83$ 로 높게 나타났다(Fig. 11A). Total flavonoids와 total phenolic의 함량 예측 모델링을 통한 선형회귀분석 상관계수는 각각  $R^2 = 0.86$ (Fig. 11B) 및  $0.72$ (Fig. 11C)로 나타났다. 이들 결과는 양 시료의 FT-IR 스펙트럼 데이터만으로도 이들 시료 내에 존재하는 total carotenoids, flavonoids 및 phenolic의 함량을 80-90%정도 정확성으로 함량을 예측할 수 있는 결과라 사료된다. 아직 아프리카 양 시료에서 carotenoid 함량 예측 모델링 연구가 보고된 바 없어 정확한 비교는 어렵지만 carotenoid 함량이 높은 당근의 경우 FT-IR 스펙트럼과 total carotenoids 함량의 상관관계 분석을 통해 상관계수( $R^2$ )가 0.9정도임이 보고된 바 있다(Zude et al., 2008). Total flavonoids와 total phenolics의 함량 예측의 경우 커피(Páscoa et al., 2013)에서 보고된 바 있으며 이 경우 함량 예측 모델링의 상관계수는 각각  $R^2 = 0.94$ ,  $R^2 = 0.94$ 로 본 연구에서 보여준 아프리카 양의 경우보다 상관계수가 높게 나타났다. 향후 PLS 모델링의 정확도 향상을 위한 시료의 분석조건 표준화, 모델링에 사용되는 분석시료의 개체 수 증대, 기능성 성분의 양적 변화가 큰 시료가 모델링에 활용되는 등의 보완이 이루어진다면 보다 정확한 성분 예측 모델링으로 활용이 가능할 것으로 기대된다. 아울러 양의 경우 종간에 기능성 성분들의 양적, 질적 차이가 크게 이루어짐이 보고(Yang et al., 2009)된 바 있다. 또한 양이 함유하고 있는 기능성 성분들에 대한 약리적 효과가 연구되고 있으며(Song et al., 2010; Yoon et al., 1989), 최근에는 기능성 성분들에 대한 연구가 매우 활발히 진행되고 있다(Kim et al., 2009; Kum et al., 2006; Kwon et al., 2010). 그러나, 기능성 성분을 정량 분석하기 위해서는 많은 시간이 필요하고, 대량의 시료를 분석하는데 어려움이 있다. 따라서 본 연구 결과를 토대로 기능성

성분 예측 모델링을 구축하고, 활용한다면 보다 쉽고, 빠르고 그리고 비교적 정확하게 기능성 성분을 예측할 수 있기 때문에 시간적, 비용적인 측면에서 보다 손쉽게 활용이 가능할 것으로 사료된다.

따라서 본 연구에서 확립된 대사체 수준에서 아프리카 양의 유용 기능성 성분 함량 예측 모델링은 아직 동정 및 표준화가 이루어지지 않은 다수의 아프리카 양 시료로부터 품질특성 평가 수단으로 활용이 가능하여 우수한 양 품종이나 계통의 신속한 선발 수단으로서 활용이 가능할 것이다. 또한, 선발된 라인은 육종의 직접적인 소재로 활용이 가능하여 아프리카 양의 기능성 품종개발 가속화에 기여 할 수 있을 것으로 예상된다.



**Fig. 11** Linear regression analysis between estimated and predicted values of total carotenoids (A), flavonoids (B), and phenolic contents (C) by the PLS regression model from FT-IR spectral data. Regression coefficient values ( $R^2$ ) are 0.83, 0.86 and 0.72 respectively.

**5. FT-IR 스펙트럼과 HPLC 정량 분석을 이용한 감귤과실로부터  
carotenoid 성분 예측**

**(Prediction of Carotenoid Content from Citrus Fruit by Combined Fourier  
Transform Infrared Spectroscopy and Quantitative HPLC Analysis)**

## 5. 1. Abstract

This work examined the feasibility of prediction of carotenoid content without HPLC analysis using prediction modeling of carotenoid content by multivariate analysis combined with FT-IR and HPLC data. FT-IR spectra from peels and flesh of citrus (*Citrusunshiu* Marc. cv. Miyagawa) fruit were measured at monthly intervals. Quantitative analysis of carotenoids from the same citrus fruit was confirmed by quantitative HPLC analysis. FT-IR spectroscopy showed that remarkable increase in carbohydrate region ( $1,000-1,200\text{ cm}^{-1}$ ) and decrease in amide region ( $1,500-1,700\text{ cm}^{-1}$ ). HPLC also showed that increase in  $\beta$ -cryptoxanthin and decrease in lutein content during citrus fruit development. Reliable prediction of antheraxanthin ( $R^2=0.9117$ ),  $\beta$ -carotene ( $R^2=0.8816$ ),  $\beta$ -cryptoxanthin ( $R^2=0.8856$ ), and violaxanthin ( $R^2=0.7305$ ) was obtained from partial least square (PLS) regression modeling. Considering these results, FT-IR might be applied for metabolic evaluation of citrus fruit during ripening.

## 5. 2. Introduction

감귤은 세계에서 가장 경제적으로 중요한 과일 작물이고, 과일 중에서 복잡하고 많은 수의 carotenoid가 발견되었다(Gross 1987). 감귤은 일반적으로 유전적, 성숙 단계, 환경 조건 등 여러 가지 요인에 따라 carotenoid가 다르게 나타난다(Lee and Castle 2001; Kato et al. 2004; Fanciullino et al. 2006). 일반적인 carotenoid 성분 외에 감귤은 인간의 식단에 필요한 많은 유기 화합물을 포함하고, 건강에 가치 있는 대사물질을 함유하고 있다(Tadeo et al. 2008). Carotenoid는 식물의 chromoplasts와 다양한 광합성 생물 조류, 곰팡이, 박테리아 등에서 tetraterpenoid 유기 색소가 자연스럽게 엽록체에서 합성되어 생성된다(Moran et al. 2010). 600 개 이상의 알려진 carotenoid는 lutein과 zeaxanthin 같은 산소가 포함된 xanthophylls과  $\alpha$ -carotene,  $\beta$ -carotene, lycopene 같은 산소가 포함되지 않은 carotenoid로 분류 된다(Armstrong et al. 1996). 일반적으로 carotenoid는 식물에서 광합성의 빛에너지 흡수와 광손상으로부터 chlorophyll 보호 역할을 한다. carotenoid 색소는 과실의 특성을 결정하는데 중요하게 작용한다.1 게다가, carotenoid는 인간과 동물에 필수적인 비타민 A (mainly  $\beta$ -carotene,  $\alpha$ -carotene,  $\beta$ -cryptoxanthin)의 전구체와 항산화활성 억제와 같이 영양적인 가치가 있다(Olson 1989).

식물의 carotenoid 생합성 경로는 광범위하게 연구되고 있다. carotenoid 생합성의 첫 번째 개입 단계는 무색 phytoene 으로부터 geranylgeranyl pyrophosphate의 두 분자를 축합시킨다. Phytoene은 lycopene으로 변환된다.  $\alpha$ -carotene과  $\beta$ -carotene은 lycopene 화학 반응에서 중요한 역할을 한다.  $\alpha$ -Carotene은 순차적으로 hydroxylation되면서 lutein으로 변환된다.  $\beta$ -Carotene은 두 번째 단계의 hydroxylation에 의해  $\beta$ -cryptoxanthin을 거쳐 zeaxanthin으로 변환됩니다. Zeaxanthin은 antheraxanthin가 에폭시화를 거쳐 violaxanthin로 변환된다(Cunningham and Gantt 1998; Melendez-Martinez et al. 2003).

일반적으로 HPLC (high-performance liquid chromatography) 는 편리하고 신뢰할 수 있는 carotenoid 분석이다(Matsumoto et al. 2007). Atmospheric pressure chemical ionization를 사용한 LC-MS (liquid chromatography

-massspectrometry) 방법은 carotenoid 생합성 경로의 18개 carotenoid 기본 경로를 정량하기 위해 발했다(Kim et al. 2004). 그러나, carotenoid 생합성의 기본 경로에 감귤의 일부 carotenoid들은 장기적 분석, 여러 가지의 gradient elution, 검출기의 부족한 감도, 추출물에 들어있는 불순물의 간섭 등 여러 가지 문제가 있어 정량되지 않았다(Kato et al. 2004; Matsumoto et al. 2007). 이러한 약점들을 보완하기 위해 <sup>1</sup>HNMR, FT-IR 기술들이 대사체 분석에 사용되었다. 이러한 분광 기법은 일반적으로 이해하고 복잡한 대사체 데이터 세트로부터 생체 정보를 추출하기 위한 다변량통계분석이 결합된다. 이 기법의 큰 장점은 시료의 준비가 간단하고 빠르며 감도가 낮더라도 chromatographic 기법과 비슷하게 나타난다. 이러한 분광 기술은 식물 종과 품종의 신속한 식별(Kim et al. 2004), GM 식물의 대사체 평가(Roessner et al. 2001), 상업식품저장(Son et al. 2009), 약초(Lee et al. 2009) 등 다양한 식물 연구 분야에 적용되고 있다. 더욱이, 이러한 분광 기법은 스트레스 반응과 유전자 기능의 결정 같은 여러 가지 생리적으로 이해하는데 적용될 수 있다(Fiehn et al. 2000).

이와 같이, 스펙트럼 기법과 HPLC 기법을 결합하여 사용하면 과실의 성숙과 익어가는 과정에서 변화되는 과실에 대사체를 쉽게 관찰할 수 있고, 과실의 성숙 단계를 빠르게 식별할 수 있다. 본 연구는 감귤의 녹색에서 주황색으로 껍질 색 변화와 이 과정에서 생길 수 있는 carotenoid 함량 관계를 분석하여 신속한 대사체 식별체계 확립을 위해 수행하였다. 본 연구의 결과는 감귤 육종과 carotenoid 가 향상된 감귤 생산에 신속한 선발체계를 확립하고자 하였다.

## 5. 3. Material and Method

### 5. 3. 1. 식물재료

궁천조생(*Citrusunshiu* Marc. cv. Miyagawa)과 방사선 돌연변이 궁천조생은 2010년 6월부터 12월까지 매월 수확하여 사용하였다. 과실은 증류수(D.W)를 이용하여 표면에 묻어있는 불순물을 세척한 후, 실온에서 건조하였고, 과실의 과육과 과피를 나눠 실험에 사용하였다. 액체질소를 사용하여 빠르게 동결건조시킨 후, 막자와 막자사발을 이용하여 미세한 분말로 분쇄하였으며 -70°C 초저온냉동고에 보관하면서 실험에 사용하였다.

### 5. 3. 2. 전세포추출물 제조

FT-IR 스펙트럼 분석을 위해 전세포추출물을 준비하였다. 준비된 감귤 분말 시료 20 mg을 1.5 ml Eppendorf tube에 넣고 20%(v/v) methanol 용액을 200  $\mu$ L씩 첨가하여 잘 섞어주었다. 각 tube를 50°C water bath에서 20분간 반응시킨 다음 13,000rpm에서 15분간 원심분리한 후 상층액을 새로운 tube에 옮겼다. 새로운 tube에 옮긴 시료는 다시 한 번 13,000rpm에서 5분간 원심분리하고 찌꺼기가 들어가지 않도록 상층액을 조심스럽게 새로운 tube에 옮겼다. 추출된 추출물은 -20°C에서 보관하여 FT-IR 스펙트럼 분석에 사용하였다.

### 5. 3. 3. FT-IR 스펙트럼과 데이터 전처리

FT-IR(Fourier transform infrared) 스펙트럼 조사는 Tensor 27(Bruker Optics GmbH, Ettlingen, Germany)를 사용하였고, DTGS(deuterated triglycine sulfate) 검출기로 분석하였다. 추출된 genomic DNA 각각의 시료 5 $\mu$ l를 384-well ZnSe plate에 분주하여, 37°C hot plate에서 약 20분간 건조하였다. 건조된 ZnSe plate는 Tensor 27에 장착된 HTS-XT(Bruker Optics GmbH) 고효율 자동화 장치를 이용하여 스펙트럼을 조사하였다. FT-IR 스펙트럼은 OPUS software

(version 6.5)을 통해 획득하였다. 각 시료의 스펙트럼은 총  $4,000 \sim 400\text{cm}^{-1}$  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. IR 스펙트럼은 plate 스펙트럼 (background)을 뺀 샘플의 스펙트럼을 나타냈다. 다변량 분석을 위해 기본 FT-IR 스펙트럼의 baseline correction, normalization, smoothing 그리고 2차 미분을 OPUS software (ver. 7.0)를 사용하여 전 처리하였다. 전 처리된 스펙트럼 데이터는 다변량통계 분석에 사용하였다.

### 5. 3. 4. carotenoid 추출과 HPLC 분석

HPLC를 이용한 carotenoid의 추출, 분리 및 측정은 다음과 같이 수행 하였다(Kim et al. 2010). Carotenoid 추출은 감귤 시료 0.1g에 0.1% ascorbic acid (w/v)가 첨가되어 있는 ethanol 3mL을 넣고, 20초간 vortexing한 후, 85 °C water bath에서 5분간 추출하였다. 5분간 추출 후, potassium hydroxide (120  $\mu$ L, 80% w/v)을 첨가하고 85 °C water bath에서 10분간 감화하여 carotenoid를 추출하였다. 가수분해 후, 샘플을 얼음에 즉시 놓고, 차가운 deionised water (1.5 mL)을 첨가 하였다. Internal standard로  $\beta$ -Apo-8-carotenal (0.05 mL, 25  $\mu$ g/mL)를 첨가하였다. carotenoid는 1200  $\times$ g에서 원심 분리하여 층 분리하였고, hexane (1.5 mL)로 2 회 추출 하였다. 추출물의 정제는 질소 기류하여 건조하고, HPLC 분석 전에 50:50 (v/v) dichloromethane/methanol에 다시 녹여 사용하였다. HPLC에 C<sub>30</sub> YMC column(250  $\times$  4.6 mm, 3  $\mu$ m; Waters Corporation, Milford, MA, USA)을 이용하여 carotenoid를 분리하였다. 크로마토그램은 450 nm로 하였다. Solvent A는 10 mM ammonium acetate가 첨가된 methanol/water (92:8 v/v)를 사용하였다. Solvent B는 100% methyl *tert*-butyl ether을 사용하였다. Gradient elution은 1mL/min 이하 조건에서 수행하였다; 0min, 83% A/17% B; 23min, 70% A/30% B; 29min, 59% A/41% B; 35min, 30% A/70% B; 40min, 30% A/70% B; 44min, 83% A/17% B; 55min, 83% A/17% B. 정량 분석을 위해,  $\beta$ -Apo-8-carotenal의 피크 면적 비율을 사용하여 기준 carotenoid 검정 곡선을

나타내 사용하였다.

### 5. 3. 5. 다변량통계분석

다변량통계 분석은 전체 스펙트럼에서 1,800-800cm<sup>-1</sup> 영역의 FT-IR 스펙트럼 데이터를 이용하여 principal component analysis (PCA)와 partial least squares discriminant analysis (PLS-DA)를 실시하였다. 2차 미분된 FT-IR 스펙트럼 데이터는 R 통계 분석 프로그램 (버전 2.15.0, Auckland, New Zealand)을 통해 분석되었다. PCA는 non-linear iterative partial least squares (NIPALS) 알고리즘에 따라 수행하였다(Wold 1966). 고유값 (eigenvalues) > 1.0 으로 주성분(Principal components)을 추출하였다. PCA 분석은 PCA scores를 추출하여 상관 행렬을 계산하였다. 월간 대사 변화에 대한 중요한 FT-IR 변수를 찾기 위해 PCA loading을 조사 하였다. PLS-DA는 모든 데이터를 사용하는 PCA는 달리 대사체 프로파일에 숨겨져 있는 정보에 대한 차이를 확인할 수 있는 분석방법이다(Fiehn et al. 2000; Trygg et al. 2007). PLS-DA는 PCA 분석 후 감귤류에 월별 대사체 변화를 R program 사용하여 clustering 실시하였다.

감귤류의 carotenoid 함량 예측 모델을 확립하였다. PLS regression을 실시하기 위해서 2개의 변수 X와 Y에 대한 정보를 연결시켰다; X 변수에는 FT-IR 스펙트럼 데이터를 적용하였고, Y 변수에는 HPLC 정량 분석 데이터를 적용하였다. 데이터 세트는 2개의 파트로 나뉘었다; 예측 모델의 예측 능력을 테스트하기 위해 트레이닝 세트를 만들어 모델 구축에 사용하였고, 트레이닝 세트는 테스트 세트에 사용되지 않았다. X 변수와 Y 변수간의 측정 모델을 만들기 위하여 상관관계를 예측한 후 회귀계수를 사용하여 상관관계를 계산하였다. 모든 분석은 통계 분석을 위해 3반복 수행하였다. 실험 데이터는 ANOVA (analysis of variance) 분석을 실시하였고, 유의성 판별은 P<0.01 수준에서 t-test를 실시하였다.

## 5. 4. Results and Discussion

### 5. 4. 1. FT-IR 스펙트럼 기반 다변량통계분석을 이용한 감귤류의 신속한 대사체 식별

감귤의 간편하고 신속한 대사체 식별 시스템은 FT-IR 스펙트럼과 다변량통계 분석을 이용하여 확립하였다. 감귤 과피로부터 녹색에서 오렌지색으로 변화하는 FT-IR 스펙트럼 데이터 나타냈다(Fig. 12A). FT-IR 스펙트럼을 보면 polysaccharide ( $900-1,200\text{cm}^{-1}$ )와 protein/amide I, II ( $1,500-1,700\text{cm}^{-1}$ )가 감귤의 성숙 과정에서 많은 변화가 나타냈다.  $1,750\text{cm}^{-1}$ 에서 나타나는 ester bond의 C=O stretching band가 8월 시료에서만 나타났다. 스펙트럼 상의 phospholipid/DNA/RNA(Parker 1983)에 대응하는 다른 부위 ( $1,200-1,500\text{cm}^{-1}$ )에서 월간 시료들이 유사한 패턴을 보였다(Fig. 12A). 감귤의 과실은 성숙과정에서 세포 분열과 세포 확대 과정에서 carbohydrate를 활성화하여 수용성 당류를 축적하는 것을 주로 구성한다(Mehouachi et al. 1995). 감귤 과실의 발달과 성숙하는 동안 발생하는 가장 중요한 대사체 변화는 수분 축적, carbohydrate 축적, 색소의 변화 (carotenoid 축적과 chlorophyll 감소), organic acid 축적 (citrate과 ascorbic acid), amino acid 감소 (glutamate and GABA)에 관련이 있다 (Fanciullino et al. 2006, Cercós et al. 2006). 돌연변이 미성숙 과실의 경우, lycopene 축적, 낮은 citric acid, 높은 sucrose의 대사체 변화가 있다고 오렌지를 통해 보고된바 있다(Liu et al. 2007).

궁천조생 과실로부터 FT-IR 스펙트럼의 7월부터 8월까지 중요한 월별 변화를 관찰하였다. 감귤 과피로부터 FT-IR 데이터에서 과실 성숙 과정 중 polysaccharide ( $900-1,200\text{cm}^{-1}$ ) 부위가 증가하고 protein/amide I, II ( $1,500-1,700\text{cm}^{-1}$ ) 부위가 감소하는 모습을 볼 수 있었다. 방사선 돌연변이 역시 궁천조생 과실과 비슷하게  $1,500-1,700\text{cm}^{-1}$  부위에서 빠르게 감소하는 모습을 FT-IR 스펙트럼 변화로 볼 수 있었다. 세포벽에서 cellulose, hemicellulose, pectic polysaccharides의 당류와 수용성 당 성분들이 망고 과실에서는 증가하였다 (Duarte et al. 2002). 이런 당 화합물들은 IR 스펙트럼의  $900-1200\text{cm}^{-1}$  부위에서

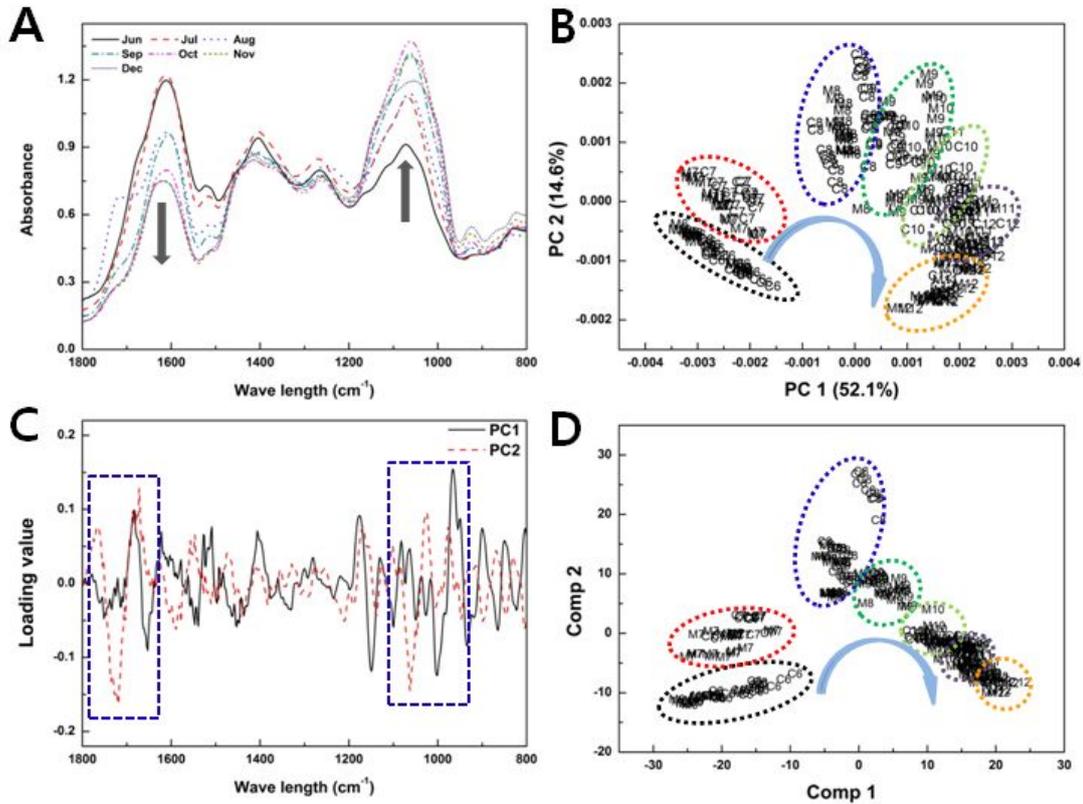
특징적 밴드로 복합적으로 나타난다(Duarte et al. 2002; Lopez-Sanchez et al. 2010). 세포벽에 holocellulose는  $900-1200\text{cm}^{-1}$  부위에서 전형적인 밴드를 나타낸다(Pandey 1999). 감귤 과피에서 이러한 FT-IR 스펙트럼의 특성을 고려하면, polysaccharide ( $900-1,200\text{cm}^{-1}$ ) 밴드의 증가는 감귤 과실의 발달과 성숙 과정에서 세포 확대 단계는 carbohydrate 축적과 관련이 있을 수 있다. 또한, protein/amide I, II ( $1,500-1,700\text{cm}^{-1}$ ) 밴드의 감소는 감귤 과실의 발달과 성숙 과정에서 organic acid와 amino acid의 감소와 관련이 있을 수 있다. 이러한 결과는 FT-IR 스펙트럼 변화가 감귤 과피에 질적, 양적 대사체의 변화와 밀접한 관계가 있음을 나타낸다.

감귤 과피로부터 FT-IR 데이터의 PCA는 PC 1, 2 score를 사용하여 나타냈다(Fig. 12B). PC 1, 2의 설명력은 52.1%와 14.6%이고, 전체변이량은 65.7%로 나타났다. 월별 과실 시료는 분리 집단에서 각각의 월별 과실들이 PCA 상에서 같은 수확 시기별로 그룹이 되어 나타났다. PCA 결과, 발달과 성숙 과정에서 감귤 껍질의 전체 대사체 변화가 별개의 대사 흐름이 있다는 것을 나타냈다. 궁천조생과 방사선 돌연변이 사이의 대사체 변화가 적어 동일하게 월별 변화를 나타냈다. 따라서, PCA 결과처럼 FT-IR 스펙트럼이 감귤 과실의 발달과 숙성과정에서 대사체 식별이 활용 가능할 것으로 판단된다.

샘플 분리를 위해 PC 1과 PC 2 score 사이에 가장 큰 차이를 나타내는 스펙트럼 변수를 확인하기 위해 PCA loading values 확인하였다. 감귤 과피로부터 PCA 데이터를 기반으로 PC score loading plot을 관찰하였다(Fig. 12C). 각각의 polysaccharide ( $900-1,200\text{cm}^{-1}$ )와 protein/amide I, II ( $1,500-1,700\text{cm}^{-1}$ ) 부위가 PC 1과 PC 2의 중요 FT-IR 스펙트럼 변화를 결정하는데 작용하였다. 이러한 결과는 polysaccharide와 protein/amide I, II 부위에 질적, 양적 대사체 변화가 과실 발달하는 동안 중요한 대사체 변화로 나타났다. 이러한 결과를 고려하면, 중요 FT-IR 스펙트럼에 변화가 감귤 과실의 발달과 성숙 과정에서 우수한 집단 구분으로 사용될 것이다.

PLS-DA 또한 PCA 이상의 집단 구분 패턴을 나타냈다(Fig. 12D). 같은 수확 시기에 속하는 월별 과일 샘플이 PCA보다 더 밀접하게 집단 구분이 이뤄졌고, PLS-DA 또한 감귤 과실 발달과 성숙 과정에서 변화되는 대사체를 식별할 수 있

는 것으로 나타났다. 따라서, PCA와 PLS-DA 결과 감귤 과실의 발달과 성숙 과정을 FT-IR 스펙트럼 분석을 통해 대사체 변화와 월별 집단 구분이 가능하였다.

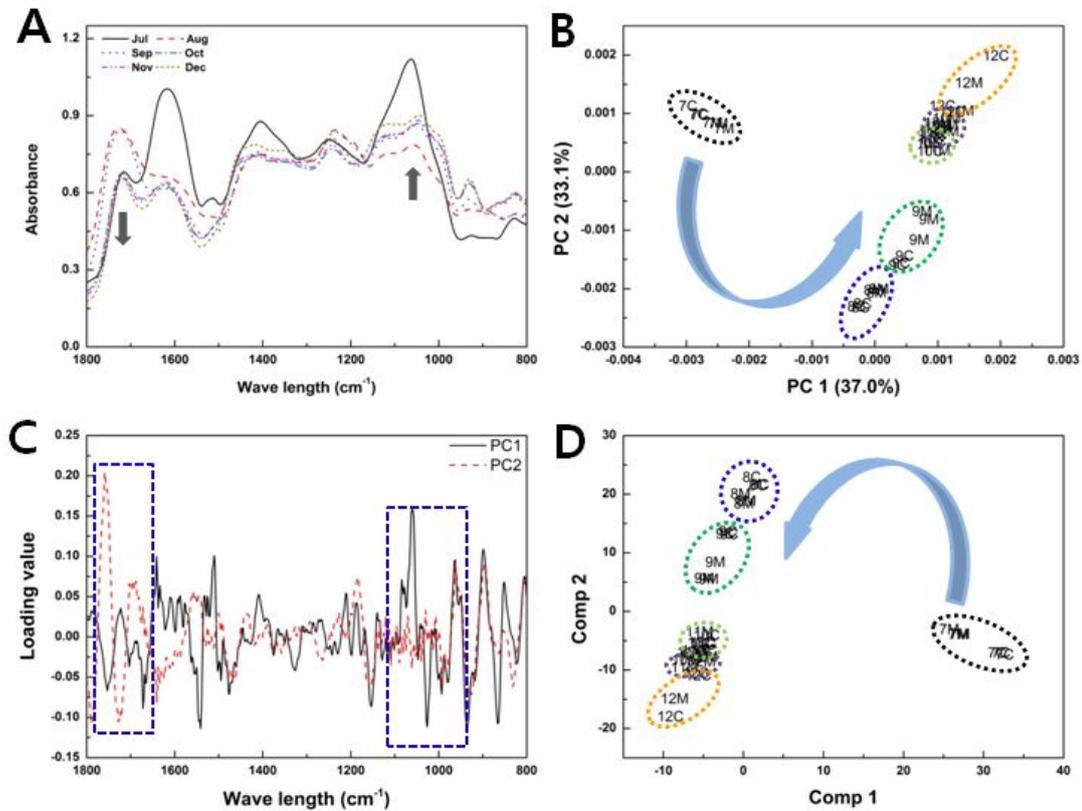


**Fig. 12** Multivariate analysis of FT-IR spectroscopic data from peels of *C. unshiu*. A: Representative FT-IR spectra of from June to December. Dotted lines represent monthly samples and arrows represent FT-IR spectroscopic range, which shows significant variation between samples. B: Two-dimensional PCA score plot of FT-IR spectroscopic data from wild-type and mutant fruits. Arrows represent the direction of whole metabolic flow from monthly samples by PCA. C: PC loading plot based on PCA data of wild-type and mutant fruit. Dotted rectangles represent significant FT-IR spectroscopic ranges for determining PC 1 and PC 2 of the PCA score plot (B). D: PLS-DA score plot based on PCA data of wild-type and mutant fruit. Dotted ellipses represent a group and shows monthly variation in wild-type and mutant fruits (B, D). Arrows represent the direction of whole metabolic flow from monthly samples by PLS-DA.

궁천조생 과육으로부터 FT-IR 스펙트럼 데이터를 이용한 다변량통계 분석을 실시하였다(Fig. 13). 7월부터 12월까지의 궁천조생 과육으로부터 중요한 FT-IR 스펙트럼을 나타냈다(Fig. 13A). 과피에서와 같은 결과로, 과육으로부터 샘플 사이에 중요한 FT-IR 스펙트럼의 변화는 polysaccharide( $900-1,200\text{cm}^{-1}$ )와 protein/amide I, II ( $1,500-1,700\text{cm}^{-1}$ ) 부위로 나타났다. 과실 발달 과정 중에 polysaccharide ( $900-1,200\text{cm}^{-1}$ ) 부위가 증가하고 protein/amide I, II ( $1,500-1,700\text{cm}^{-1}$ ) 부위가 감소하는 모습을 볼 수 있었다.  $1,750\text{cm}^{-1}$ 에서 나타나는 ester bond의 C=O stretching band가 과피에서는 스펙트럼 변화가 낮았지만, 7월에서 10월 사이에는 감귤 과육에서는 많은 변화가 나타났다. 올리브오일에서는  $1,746\text{cm}^{-1}$ 에서 나타나는 ester bond의 C=O stretching band가 특이적으로 스펙트럼에 나타났다(Lopez-Sanchez et al. 2010). 올리브 과실은 7월과 10월 사이에 특이적인 오일 밴드가 FT-IR 스펙트럼에서 급격히 증가하는 것을 볼 수 있다. 이러한 결과는 감귤 과육으로부터  $1,750\text{cm}^{-1}$ 의 FT-IR 스펙트럼 변화가 크게 나타나 오일과 지질화합물의 질적, 양적 변화에 관련이 있음을 나타냈다. 감귤 과피(Fig. 12A)와 과육(Fig. 13A)의 FT-IR 스펙트럼 결과를 보면 과실의 발달과 성숙시 세포 확장 과정에서 carbohydrate 축적되는 공통적인 대사체 변화가 나타나는 것을 알 수 있다. 그러나, proteins/amides 감소, organic acid와 amino acid의 축소, 오일 화합물의 변화에 대사체 변화는 두 과실이 다른 관계를 나타냈다.

감귤 과육으로부터 FT-IR 데이터의 PCA는 PC 1, 2 score를 사용하여 나타냈다(Fig. 13B). PC 1, 2의 설명력은 37.0%와 33.1%이고, 전체변이량은 70.1%로 나타났다. 과피에서와 비슷한 결과로 과육 시료 역시 각각의 월별 과실들이 같은 수확 시기별로 그룹이 되어 나타났다. PCA 결과, 궁천조생과 방사선 돌연변이 사이의 대사체 변화가 적어 동일하게 월별 변화를 나타냈고, 감귤 과실의 발달과 성숙 과정에서 대사체 식별이 FT-IR 스펙트럼으로 가능한 것으로 나타났다. 감귤 과육으로부터 PCA 데이터를 기반으로 PC score loading plot을 관찰하였다(Fig. 2C). 과피에서와 비슷한 결과로 각각의 polysaccharide ( $900-1,200\text{cm}^{-1}$ ) 부위와 protein/amide I, II ( $1,500-1,700\text{cm}^{-1}$ ) 부위가 PC 1과 PC 2의 중요 FT-IR 스펙트럼 변화를 결정하는데 작용하였다. 감귤 과육으로부터 PLS-DA 또한 PCA 이상의

집단 구분 패턴을 나타냈다(Fig. 13D). 따라서, 과피(Fig. 12)와 과육(Fig. 13)의 PCA와 PLS-DA 결과 감귤 과실의 발달과 성숙 과정을 FT-IR 스펙트럼 분석을 통해 대사체 변화와 월별 집단 구분이 가능하였다.



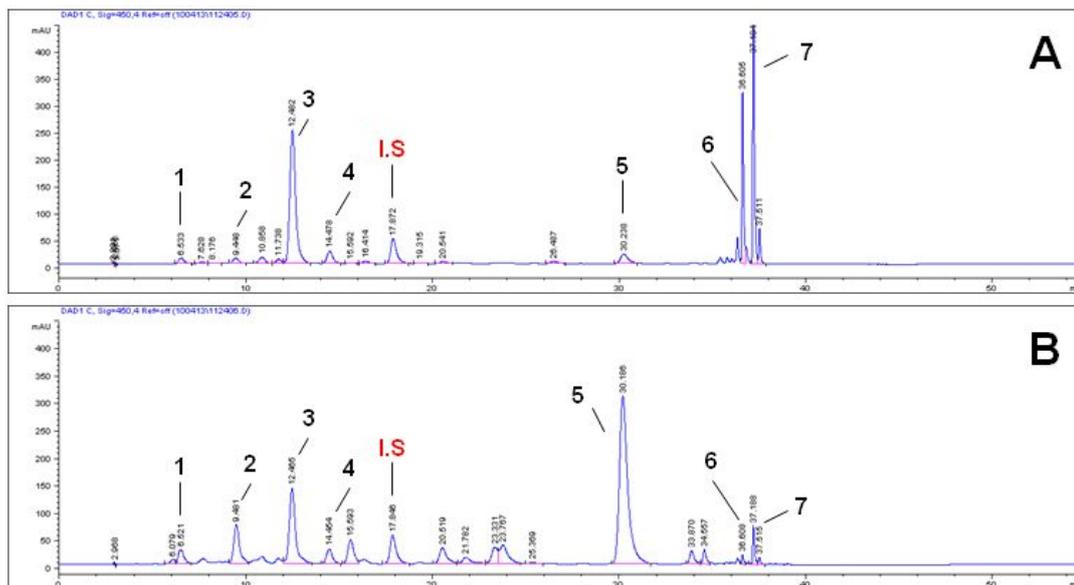
**Fig. 13** Multivariate analysis of FT-IR spectroscopic data from flesh of *C. unshiu*. (A) Representative FT-IR spectra of from June to December. Dotted lines represent monthly samples and arrows represent FT-IR spectroscopic range showed significant variation between samples. (B) Two-dimensional PCA score plot of FT-IR spectroscopic data from wild-type and a radiation mutant of *C. unshiu*. Arrows represents the direction of whole metabolic flows from monthly samples by PCA. (C) PCs loading plot based on PCA data of wild-type and a radiation mutant line of *C. unshiu*. Dotted rectangles represent significant FT-IR spectroscopic ranges for determining PC 1 and PC 2 of PCA score plot (B). (D) PLS-DA score plot based on PCA data of wild-type and a radiation mutant line of *C. unshiu*. Dotted ellipses represent a group showed monthly variation from wild-type and a radiation mutant line of *C. unshiu* (B, D). Arrows represents the direction of whole metabolic flows from monthly samples by PLS-DA.

#### 5. 4. 2. 감귤 과실 발달 과정에 따른 월별 carotenoid 함량의 변화

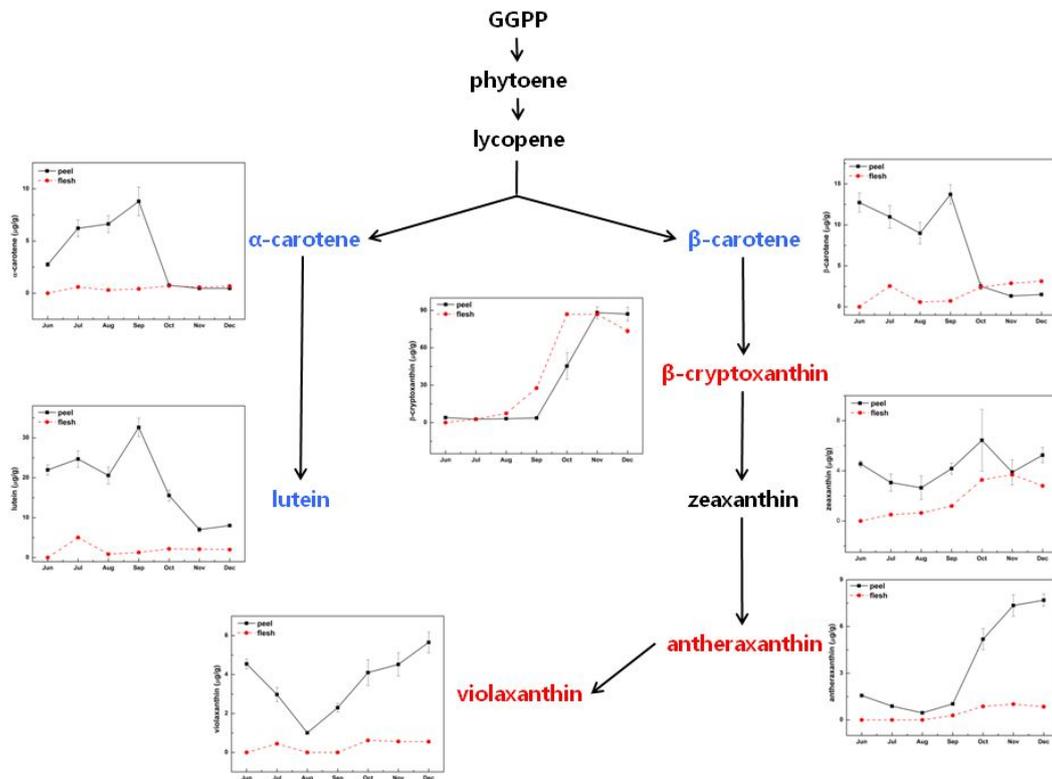
궁천조생과 방사선 돌연변이의 과피와 과육에서 carotenoid 함량을 조사하였다. 감귤 과실이 발달하는 동안, chlorophyll이 저하되면서 carotenoid의 많은 축적이 나타났다(Matsumoto et al. 2007). 궁천조생의 9월(A)과 10월(B) 과피 시료에서 추출한 대표적인 추출물을 나타냈다(Fig. 14). 감귤 과실의 대표적 carotenoid 화합물인 Violaxanthin, antheraxanthin, lutein, zeaxanthin,  $\beta$ -cryptoxanthin,  $\alpha$ -carotene,  $\beta$ -carotene를 동정하였다. 과피와 같이, 과육에서도 대표적 carotenoid 화합물을 추출하였다. 그러나 각각의 carotenoid 화합물에 양적 변화는 조직의 유형(과피와 과육)과 과실 발달하는 과정에서 수확시기에 따라 큰 차이가 나타났다.

감귤 과실의 carotenoid 생합성 추정 경로와 궁천조생의 과피에서 carotenoid 함량의 월별 변화를 나타냈다(Fig. 15). 감귤 과실로부터 FT-IR 스펙트럼의 중요한 변화를 7월과 8월 사이에서 관찰하였다(Fig. 12, 13). 하지만, carotenoid 화합물의 중요한 질적, 양적 변화를 과피에서는 9월과 10월에, 과육에서는 7월과 8월에 관찰할 수 있었다. 궁천조생의 과피로부터 7개의 carotenoid 화합물들은 과실 발달과정에서 양적 패턴이 3개의 과정으로 분류되었다. 첫 번째 과정은  $\alpha$ -carotene,  $\beta$ -carotene, lutein이 포함된 carotenoid 생합성 경로의 초기 전구체로 구성되어있고, 이 성분들은 과실의 성숙과정을 통해 감소되었다. 두 번째 과정은 zeaxanthin과 같은 carotenoid 생합성의 중간 전구체로 구성되어있고, 이 성분들은 과실의 성숙과정 동안 큰 변화가 없었다. 세 번째 과정은  $\beta$ -cryptoxanthin, violaxanthin, antheraxanthin와 같은 carotenoid 생합성의 최종 전구체로 구성되고, 이 성분들은 과실의 성숙과정 동안 증가한다(Fig. 15). 과육에서 lutein을 제외한 대부분의 carotenoid 성분 함량이 과실에 발달 과정에서 증가했다. 하지만, 과육에 있는 carotenoid 함량의 매달 변화하는 양적 정도가 과피에 비해 낮게 나타났다(Fig. 15). 과피에서 가장 큰 양적 변화를 보여준 carotenoid 성분은 lutein과  $\beta$ -cryptoxanthin으로 나타났다. Lutein 성분은 7월(21.9  $\mu\text{g/g}$  dry weight [dw] of peels)에서 11월(7.0  $\mu\text{g/g}$  dw of peels)로 가면서 급격히 감소하였고,  $\beta$ -cryptoxanthin 성분은 6월(4.0  $\mu\text{g/g}$  dw of peels)에서 11월

(88.2  $\mu\text{g/g dw of peels}$ )로 가면서 크게 증가하였다. Lutein과  $\beta$ -cryptoxanthin 성분의 비슷한 양적 변화를 과육에서 확인할 수 있었다. 과피의  $\alpha$ - 와  $\beta$ -carotene 성분은 9월과 10월 사이에 급격히 감소하였다. 하지만, 과육의  $\alpha$ - 와  $\beta$ -carotene 성분은 비슷한 시기에 변화가 없거나 약간 증가하는 것을 알 수 있었다. 이러한 결과를 보면 과피와 과육의 carotenoid 성분에 질적, 양적 변화가 서로 다르다는 것을 알 수 있었다.

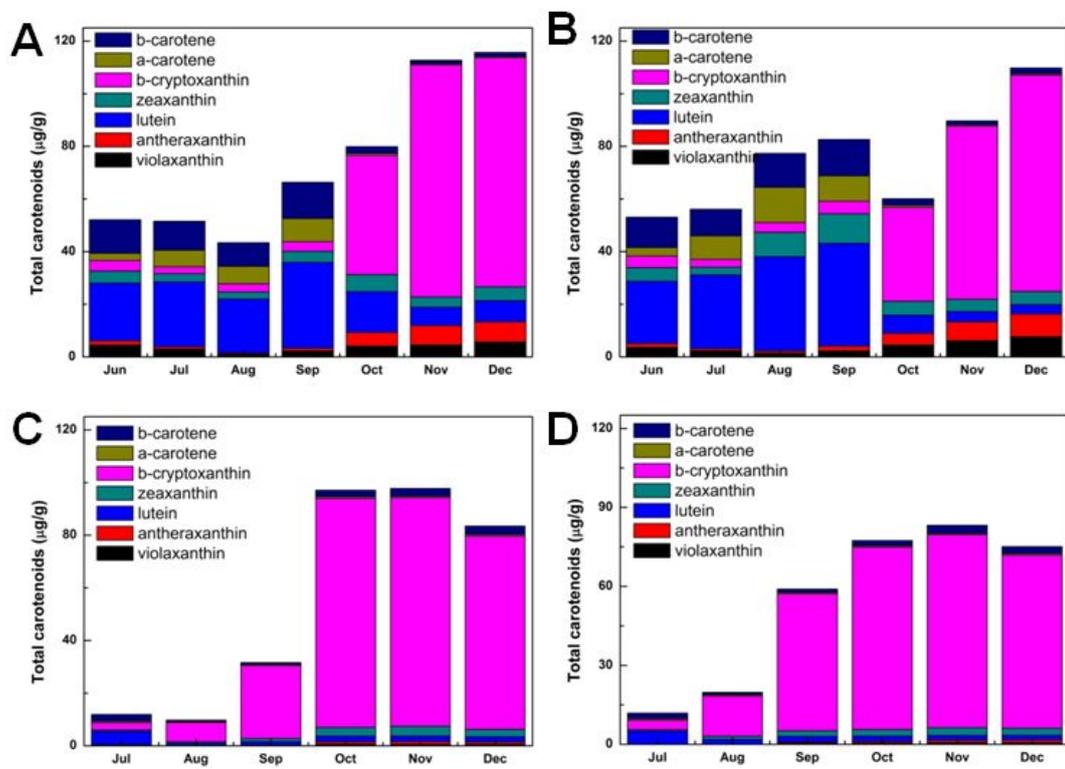


**Fig. 14** Carotenoid biosynthetic pathway and monthly carotenoid change from peels of wild-type *C. unshiu* fruits. Peels were collected in (A) September and (B) October. Numbers represent identified carotenoid compounds and I.S. represents the internal standard for qualitative and quantitative analysis of carotenoids. 1, violaxanthin; 2, antheraxanthin; 3, lutein; 4, zeaxanthin; 5,  $\beta$ -cryptoxanthin; 6,  $\alpha$ -carotene; 7,  $\beta$ -carotene.



**Fig. 15** Carotenoid biosynthetic pathway and monthly carotenoid change from peels of wild-type *C. unshiu* fruits. Arrows represent carotenoid biosynthetic pathway in *C. unshiu* fruits. Each graph showed monthly variation of each carotenoid compound. Columns and bars represent the means and SE, respectively. Means and standard deviations ( $n = 3$ ) were calculated from triplicate HPLC analysis.

궁천조생과 방사선 돌연변이의 carotenoid 성분 함량을 비교하여 나타냈다 (Fig. 16). 과피(Fig. 16A, B)와 과육(Fig. 16C, D)의 carotenoid 성분에 양적, 질적 변화를 확인한 결과 궁천조생과 방사선 돌연변이 사이의 차이가 크게 나타나지 않았다. 과일은 일반적으로 유전적, 성숙 단계, 환경 조건 등 여러 가지 조건에 따라 복잡한 carotenoid profiles이 나타난다. Kato et al. 보고에 따르면 carotenoid 축적은 감귤 과실의 성숙 과정에서 유전자 발현에 의해 많은 carotenoid 생합성이 이뤄지는 것으로 보고되었다(Kato et al. 2004). 여러 연구에 따르면 carotenoid 성분이 유전적으로 연결되어 있어 carotenoid 구조에 따라 분종 구분이 가능한 것으로 알려져 있다. Goodner et al. 보고에 따르면 성숙한 과일의 주스에서  $\beta$ -cryptoxanthin 함량의 차이가 오렌지, 귤, 잡종 사이에 다르게 나타나 구분이 가능하다고 보고되었다(Goodner et al. 2001). Lee and Castle 보고에서도 violaxanthin이 오렌지 품종의 주요 색소이지만, 이 성분은 성숙 기간 동안 다른 종류로 변경된다고 보고하였다(Lee and Castle 2001). 최근에, Fanciullino et al. 은 성숙한 과일 주스의  $\beta$ -cryptoxanthin와 violaxanthin 성분에 따라 세 개의 집단에서 여덟 종에 속하는 25개의 감귤 품종을 분류하였다 (Fanciullino et al. 2006). 서로 다른 carotenoid 축적에 따라 감귤 품종 분류에 대한 포괄적인 연구가 최근 Matsumoto et al. 에 의해 수행되었습니다. 이 연구들에 따르면 violaxanthin와  $\beta$ -cryptoxanthin는 중요 carotenoid이며, 오렌지 품종의 특성에 중요한 역할을 갖는 것으로 나타났다(Matsumoto et al. 2007). 최근 Lopez-Sanchez et al. 보고에서는 Satsuma mandarin와 Valencia orange에서 과피의 색이 녹색에서 주황색으로 변할 때 carotenoid의 질적 패턴 변화가 일어난다고 보고하였다(Lopez-Sanchez et al. 2010). 이러한 결과들을 보면, 감귤 과실이 숙성 동안 계절적 carotenoid 차이를 통해 감귤을 분류하는데 사용될 수 있을 것으로 판단된다(Lopez-Sanchez et al. 2010). 본 연구에서는,  $\beta$ -cryptoxanthin이 궁천조생과 방사선 돌연변이 과실 발달 과정에서 중요한 carotenoid임을 확인하였다. 과실의 발달과 성숙과정에서 carotenoid 성분들이 서서히 증가하였다. 본 연구에 carotenoid 분석 결과는 이전 보고들과 약간에 차이는 있지만 매우 유사하였다(Lee et al. 2001; Fanciullino et al. 2006; Matsumoto et al. 2007; Goodner et al. 2001).



**Fig. 16** Monthly variation of carotenoid content from citrus. Carotenoids contents from peels of wild-type fruits (A) and a radiation mutant of *C. unshiu* (B), and flesh of wild-type fruits (C) and a radiation mutant of *C. unshiu* (D). Symbols represent each carotenoid compound.

#### 5. 4. 2. 감귤 과실의 carotenoid 함량 예측 PLS regression 모델링

FT-IR 스펙트럼과 HPLC 분석을 이용하여 감귤 과피의 carotenoid 함량 PLS regression 예측 모델링을 확립하였다. IR 스펙트럼은 빠르고 비파괴적인 분석 방법으로 망고 주스의 당을 정량 분석(Duarte et al. 2002), 토마토 과실의 lycopene, beta-carotene 성분 분석에 응용되었다(Baranska et al. 2007). PLS regression은 FT-IR 스펙트럼과 HPLC 분석 사이에 상관관계 분석하기 위하여 사용하였다. 감귤의 과피와 과육에 함유되어 있는 carotenoid 성분의 월별 변화를 확인하였다(Fig. 15). 시료 안에 있는 일부 carotenoid는 HPLC 분석으로 검출되지 않았다. 따라서 측정값을 PLS regression 변수에서 0으로 처리 하였다. PLS regression 모델을 구축하기 위해, 각 그룹의 대표 데이터 세트를 검증에 위한 테스트 세트로 선택했고, 남아있는 데이터를 트레이닝 셋으로 사용하여 PLS 모델을 만들었다. FT-IR 스펙트럼 데이터와 HPLC 분석을 이용한 carotenoid 성분의 PLS regression 모델링에 예측결과를 나타냈다(Table 2.). 감귤 과피에 모두 7개의 carotenoid 화합물들이 트레이닝 셋으로부터 높은 예측의 PLS regression 모델링이 나타났다. Antheraxanthin ( $R^2=0.9117$ ),  $\beta$ -carotene ( $R^2=0.8816$ ),  $\beta$ -cryptoxanthin ( $R^2=0.8856$ ), violaxanthin ( $R^2=0.7305$ )의 테스트 셋에서 높은 예측이 나타났다(Fig. 17). 이러한 결과는 antheraxanthin,  $\beta$ -carotene,  $\beta$ -cryptoxanthin, violaxanthin 성분은 추가적인 HPLC 분석 없이도 FT-IR 스펙트럼 데이터로부터 비교적 높은 예측값을 나타냈다. 이들 4개의 carotenoid 역시 예측값과 실측값 사이에 linear regression을 통해 높은 상관관계를 나타냈다(Table 2.). Antheraxanthin,  $\beta$ -carotene,  $\beta$ -cryptoxanthin, violaxanthin의 R-squared (RSQ) 값이 각각 0.94758, 0.85301, 0.89147, 0.77467로 나타났다. 이들 4개의 carotenoid는  $p < 0.01$ 로 의미 있는 결과로 나타났다(Table 2.). 이러한 통계 결과는 PLS regression 모델링이 감귤 과피에 있는 carotenoid 성분에 대한 좋은 예측 가능성을 가지고 있음을 나타냈다. 일부 예외를 제외하고, carotenoid 함량의 예측(FT-IR spectra data set)과 carotenoid 함량의 실측(HPLC analysis data set)의 관계가 전반적인 높은 상관관계임을 나타냈다.

일반적인 carotenoid peak은  $1,525-1,155\text{cm}^{-1}$ 의 FT-IR 스펙트럼 부위에서 관

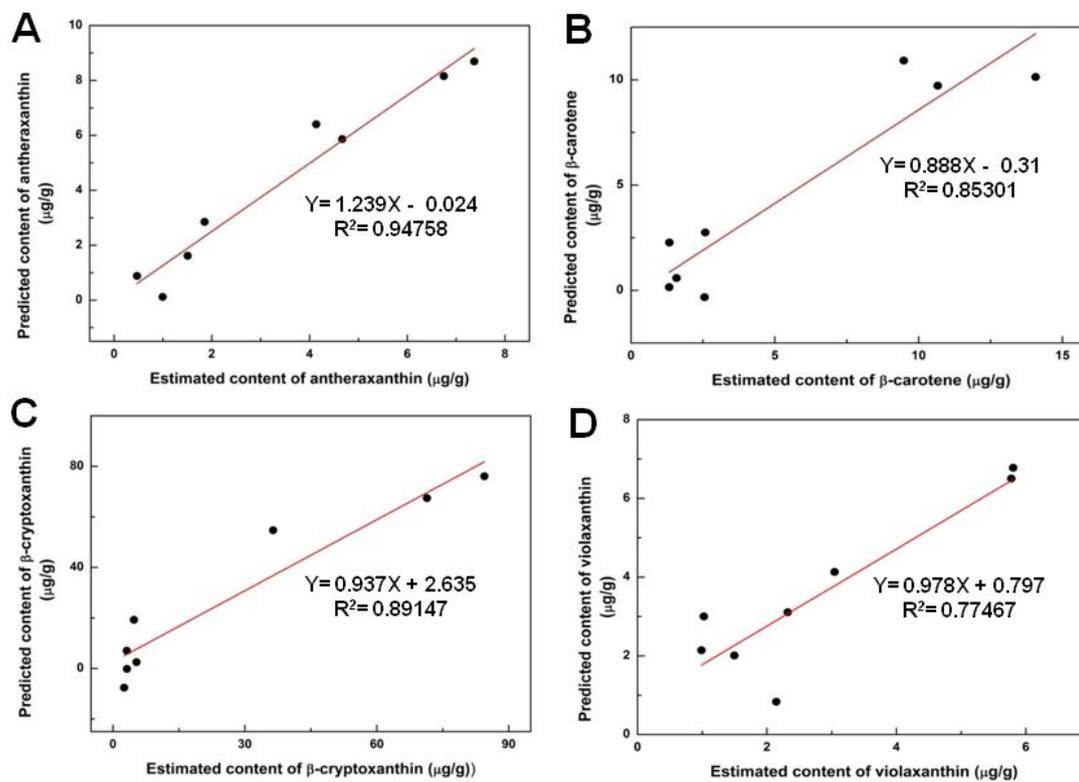
찰이되고, carotenoid의 전형적인 C=C과 C-C stretching에 일치한다.<sup>26</sup> 과실이 성장하면서 carotenoid 성분은 점진적으로 증가하였다. 과피의 FT-IR 스펙트럼을 보면, 과실의 발달과 성숙과정에서  $1,525\text{cm}^{-1}$  peak은 감소하고,  $1,150\text{cm}^{-1}$  peak은 현저하게 증가하였다(Fig. 12A). 이러한 결과는 과피에서 FT-IR 스펙트럼 데이터가 순수 carotenoid의 정확한 정량적인 정보를 제공할 수 있음을 알 수 있었다. 그러나, 본 연구에서 FT-IR 스펙트럼에 대한 전세포 추출물을 준비했기 때문에 carotenoid peak은 일반적인 carotenoid 부위에서 sugar와 phenolic 화합물 등 다른 주요 화합물과 중첩되어 나타날 수 있다. Fig. 12에서 나타난 것과 같이, polysaccharides ( $900\text{-}1,200\text{cm}^{-1}$ )과 proteins/amides ( $1,500\text{-}1,700\text{cm}^{-1}$ )이 FT-IR 스펙트럼 변화가 매우 강하게 나타났다. 따라서, polysaccharide ( $900\text{-}1,200\text{cm}^{-1}$ ) peak에 직, 간접적으로 작용하는 여러 주요 화합물의 영향을 제거한다면 FT-IR 스펙트럼 데이터 기반 carotenoid 성분의 PLS 모델링이 가능할 것이다. 이 결과는 sugar 부분과 주요 화합물은 과실 발달과 성숙 과정에서 carotenoid 성분들의 질적, 양적 변화와 관련이 높은 것으로 판단된다.

결론적으로, FT-IR 스펙트럼 기술은 과실이 발달과 성숙 과정에서 감귤의 대사체 변화를 관찰하는데 사용될 수 있다. polysaccharides, phenolic compounds, carotenoid들이 다른 화합물의 성분으로 변화하는 것을 간단하고 신속한 방식으로 관찰 할 수 있다. 특히, 성숙 시기에 일부 carotenoid 축적되는 것을 전통적인 chromatographic 분석 방법을 사용하지 않고, FT-IR 스펙트럼 기술에 HPLC와 PLS 모델링을 접목시켜 쉽게 관찰할 수 있다. FT-IR은 신속하고 간편하기 때문에 감귤 품종의 생식 특성에 신속한 평가를 위한 훌륭한 도구가 될 수 있습니다. 또한, FT-IR 스펙트럼으로부터 carotenoid 성분의 예측은 최적의 수확시기를 결정하는데 중요한 지표가 될 수 있습니다. 이러한 새로운 기술의 적용은 감귤 육종 연구에 제한적인 육종 방법을 극복하는데 도움이 될 것이다. 본 연구는 감귤 산업의 궁극적 목표인 감귤에 영양 가치 증가와 생산의 최적화에 기여할 수 있을 것이다.

**Table 2.** Summary of Carotenoid Content Prediction from FT-IR Spectroscopic Data by PLS Regression Modeling<sup>a</sup>

	a-carotene	antheraxanthin	b-carotene	b-cryptoxanthin	lutein	violaxanthin	zeaxanthin	Total carotenoids
RSQ value of training set	0.98905576	0.99324019	0.995874139	0.99088667	0.983475676	0.99019493	0.927574257	0.969285806
RSQ value of test set	0.696095929	0.911741598	0.881687131	0.885678425	0.581078534	0.730553899	0.115526478	0.398342836
SD of RSQ value of test set (10 times)	0.189611572	0.010209505	0.073395503	0.037346752	0.544885618	0.168078636	0.152641436	0.30388006
P-value	0.08276	2.88455E-5	6.56936E-4	2.60844E-4	0.01589	0.00244	0.03706	0.05896

<sup>a</sup>Averages and standard deviations of R-squared (RSQ) values from the test set are represented after 10 iterations.



**Fig. 17** Linear regression plots of predicted and estimated carotenoid content. Plots are shown for (A) antheraxanthin, (B)  $\beta$ -carotene, (C)  $\beta$ -cryptoxanthin, and (D) violaxanthin. Predicted values from FT-IR spectroscopic data by PLS regression and estimated values for carotenoid compounds from averages of HPLC analysis were used in linear regression analysis.

**6. FT-IR 스펙트럼 데이터의 다변량통계분석를 이용한 돌연변이  
감귤의 당산도 예측 모델링**

(Sugar content and acidity of mutant Citrus (*Citrus unshiu* Marc. cv. Miyagawa) using FT-IR Fingerprinting Combined by Multivariate Analysis and Quantitative Prediction of Functional Compounds by PLS Regression Modeling)

## 6. 1. Abstract

We established a high throughput screening system of Citrus fruits lines which contain higher contents of sugar content and acidity content using HORIBA and Fourier transform infrared (FT-IR) spectroscopy in combination with multivariate analysis. The sugar contents from 5 citrus fruits varied from 0.84 to 10.3°Brix. The acidity contents also varied from 0.61 to 0.99%. FT-IR spectra confirmed typical spectral differences between the frequency regions of 1,700–1,500, 1,500–1,300 and 1,100–950  $\text{cm}^{-1}$ , respectively. These spectral regions were reflecting the quantitative and qualitative variations of amide I, II from amino acids and proteins (1,700–1,500 $\text{cm}^{-1}$ ), phosphodiester groups from nucleic acid and phospholipid (1,500–1,300 $\text{cm}^{-1}$ ) and carbohydrate compounds (1,100–950 $\text{cm}^{-1}$ ). Principal component analysis (PCA) and subsequent partial least square-discriminant analysis (PLS-DA) were able to discriminate 5 citrus fruits lines into three separate clusters corresponding to their taxonomic relationship. The quantitative prediction modeling of sugar contents and acidity contents from citrus fruits lines were established using partial least square regression algorithm from FT-IR spectra. The regression coefficients( $R^2$ ) between predicted values and estimated values of sugar contents and acidity contents were 0.99 respectively. These results showed that quantitative predictions of sugar contents and acidity contents were possible from FT-IR spectra of citrus fruits lines with higher accuracy. Therefore we suggested that quantitative prediction system established in this study could be applied as a rapid selection tool for high yielding citrus fruits lines.

## 6. 2. Introduction

감귤은 제주도에서 2008년 이후 6,000억원이상의 생산액을 꾸준히 나타냈고, 국내 과수 총생산량의 1/4 이상을 차지할 만큼 생산량이 높은 과실이다. 최근 중국, 일본, 미국, 유럽과의 FTA 등 수입개방화로 인해 감귤류의 수입량이 증가하면서 국내 시장에 많은 비중을 차지하고 있다. 이에 따라 국내 감귤산업의 생존이 위협받고 있는 실정이다. 또한, 2012년부터 시행된 UPOV 협약에 따라 제주도에서 재배되는 99%의 일본품종 감귤이 품종 보호되면서 로열티에 대한 문제가 발생하고 있다. 그 중 일부 품종인 궁천, 흥진, 일남 1호 등이 품종보호기간 25년이 만료되었지만, 한라봉, 천혜향, 감평, 진지향 등은 품종보호가 되고 있다(Jun and Yoon, 2013).

품종보호뿐만이 아니라 2003년 이후 급격히 증가하고 있는 오렌지 수입량이 12만 톤 이상으로 증가하고 있고, 한·미 FTA 발효에 따라 가격이 싸지면서 국내 시장을 위협하고 있다. 이것은 국내 감귤농가에 큰 피해로 다가오고 있고, 비상품 감귤의 가공적인 측면에서도 많은 피해가 발생하고 있다.

이러한 문제점들을 극복하기 위한 한 가지 방법으로 돌연변이 육종이 대두되고 있다. 이 방법은 인위적으로 유전적 변이를 유도하여 개체의 형질을 변화시키는 방법이다. 전세계적으로 80년 이상 사용되었고, 1960년대부터 국내에서도 벼, 콩, 마늘 등의 다양한 품종에 활용되고 있다. 하지만 과수에서는 주로 교배 육종이 사용되고 있어 돌연변이 육종은 거의 이루어지고 있지 않다. 또한, 감귤에서는 당산도 우수비가 다른 작물에 비해 매우 중요하게 작용하고 있고, 아직까지 당산도에 대한 예측 모델이 보고된바 없다.

최근 오믹스 연구기법의 발달에 힘입어, 대사체 수준에서 시료간의 양적, 질적 패턴 차이를 구명하는 대사체학 연구가 활발하게 이루어지고 있다. 대사체 연구는 주로 Fourier transform infrared (FT-IR) spectroscopy, proton nuclear magnetic resonance (HNMR) spectroscopy 및 mass spectrometry (MS) 등의 분석 데이터를 이용하여 진행되고 있으며 각종 시료의 식별 및 대사체 마커 개발에 활용이 되고 있다(Krishnan et al., 2005). 특히, FT-IR 분석 방법은 스펙트럼이 가지고 있는 다양한 정보를 다변량통계분석을 이용하여 신속하고

정확하게 구분할 수 있다(Gallardo-Velázquez et al., 2009). 따라서 대사체학은 대사성분의 작용기작 구명 등 기초학문적 활용은 물론 농산물의 품종, 원산지, 재배연령, 진위 식별, 식품이나 의약품 산업의 표준화 및 안정성 관련 다양한 산업적 활용이 이루어지고 있다. 대사체 분석데이터로부터 보다 유용한 생물학적 정보를 추출하기 위하여 다양한 다변량통계분석 기법이 필수적으로 동반되고 있다. 이중에서 대표적인 다변량통계분석기법중 하나인 partial least squares (PLS) regression은 시료로부터 정확한 정량분석 데이터와 동일 시료의 스펙트럼 데이터의 상관분석을 통해 다양한 성분들의 함량 예측 모델링 방법으로 활용이 이루어지고 있다( Bastiena et al., 2005; Höskuldsson, 1988; Mevik and Wehrens, 2007; Wold et al., 2001). 최근에는 PLS regression 기법을 이용하여 동백나무속 식물(*Camellia oleifera*)에서 지방산 함량 예측 모델링(Yuan et al., 2013), 사과, 오렌지 그리고 복숭아 주스에 함유되어 있는 carotenoid 성분 함량 예측 모델링(Leopold et al., 2011), 버섯에 존재하는 단당류와 다당류 함량 예측 모델링(Chen et al., 2012), 육류의 부패 및 손상에 대한 예측 모델링(Argyri et al., 2013) 이 가능한 것으로 보고되고 있으며 예측정확도 또한 매우 높아 산업적 활용 측면에서 실용화 가능성이 매우 높아지고 있는 실정이다.

본 연구에서는 돌연변이 육종을 통해 얻어진 과실을 대상으로 외형적인 특징이나 당도, 산도 등의 형질적인 특성을 관찰하여 조사하였고, FT-IR 스펙트럼 데이터의 다변량통계분석 기법을 이용하여 돌연변이 감귤의 신속한 구분 및 식별체계를 확립하고 더 나아가 돌연변이 감귤에 당도와 산도를 신속한 함량예측 모델링을 개발하여 수확 전 신속한 선별체계를 확립하고자 하였다.

## 6. 3. Material and Method

### 6. 3. 1. 식물재료

본 연구에서 사용된 감귤품종은 제주도 내에서 가장 많이 재배되는 궁천조생(*Citrus unshu* Marc. Cv. Miyagawa-Wase)의 접수를 이용하여 돌연변이 육종에 사용하였다. 방사선 조사에 사용된 접수는 매년 3월과 8월에 2회 채취한 다음 비닐봉지에 밀봉하여 냉장보관 하고, 제주대학교 방사선응용과학연구소에서 방사선( $^{60}\text{Co}$ )을 조사하여 돌연변이를 유도하였다. 감귤 시료는 2013년 8월초와 11월말에 궁천조생 1품종과 돌연변이로 유도된 4품종을 최소 3개에서 최대 5개까지 채취하고, 저온저장고(4°C)에 보관하면서 형질측정을 수행하였다. 또한, 돌연변이 감귤을 동결건조하고 건조된 감귤은 막자와 막자사발을 이용하여 미세한 분말로 분쇄하였으며 -70°C 초저온냉동고에 보관하면서 실험에 사용하였다.

### 6. 3. 2. 당산도 측정

감귤 시료의 당산도 측정은 착즙기를 이용하여 채취하였고, mesh를 이용하여 과즙을 통과 시켰다. 통과된 과즙은 10mL 주사기를 이용하여 약 4~5mL를 산당도분석장치(NH-2000, HORIBA, Japan)에 삽입하여 기기사용 매뉴얼에 따라 당도와 산도를 측정하였다.

### 7. 3. 3. FT-IR 스펙트럼 조사 및 스펙트럼 데이터 다변량 통계분석

감귤 5품종의 시료 분말 20mg을 tube에 넣고 20% methanol 용액을 200mL씩 첨가하여 잘 섞어주었다. 각 tube를 50°C 수조에서 20분간 반응시킨 다음 13,000rpm에서 15분간 원심분리한 후 상징액을 새로운 tube에 옮겼다. 회수한 상징액은 원심분리과정을 반복하여 찌꺼기가 들어가지 않도록 조심스럽게 새로운 tube에 상징액을 옮겼다. 회수된 상징액은 -20°C에서 보관한

다음 FT-IR 스펙트럼 조사에 사용하였다.

FT-IR(Fourier transform infrared) 스펙트럼 조사는 Tensor 27(Bruker Optics GmbH, Ettlingen, Germany)를 사용하였고, DTGS(deuterated triglycine sulfate) 검출기로 분석하였다. 추출된 각각의 시료  $5\mu\text{l}$ 를 384-well ZnSe plate에 분주하여,  $37^{\circ}\text{C}$  hot plate에서 약 20분간 건조하였다. 건조된 ZnSe plate는 Tensor 27에 장착된 HTS-XT(Bruker Optics GmbH) 고효율 자동화 장치를 이용하여 스펙트럼을 조사하였다. 각 시료의 스펙트럼은 총  $4,000 \sim 400\text{cm}^{-1}$  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. 6.5, 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, Auckland, New Zealand)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. 가공된 FT-IR 스펙트럼 데이터는 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) 분석(Fiehn et al., 2000; Trygg et al., 2007)을 수행하였다.

#### 6. 3. 4. 감귤 시료로부터 당산도 함량 예측 PLS 모델링

감귤 시료의 FT-IR 스펙트럼 데이터로부터 당산도 함량 예측 모델을 개발하였다. 5품종 감귤 시료의 FT-IR 스펙트럼 데이터와 동일 시료로부터 얻어진 당산도 함량 실측 정량 데이터를 이용하여 X 변수에는 FT-IR 스펙트럼 데이터를 사용하였고, 2개의 Y 변수에는 산당도분석장치(NH-2000, HORIBA,

Japan)를 이용하여 측정된 당산도 함량 정량 데이터를 각각 사용하였다. PLSR(partial least square regression) 분석은 R 프로그램(version 2.15.0)을 이용하였다. 예측 모델링의 정확도를 향상시키기 위하여 1개의 X 변수에 대해 2개의 Y 변수를 각각 교차검정을 실시하였다. 확립된 예측 모델링을 이용하여 각각의 감귤 시료의 당산도 함량 예측을 수행하였다. 함량 예측 모델링의 정확도를 조사하기 위하여 감귤 시료로부터 당산도 함량 실측치와 예측치에 대한 선형회귀분석을 수행하여 상관계수를 조사하였다.

## 6. 4. Results and Discussion

### 6. 4. 1. 감귤 시료로부터 당산도 함량 조사

산당도분석장치를 이용하여 감귤 시료에 존재하는 8월과 11월의 당산도 함량을 조사하였다(Table 3). 8월 감귤 시료의 전체적인 당도는 6.2~6.3°Brix로 나타났다. 이 시기는 감귤의 수분이 증발하고 당도를 축적하는 단계여서 낮은 당도 함량을 보였다. 또한, 산도는 M4가 1.75%로 가장 낮게 나타났고, 궁천조생이 2.19%로 가장 높게 나타났다. 돌연변이 감귤과 궁천조생이 당도보다는 산도에서 차이를 확연한 차이를 나타냈다(Table 3). 11월 감귤 시료에서는 M1이 10.3°Brix로 궁천조생 8.4°Brix에 비해 약 2°Brix 높게 나타났고, 조사된 감귤 시료 중에 가장 높은 당도를 나타냈다. 또한, M2 시료는 8.4°Brix로 궁천조생과 비슷한 당도를 나타냈고, M3와 M4 시료는 각각 9.7, 9.5°Brix로 궁천조생에 비해 약 1°Brix 높게 나타났다. 산도 또한 궁천조생과 돌연변이체가 차이를 나타냈는데, 궁천조생의 산도는 0.64%, M1에 경우 산도가 0.99%로 1%미만으로 나타났고, M2, 3 그리고 4는 산도가 각각 0.79, 0.61 과 0.65%로 1%미만의 산도 함량을 나타냈다(Table 3). 전체적인 당도와 산도 함량이 궁천조생과 M2가 비슷하게 나타나고 있고, M1, 3와 4가 높게 나타났다. 이 결과를 보면 돌연변이 감귤이 궁천조생에 비해 당도는 높고 산도는 낮게 나타나는 것을 보아 돌연변이 품종이 높은 가치를 나타냈다. 또한, Song et al. (2014)의 감귤 돌연변이 육종 현황 보고에 따르면 돌연변이 감귤의 경우 당도는 9~10°Brix가 대체적으로 나타난다고, 산도는 0.6-0.8%의 함량을 나타냈다는 보고와 일치하였다.

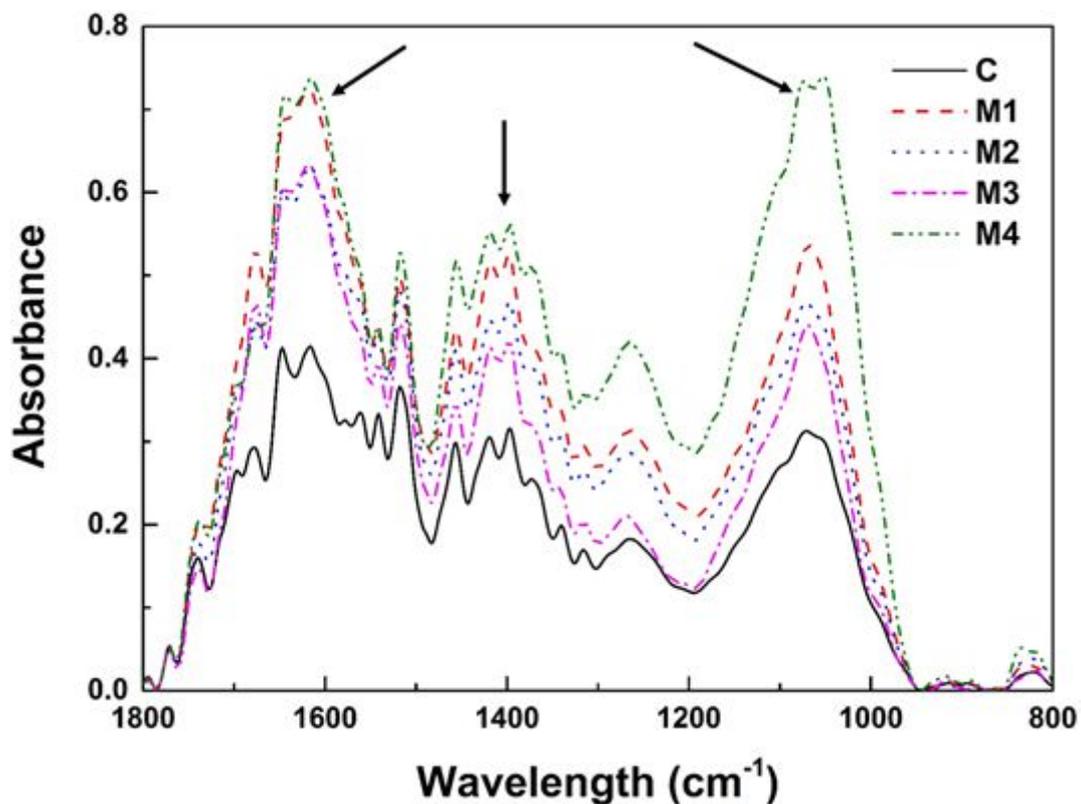
**Table 3.** Quantitative analysis of sugar content and acidity content from *citrus* using HORIBA analysis. All samples were run in five replicates. C : *Citrus unshu* Marc. Cv. Miyagawa-Wase, M1 : 6s8a68-12, M2 : 6s4c25, M3 : SHa1-2, M4 : SBa3

Sample	Lines Numbers	August		November	
		Sugar content (°Brix)	Acidity (%)	Sugar content (°Brix)	Acidity (%)
C	5	6.2±0.13	2.19±0.03	8.4±0.42	0.64±0.04
M1	5	6.2±0.29	1.82±0.08	10.3±1.88	0.99±0.25
M2	5	6.2±0.53	1.81±0.08	8.4±0.23	0.79±0.08
M3	5	6.3±0.38	2.09±0.16	9.7±0.83	0.61±0.06
M4	5	6.3±0.26	1.75±0.07	9.5±0.18	0.65±0.03

Data represent the mean±SD of measurements.

#### 6. 4. 2. 감귤 시료의 FT-IR 스펙트럼 비교분석

감귤 시료로부터 전세포 추출물의 FT-IR 스펙트럼 데이터 다변량통계분석을 통해 감귤의 대사체 수준에서 중간 유연관계 구멍 및 식별체계를 확립하였다. 각 감귤 시료의 FT-IR 스펙트럼을 비교한 결과 감귤 시료들은 FT-IR 스펙트럼상의 1,700-1,500, 1,500-1,300, 1,100-950 $\text{cm}^{-1}$ 부위에서 대사체의 양적, 질적 패턴 변화가 크게 이루어짐을 알 수 있었다(Fig. 18). FT-IR 스펙트럼의 1,700-1,500, 1,500-1,300 그리고 1,100-950 $\text{cm}^{-1}$ 부위는 각각 아미노산 및 단백질의 amide bond I 과 II, 핵산 및 인지질로부터 phosphodiester bond 및 인을 포함하는 유기산, 그리고 단당류나 복합 다당류를 포함하는 carbohydrates 계열의 화합물들의 질적, 양적 정보를 반영한다(D'Souza et al., 2008; Dumas and Miller, 2003; Lopez-Sanchez et al., 2010; Parker, 1983; Wolkers et al., 2004; Yee et al., 2004). 즉 FT-IR 스펙트럼 상의 질적, 양적 차이는 감귤에 함유되어 있는 아미노산이나 단백질, 지방산, 그리고 탄수화물계통의 화합물들의 질적, 양적 차이가 현저하게 나타남을 의미한다. 따라서 FT-IR 스펙트럼 분석은 감귤 시료들 사이에서 주요 대사체의 질적, 양적 변화를 조사하고 이들 화합물의 함량이 높은 계통이나 라인의 신속한 선별 수단으로 활용이 가능할 것으로 기대된다.



**Fig. 18** Representative FT-IR spectral from citrus fruits. FT-IR spectral ranges showed quantitative information of protein/amide I, II ( $1500-1700\text{cm}^{-1}$ ), phosphodiester group ( $1300-1500\text{cm}^{-1}$ ), and sugar compound ( $950-1100\text{cm}^{-1}$ ). Solid or dotted lines and abbreviations represent each citrus fruits species. C : *Citrus unshu* Marc. Cv. Miyagawa-Wase, M1 : 6s8a68-12, M2 : 6s4c25, M3 : SHa1-2, M4 : SBa3. Arrows indicate the FT-IR regions showing significant spectral variations between citrus fruits samples.

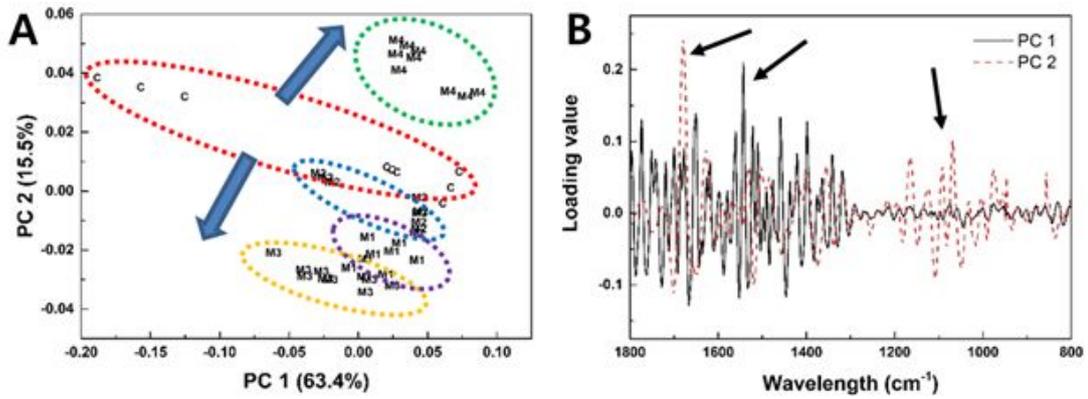
#### 6. 4. 3. 감귤 시료의 FT-IR 스펙트럼 데이터 다변량통계분석

FT-IR 스펙트럼 데이터의 PCA 분석 결과 PC 1과 2 score는 전체 변이량의 각각 63.4%, 15.5%의 설명력을 갖고 있으며 이는 전체 변이량의 약 78.9%를 반영하고 있음을 알 수 있었다(Fig. 19A). 이처럼 설명력이 높은 PC1과 PC2를 기준으로 감귤 시료들의 PCA score plot을 관찰한 결과, 감귤 시료들은 PC 2를 기준으로 크게 상하로 구분이 이루어지는 것을 관찰할 수 있었다. PCA score plot 상의 궁천조생을 중심으로 상하로 위치하는데, 궁천조생과 당도가 비슷한 M2 품종이 궁천조생 그룹과 같이 분포하고, 상하로 M3와 M4 품종 그룹들이 주로 위치하였으며, 당도가 가장 높은 M1은 M2와 M3 품종 사이에 위치하였다(Fig. 19A).

감귤 시료의 PCA score plot상에서 공간적인 분포위치가 주는 생물학적인 정보를 파악하기 위하여 먼저 감귤 시료들의 당도와 산도 함량 차이 결과(Table 4.)와 PCA score plot상의 이들 시료들의 위치와 상관관계를 조사하였다. 당도가 비교적 낮았던 시료인 궁천조생의 경우 PCA score plot상에서 중앙에 별도의 그룹을 형성하였으며, 반대로 당도 함량이 가장 높았던 시료인 M1의 경우 PCA score plot 상에서 M2와 M3 중앙에 별도의 그룹을 형성하며 분포하는 것을 관찰할 수 있었다. 이상의 결과를 미루어볼 때 PCA score plot 상에서 공간적인 분포위치와 이들 시료의 당도 함량과 밀접한 상관관계가 있음을 알 수 있었다.

PCA score plot 상에서 기능성 성분들의 함량에 따른 감귤 시료들의 집단 구분 및 클러스터링에 중요한 역할을 하는 FT-IR 스펙트럼 부위를 조사하기 위하여 PC 1과 2를 결정하는데 중요한 FT-IR 스펙트럼 부위를 조사하였다(Fig. 19B). Loading value 분석결과 감귤 시료의 상하 구분에 중요한 PC 1을 결정하는 중요한 FT-IR 스펙트럼 부위는 주로 1,700-1,500과 1,500-1,300 $\text{cm}^{-1}$  region임을 알 수 있었고, PC 2의 경우 주로 1,100-950 $\text{cm}^{-1}$  region임을 알 수 있었다(Fig. 23B). 이들 FT-IR 스펙트럼 부위는 감귤 시료의 FT-IR 스펙트럼 비교에서 보였던 차이가 큰 부위(Fig. 4)와 일치하는 것으로 amide I, II, 지방산과 carbohydrates 계열의 화합물들의 질적, 양적 차이가 감귤 시료의 대사체 수준 식별에 중요한 역할을 하고 있음을 의미하는 결과라 사료된다(Fig. 19B). 특히

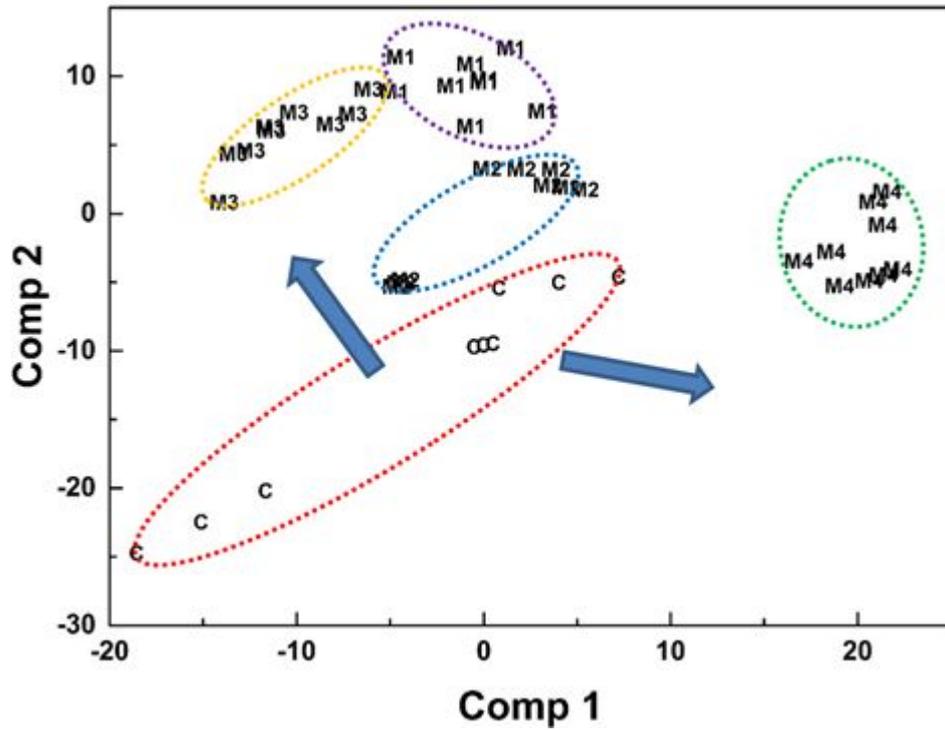
기능성 성분들의 함량 차이에 따른 감귤 시료의 구분에 있어서 중요한 역할을 하는 PC 2의 경우 carbohydrates 계열의 화합물들의 질적, 양적 차이가 크게 영향을 줌을 알 수 있었다. 이 결과로 미루어볼 때 이들 당 계열의 일차대사산물들의 양적 변화가 이차대사산물의 질적, 양적 변화와 양의 상관관계에 놓여 있음을 의미하는 결과라 사료된다.



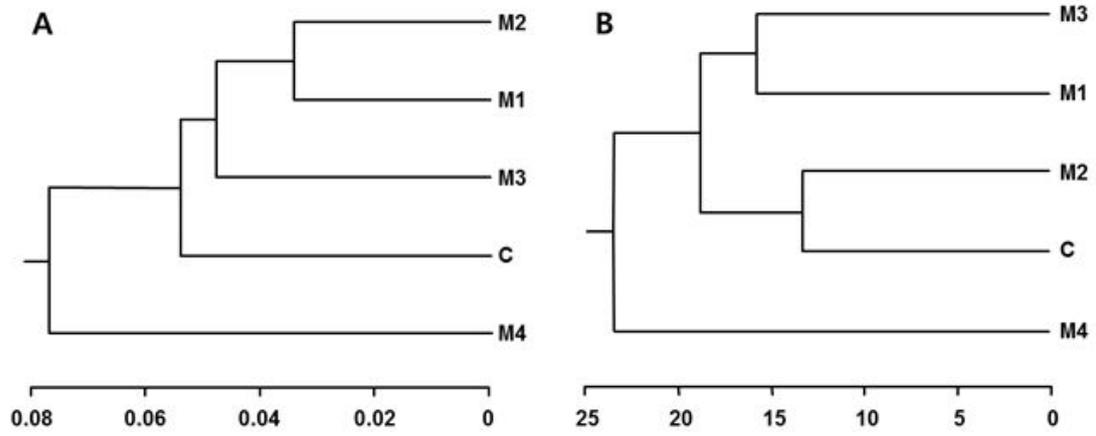
**Fig. 19** PCA scores (A) and loading values plot (B) of PCA scores from FT-IR data of citrus fruits lines. Dotted ellipses represent the clustering boundary of yam tuber samples with high (M1) and low (M2) contents of sugar contents and acidity contents. Abbreviations in the PCA score plots represent each citrus fruits samples: Solid line and dotted line represent PC 1 and PC 2 score, respectively (A); Arrows indicate the FT-IR regions playing important roles in citrus fruits sample clustering (B) C : *Citrus unshu* Marc. Cv. Miyagawa-Wase, M1 : 6s8a68-12, M2 : 6s4c25, M3 : SHa1-2, M4 : SBa3.

PLS-DA 분석의 경우 PCA 분석 보다 더욱 확실하게 감귤 시료들의 그룹 식별이 가능하였다(Fig. 20). 각각의 감귤 시료들의 그룹 경계는 PCA에 비해 더욱 작아졌으며 동일 품종에 속하는 각 시료들의 반복구들도 각 그룹 내에 위치하여 PCA보다 감귤의 식별능력이 향상되었음을 알 수 있었다. 당도 함량이 비교적 낮았던 시료인 궁천조생의 경우 PCA score plot상에서 중앙에 별도의 그룹을 형성하였으며, 당도 함량이 가장 높았던 시료인 M2의 경우 PLS-DA score plot 상에서 M3와 함께 좌측 상단에 별도의 그룹을 형성하며 분포하는 것을 관찰할 수 있었다. 당도의 함량이 궁천조생과 비슷한 M2 그룹은 주로 궁천조생 그룹과 비슷한 위치에 분포하였다(Fig. 20). 이상의 결과를 미루어볼 때 PCA와 마찬가지로 PLS-DA 역시 각 감귤 시료들의 score plot 상에서 공간적인 분포위치와 이들 시료의 당산도 함량과 밀접한 상관관계가 있음을 알 수 있었다.

HCA dendrogram 분석 결과를 보면 총 5개의 감귤 시료 간에 교배 육종 유연관계를 나타냄을 알 수 있었다(Fig. 21). PCA dendrogram을 보면 궁천조생을 중심으로 M1, M2, M3, M4가 분포되어 있다(Fig. 21A). 이는 궁천조생에서 돌연변이 되었기 때문에 궁천조생과 M1, M2, M3 3가지 품종이 형질적으로 유연관계가 높을 것으로 판단되고, M4는 형질적으로 유연관계가 낮을 것으로 판단된다. PLS-DA dendrogram에서 또한 감귤 품종 간에 유연관계를 확연히 나타냈다(Fig. 21B). 궁천조생과 M2 품종 간에 형질이 가장 높은 유연관계를 나타냈고, M1과 M3가 형질적으로 유연관계가 높게 나타났다. 하지만 M4는 형질적인 유연관계에서 낮은 관계성을 나타냈다. 이처럼 돌연변이 육종에 따라 형질 표현이 다르게 나타나 이와 같은 결과를 관찰할 수 있었다.



**Fig. 20** PLS-DA score plot of FT-IR data from citrus fruits lines. Dotted eclipses and capitals represent the clustering boundary of citrus fruits samples with high (M1) and low (M2) contents of sugar contents and acidity contents. Abbreviations in the PLS-DA score plots represent each citrus fruits samlpes. C : *Citurs unshu* Marc. Cv. Miyagawa-Wase, M1 : 6s8a68-12, M2 : 6s4c25, M3 : SHa1-2, M4 : SBa3

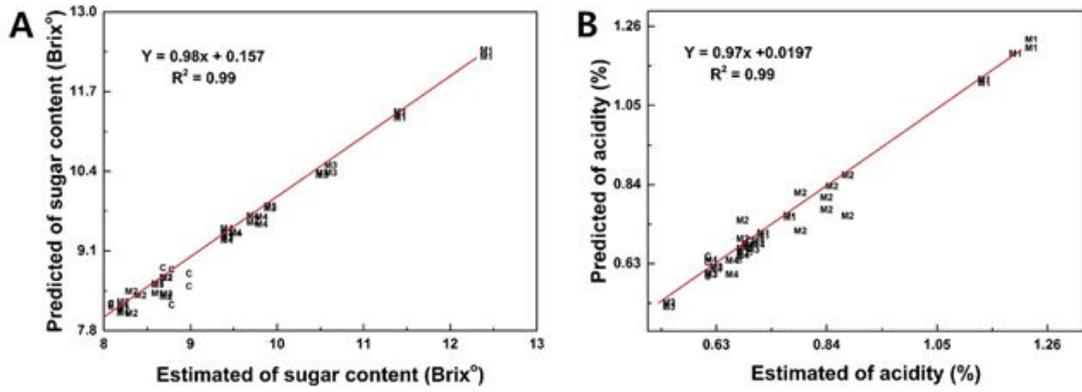


**Fig. 21** Hierarchical dendrogram of FT-IR data from Citrus fruits. C : *Citrus unshu* Marc. Cv. Miyagawa-Wase, M1 : 6s8a68-12, M2 : 6s4c25, M3 : SHa1-2, M4 : SBa3

#### 6. 4. 4. 감귤 시료의 FT-IR 스펙트럼으로부터 당도와 산도 함량 예측 PLS regression 모델링

FT-IR 스펙트럼 데이터와 산당도분석장치를 이용하여 측정된 당도와 산도 함량 데이터를 이용하여 각 성분별 함량 예측 PLS 모델링을 개발하였다(Fig. 22). 당도의 경우, PLS 모델링을 통해 FT-IR 스펙트럼 데이터로부터 예측된 함량 값과 동일 시료의 실측 함량 값을 이용하여 회귀분석을 수행한 결과 상관계수가  $R^2 = 0.99$ 로 높게 나타났다(Fig. 22A). 산도의 함량 예측 모델링을 통한 선형회귀분석 상관계수는  $R^2 = 0.99$ (Fig. 22B)로 나타났다. 이들 결과는 감귤 시료의 FT-IR 스펙트럼 데이터만으로도 이들 시료 내에 존재하는 당도와 산도 함량을 90%정도 정확성으로 함량을 예측할 수 있는 결과라 사료된다.

아직 감귤 시료에서 당산도 함량 예측 모델링 연구가 보고된 바 없어 정확한 비교는 어렵다. 하지만 네이블오렌지에 함유하고 있는 성분에 대한 모델링이 보면 색차에 대한 상관계수( $R^2$ )가 0.95로 나타났고, 과피건조정도에 대한 상관계수( $R^2$ )는 0.92로 나타났다. 또한, 과피에 함유되어 있는 fructose에 대한 상관계수( $R^2$ )는 0.8, glucose에 대한 상관계수( $R^2$ )는 0.79, sucrose에 대한 상관계수( $R^2$ )는 0.77, total carbohydrates에 대한 상관계수( $R^2$ )가 0.78로 나타났고 보고된 바 있다(Magwaza et al., 2014). 이처럼 본 연구에서 보여준 감귤의 당산도 모델링 상관계수가 높게 나타났다. 또한, 수확전 감귤의 당도와 산도를 먼저 파악할 수 있는 가능성을 확인하였고, 향후 PLS 모델링의 정확도 향상을 위한 시료의 분석조건 표준화, 모델링에 사용되는 분석시료의 개체 수 증대, 기능성 성분의 양적 변화가 큰 시료가 모델링에 활용되는 등의 보완이 이루어진다면 보다 정확한 성분 예측 모델링으로 활용이 가능할 것으로 기대된다.



**Fig. 22** Linear regression analysis between estimated and predicted values of sugar contents (A) and acidity contents (B) by the PLS regression model from FT-IR spectral data. Regression coefficient values ( $R^2$ ) are 0.99, respectively.

본 연구는 2005년에 방사선 돌연변이육종을 시작하여 2009년부터 본격적으로 과실을 수확하여 측정을 시작하였다. 돌연변이 가지에 달린 과실을 2009년 1개에서부터 점차적으로 3개 이상의 과실을 측정하였다. 2013년도는 2009년부터 2012년까지 형질이 고정되어 있는 가지를 선발하여 과실을 수확하고 형질을 측정하였다. 2013년도 돌연변이 과실 당산도 측정치는 당도는 평균  $8.4 \pm 1.4^\circ\text{Brix}$ 로 나타났고, 최소  $4.4^\circ\text{Brix}$ 에서 최대  $12.9^\circ\text{Brix}$ 의 분포도를 나타냈다. 산도는 평균  $0.73 \pm 0.2\%$ 로 나타났고, 최소 0.35%에서 최대 1.29%로 나타냈다. 당도에 따른 분포도를 보면,  $8 \sim 10^\circ\text{Brix}$ 가 많은 비중을 차지하였고, 이중에서도  $8^\circ\text{Brix}$ 이하가 42.7%로 가장 높은 비중을 차지하였다. 또한  $11^\circ\text{Brix}$ 이상의 과실이 5.5%로 당초 연구 목표였던  $11^\circ\text{Brix}$ 이상의 과실도 지속적으로 나타났다. 당도에 따른 산도가 과실에서 중요하게 작용하는데 산도의 분포는 0.6~0.8%가 56.8%로 가장 많은 분포도를 볼 수 있었다. 이는 감귤의 신맛이 많이 사라졌다고 볼 수 있는데, 산도가 감귤의 맛을 결정하는 중요한 요인으로 1%이하로 나타나게 되면 당도가 낮게 나타나도 신맛이 낮아 달콤한 맛을 느끼게 해준다. 이처럼 감귤에 당도와 산도는 감귤의 품질을 결정하는데 가장 중요한 역할을 하는 요인이다. 이러한 요인들을 수확 시기인 11월경에 확인하고 판단하는 것이 아니라, 예측 시스템이 필요하다.

감귤의 대사체 정보를 이용하여 당도와 산도를 예측하는 모델링은 현재까지 보고된바 없다. 최근 네이블오렌지에 함유하고 있는 성분 에 대한 모델링이 보고되었지만, 색차에 대한 모델링, 과피건조정도에 대한 모델링, 과피에 함유되어 있는 fructose에 대한 모델링, glucose에 대한 모델링, sucrose에 대한 모델링 그리고 total carbohydrates에 대한 모델링이 보고되었을 뿐 당도와 산도에 대한 모델링은 보고되지 않았다(Magwaza et al., 2014).

본 연구에서 확립된 대사체 수준에서 감귤의 당산도 함량 예측 모델링은 아직 동정 및 표준화가 이루어지지 않은 다수의 감귤 시료로부터 품질특성 평가 수단으로 활용이 가능하여 우수한 감귤 품종이나 계통의 신속한 선발 수단으로서 활용이 가능할 것이다. 또한, 선발된 라인은 육종의 직접적인 소재로 활용이 가능하여 감귤의 기능성 품종개발 가속화에 기여 할 수 있을 것으로 예상된다.

따라서 본 연구 결과를 토대로 기능성 성분 예측 모델링을 구축하고, 활용한다면 보다 쉽고, 빠르고 그리고 비교적 정확하게 당도와 산도를 예측할 수 있기 때문에 기능성 감귤 품종의 육성을 크게 단축시킬 수 있을 것으로 사료된다.

## 7. Genomic DNA의 FT-IR 스펙트럼을 이용한 적배추

### 부모로부터 F1 식별

(Fourier transform infrared (FT-IR) spectroscopy of genomic DNA to discriminate F1 progenies from their paternal lineage of Chinese cabbage (*Brassica rapa* subsp. *pekinensis*))

## 7. 1. Abstract

Fourier transform infrared spectroscopy (FT-IR) spectroscopy in combination with multivariate analysis was used to discriminate two different F1 hybrid lines from their parental inbred lines. Genomic DNA was isolated from leaves of three parental lines and two F1 hybrid lines of *Brassica campestris* subsp. *pekinensis*. Purified genomic DNA was analyzed by FT-IR spectroscopy in the spectral region from 4,000 to 400  $\text{cm}^{-1}$ . FT-IR spectra confirmed typical spectral differences between the frequency regions of N-H stretching (amide I) and C=O stretching vibrations (amide II) as well as  $\text{PO}_2^-$  ionized asymmetric and symmetric stretching. Principal component analysis was able to discriminate between F1 hybrid progenies depending on their parental lineages, even though they share the same maternal background. Partial least squares discriminant analysis gave a more clear discrimination between the two parental lines and their hybrid progenies. These FT-IR spectral differences might be directly related to subtle changes in the base functional group and backbone structures of genomic DNA. Considering these results, this technique could provide a solid research foundation for FT-IR spectral-based rapid diagnosis, selection, and discrimination of parental lines from their progenies. Furthermore, this technique could be applied to test purity in the hybrid seed industry.

## 7. 2. Introduction

적배추(*Brassica campestris* subsp. *pekinensis*)는 한국, 일본 그리고 중국에서 대중적으로 소비되는 채소 작물이다. 십자화과 채소는 중요한 기능성 성분인 glucosinolates, carotenoids, flavonoids 그리고 phenolic compounds를 함유하고 있다(Kim et al. 2004; Harbaum et al. 2007). 십자화과 채소에 함유되어 있는 여러 가지 성분들은 암에 위험을 감소시킨다(Giovanucci et al. 2003). 이러한 채소들의 소비는 암 또는 각종 만성 질환을 보호하는 역할을 한다(van Poppel et al. 1999; Mas et al. 2007). 인간 건강에 대한 관심의 증가로, 많은 새로운 품종과 기능성 채소의 재배가 기존 육종방법을 통해 개발되었다. 더욱이, 국제 UPOV에 따라 지적재산권 보호가 강해지면서, 새로운 품종과 재배의 특성을 구별하는 기술 개발이 필요하다.

작물의 새로운 품종을 구분하는 방법에는 형태적인 분류와 농작 방법이 일반적으로 사용되고 있다. 그러나, 이런 기술들은 여러 작물에 걸쳐 적용하기에는 많은 제한이 있다. 최근 RFLP, RAPD, AFLP(Williams et al. 1990; Thormann et al. 1994; Kwon et al. 2003), SSR(simple sequence repeat) 마커를 이용하여 새로운 품종을 찾아내고 있다. 특히, SSR 마커는 대립 유전자 형질 관련하여 양적 형질 유전자를 포함하는 특성화 게놈 영역에 대한 연구가 증가하고 있다. 또한 SSR 기술은 십자화과 채소 작물에도 적용되었다(Ramchiary et al. 2011). 게다가, Single Nucleotide polymorphisms (SNPs)에 따른 분자 마커의 사용이 작물의 분자 육종 구현을 빠르게 촉진하고 식물 유전자 분석에 널리 퍼져있다(Mammadov et al. 2012).

농작물의 생산성을 높이기 위해, 잡종 강세는 F1 종자를 생산하는데 이용되고 있다. 그러나, 잡종 강세는 신품종 육성에서 품종에 대한 분쟁을 완전히 피할 수 없다. F1 잡종 종자의 상태는 종자 생산의 표준화 부족, 종자 생산 분야의 불안정, 종자의 기계적 혼합물 등의 여러 가지 요인으로 인해 종종 낮게 나타난다. 재배 실험은 정확하고 신뢰할 수 있지만, 전체 과정은 너무 길고 비용이 많이 발생한다. 분자 마커의 동정은 매우 정확 하지만, 이들은 표적 유전자 서열에 대한 정확한 정보가 필요로 한다. 따라서 정확하고 빠른 비파괴 방법을 개발할 필요가

있다. FT-ICR MS(Fourier transform ion cyclotron resonance mass spectrometry)은 대사체 분석에 대한 강력한 방법이고, 유전자의 변화에 결과를 확인할 수 있는 기술이다(Mungur et al. 2005). FT-ICR MS의 단점은 대사체 분석 실험에 일반적으로 사용되기 힘든 높은 비용에 있다(Lei et al. 2011). 그것에 비해 FT-IR은 화학, 농업, 제약, 식품 산업 등 다양한 연구 분야에서 널리 사용되고 있다.

FT-IR은 분자 안에 화학결합의 형태를 확인할 수 있는 강력한 기술이다. FT-IR 스펙트럼의 각각의 영역은 분자내에 특징적인 화학 작용기의 진동과 연관성이 높다. IR 스펙트럼은 각각의 샘플에 lipids, proteins, nucleic acids 그리고 polysaccharides을 포함하고 있는 cell의 전체 화학구조를 매우 구체적으로 나타낸다(Lee and Chapman 1986). FT-IR 스펙트럼은 DNA에 전체 화학 구조와 서로 다른 작용기작 또는 화학적 기본 구조를 스펙트럼에 반영된다(Naumann et al. 1991). IR 스펙트럼의 변화는 nucleic acid의 분자 모양에 의해 결정되며, 가령 광범위한 수소 결합, 대칭구조의 phosphate 증가, 단백질의 C-OH 결합 감소를 나타낸다(Dukor 2002). FT-IR은 미생물을 구별하는데 강력한 툴로 사용되었고(Ngo-Thi et al. 2003; Naumann et al. 2005; Fischer et al. 2006; Preisner et al. 2007; Sundaram et al. 2012), 그리고 생식 세포질의 대사 평가(Yesudas et al. 2013), 질병과 암의 DNA 표현형을 식별하는데 이용되었다(Malins et al. 1997, 2003, 2005). 또한 암 진단에 중요한 식별 바이오 마커로 활용되고 있다(Argov et al. 2002; Sahu et al. 2008). 옥수수 식물의 유전자형으로부터 DNA 판별에 성공적으로 적용 되었다(Emura et al. 2006). 또한, 쌀에서 동일 라인에 단일 유전자와 대립 차이의 작은 변형을 구분하는데 사용되었으며, DNA 수준에서 *indica*와 *japonica*의 유전적 차이가 쌀 품종 간에 높음을 확인하였다(Emura et al. 2006). 그러나 아직까지 F1에 대한 식별은 보고된바 없다.

본 연구는 FT-IR 스펙트럼을 이용하여 genomic DNA를 통해 F1 교배 라인을 신속하게 구분함으로써 다양한 육종에 기여할 수 있는 식별 시스템 구축하는데 있다.

## 7. 3. Material and Method

### 7. 3. 1. 식물재료

적배추(*Brassica campestris* subsp. *pekinensis*)의 모든 식물 재료는 아시아 종묘에서 제공되었다. 본 연구에서는, 모든 샘플 그룹은 1개의 모계라인, 2개의 서로 다른 부계 라인, 그리고 교배들 통해 얻은 2개의 서로 다른 F1 자손 라인을 포함한 서로 다른 5개의 적배추 라인을 사용하였다. 2개의 F1 자손 라인은 같은 모계이지만 서로 다른 부계와 교배되었다. 각각의 라인 샘플은 서로 다른 5개의 식물로 구성되었다. 4주간 온실에서 성장 시킨 후, 잎을 수집하여 액체 질소를 이용하여 동결 시켜 사용하였다.

### 7. 3. 2. FT-IR 스펙트럼에 대한 Genomic DNA 추출

수집된 식물 잎 시료를 genomic DNA 추출에 사용하였다. 잎 시료는 막자사발에 액체 질소를 부어 막자로 갈아 분말형태로 만들었다. 잎 분말 시료를 1.5ml EP tube에 넣고, -70°C에서 24시간 동결건조 시켰다. 동결 건조된 시료 10mg을 사용하여 G-spin genomic DNA extraction kit (iNtRON Biotechnology Inc., Korea) 방법에 따라 DNA를 추출하였다. 이 buffer 시스템은 column을 사용하여 빠르고 쉽게 DNA를 선택적으로 추출할 수 있는 효과적인 방법이다. 추출된 DNA는 TE buffer를 이용하여 현탁했다. DNA 농도와 순도는 NanoDrop 2000을 이용하여 280, 260와 230nm에서 측정하였다. 추출된 DNA 순도를 확인하기 위해서 흡광비( $OD_{260}/OD_{280}$ )를 계산하였다. 이 비율은 polyphenol, carbohydrate contamination과 함께 protein을 측정하여 제공한다(Sambrook et al. 1989). TE buffer를 이용하여 DNA 최종 농도를  $40\text{ng}/\mu\text{l}$ 으로 조정하였다. 효과적인 FT-IR 스펙트럼 측정을 위해 genomic DNA를 체크하여 DNA 오염도를 최소화 하였다. Genomic DNA 농도는 아가로스 겔 전기영동을 통해 확인하였다. 확인된 genomic DNA는 -20°C에 보관하여 실험에 사용하였다.

### 7. 3. 3. Genomic DNA의 FT-IR 스펙트럼 조사

FT-IR(Fourier transform infrared) 스펙트럼 조사는 Tensor 27(Bruker Optics GmbH, Ettlingen, Germany)를 사용하였고, DTGS(deuterated triglycine sulfate) 검출기로 분석하였다. 추출된 genomic DNA 각각의 시료 5 $\mu$ l를 384-well ZnSe plate에 분주하여, 37°C hot plate에서 약 20분간 건조하였다. 건조된 ZnSe plate는 Tensor 27에 장착된 HTS-XT(Bruker Optics GmbH) 고효율 자동화 장치를 이용하여 스펙트럼을 조사하였다. 각 시료의 스펙트럼은 총 4,000 ~ 400cm<sup>-1</sup> 범위에서 그리고 4cm<sup>-1</sup> 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. 6.5, Bruker Optics Inc.)를 사용하였다. IR 스펙트럼은 plate 스펙트럼 (background)을 뺀 샘플의 스펙트럼을 나타냈다. 다변량 분석을 위해 기본 FT-IR 스펙트럼을 OPUS software (ver. 7.0)를 사용하여, baseline correction, normalization 그리고 smoothing 실시하여 전 처리하였다. 전 처리된 스펙트럼의 2차 미분을 OPUS software (ver. 7.0)를 통해 실시하였다. 처리 된 스펙트럼 데이터는 Pearson's 상관 분석과 다변량통계 분석을 실시하였다.

### 7. 3. 4. FT-IR 스펙트럼의 다변량통계분석

다변량통계 분석은 전체 스펙트럼에서 1,800-800cm<sup>-1</sup> 영역의 FT-IR 스펙트럼 데이터를 이용하여 principal component analysis (PCA), hierarchical clustering analysis (HCA) 그리고 partial least squares discriminant analysis (PLSDA)을 실시하였다. 2차 미분된 FT-IR 스펙트럼 데이터는 R 통계 분석 프로그램 (버전 2.15.0, Auckland, New Zealand)을 통해 분석되었다. PCA는 non-linear iterative partial least squares (NIPALS) 알고리즘에 따라 수행하였다(Wold 1966). PLS-DA는 PCA 다음에 R 프로그램을 사용하여 수행 하였다. PCA 분석은 PCA scores를 추출하여 상관 행렬을 계산에 사용하고, PLS-DA는 F1 교잡 라인과

부모 교배 라인간의 대사체 식별하였다. PCA loading value를 실행하여, F1 교잡 라인과 부모 교배 라인간의 대사체 식별에 대한 중요 변수를 확인하였다. Hierarchical dendrogram은 FT-IR 데이터의 PLS-DA로부터 UPGMA (unweighted pair group method with arithmetic mean analysis)를 R program (version 2.15.0)으로 분석하여 euclidean distance를 측정하고 각 시료의 유연관계를 나타냈다.

#### **7. 3. 4. FT-IR 스펙트럼 데이터를 이용한 Classification, prediction 그리고 cross-validation**

FT-IR 스펙트럼 데이터를 기반으로 DNA를 통한 교잡육종에 대한 예측을 위해서 PLS regression 분석을 적용하여 확립하였다. PLS regression을 실시하기 위해서 2개의 변수 X와 Y에 대한 정보를 연결시켰다; X 변수에는 FT-IR 스펙트럼 데이터를 적용하였고, Y 변수에는 DNA의 번호를 그룹화 시켜 적용하였다. 데이터 세트는 2개의 파트로 나뉘었다; 예측 모델의 예측 능력을 테스트하기 위해 트레이닝 세트를 만들어 모델 구축에 사용하였고, 트레이닝 세트는 테스트 세트에 사용되지 않았다. X 변수와 Y 변수간의 측정 모델을 만들기 위하여 상관관계를 예측한 후 회귀계수를 사용하여 상관관계를 계산하였다. 전체의 데이터는 임의의 두 부분으로 나누웠다: 먼저 트레이닝 세트를 선별하고 트레이닝 세트에 사용되지 않은 데이터를 테스트 세트로 사용하여 예측 모델을 구축하는데 사용되었다. PLS 모델의 예측 능력은 실측값과 PLS 예측 값 사이의 정확도를 10 bootstrap 테스트에 P-values와 RMSEP (root mean square error of prediction)로 나타냈다(Terhoeven-Urselmans et al. 2010; Liu et al. 2006).

## 7. 4. Results and Discussion

### 7. 4. 1. FT-IR 스펙트럼 데이터의 다변량통계 분석을 이용한 부모 라인에서 F1의 자손 식별

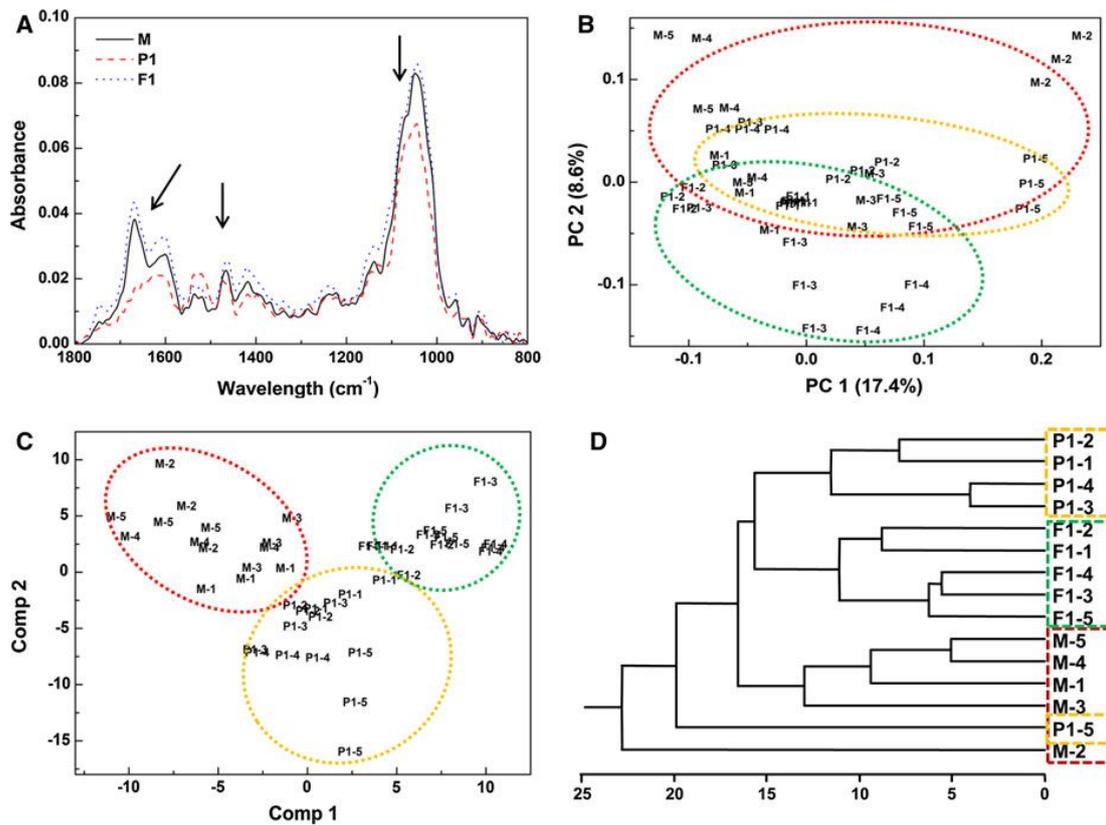
FT-IR 스펙트럼 데이터의 다변량통계 분석을 이용하여 부모 라인으로부터 F1 자손을 식별하였다(Fig. 23). *B. campestris* subsp. *pekinensis* 의 부모 교배 라인과 F1 자손 라인으로부터 genomic DNA의 중요한 FT-IR 스펙트럼을 확인하였다(Fig. 23a). 부모 라인과 F1 자손 라인에 genomic DNA의 FT-IR 스펙트럼을 보면 1,750–1,600, 1,550–1,400, 1,150–950 $\text{cm}^{-1}$  부위에서 특징적인 스펙트럼 변화를 볼 수 있었다(Fig. 23a). 많은 연구에서 FT-IR 스펙트럼에 의해 DNA의 특성을 해석하기 위한 시도가 있었다(Falk et al. 1963; Liquier et al. 1991; Brewer et al. 2002; Banyay et al. 2003; Zhou-Sun et al. 1997; Gonzalez-Ruiz et al. 2011). Nucleic acids의 IR 스펙트럼은 nucleic acid에 상호작용과 구조에 의해 4개의 부위에서 특징적으로 나타난다(Banyay et al. 2003; Gonzalez-Ruiz et al. 2011). FT-IR 스펙트럼의 1,750-1,600 $\text{cm}^{-1}$  부위는 DNA의 염기구조인 C=O, C=N 그리고 C=C 형태와  $-\text{NH}_2$  결합 형태를 나타낸다(Brewer et al. 2002; Banyay et al. 2003). 1,550-1,400 $\text{cm}^{-1}$  부위는 purine과 pyrimidine ring 모델에서 유래되었다(Falk et al. 1963; Liquier et al. 1991; Zhou-Sun et al. 1997). 그리고 1,200-1,000 $\text{cm}^{-1}$  부위는 phosphodiester-deoxyribose backbone의  $\text{PO}_2^-$  그룹에 대칭적 구조와 비대칭적 구조를 나타낸다(Falk et al. 1963; Zhou-Sun et al. 1997). 이처럼 부모 라인과 F1 자손으로부터 FT-IR 스펙트럼의 1,750-1,600, 1,550-1,400 그리고 1,150-950 $\text{cm}^{-1}$  3가지 부위에서 스펙트럼의 변화를 관찰할 수 있었다(Fig. 23a). 이 중요한 스펙트럼의 부위들은 backbone과 sugar pucker 사이의 base pairing, base stacking 그리고 glycosidic bond 회전이 민감하게 작용한다고 보고되어 있다(Brewer et al. 2002; Gonzalez-Ruiz et al. 2011). 이 결과, FT-IR 스펙트럼의 특징적 차이는 genomic DNA가 교배되더라도 부모 교배 라인(P1과 M)과 F1 자손의 고유한 genomic DNA 특성에 나타날 수 있다는 것을 의미한다. 이는 유성 생식을 통해 DNA 재조합이 일어나도 부모 라인과 F1 자손에

서 genomic DNA의 고유한 스펙트럼 변화가 발생하더라도 DNA 배열을 확인할 수 있음을 나타낸다. 따라서, FT-IR 스펙트럼 분석은 식물 종의 유성 생식을 통해 혼성 DNA의 변화를 반영하는 새로운 도구로 적용 할 수 있는 것으로 판단된다.

부모 라인(P1과 M)과 F1 자손 라인(F1)으로부터 FT-IR 데이터의 PCA score plot을 PC 1, 2를 이용하여 확인하였고, PC 1과 2 score는 전체 변이량의 각각 17.4%, 8.6%의 설명력을 갖고 있으며 이는 전체 변이량의 약 28%를 반영하고 있음을 알 수 있었다(Fig. 23b). 부모 라인(P1과 M)에 속하는 각각의 샘플은 완전히 중복되어 나타났다. 모계 라인의 단일 집단은 부계 라인보다 크게 나타났다. 또한, PCA score plot에서 각각의 동일한 라인들이 폭 넓게 집단을 구성하고 있다. 이러한 결과는 동일한 라인에서 스펙트럼 변화가 동일 라인 집단 사이에서 매우 많이 나타난 것으로 판단된다. 또한, F1 자손 라인(F1)이 부모 라인 집단과 중첩을 이루는 부분도 있지만 개별적인 집단으로 그룹화 되는 것을 관찰할 수 있었다. 이러한 결과는 PCA에서 시료간 큰 변화를 동일한 라인에서 완벽하게 식별할 수 없음을 나타냈다.

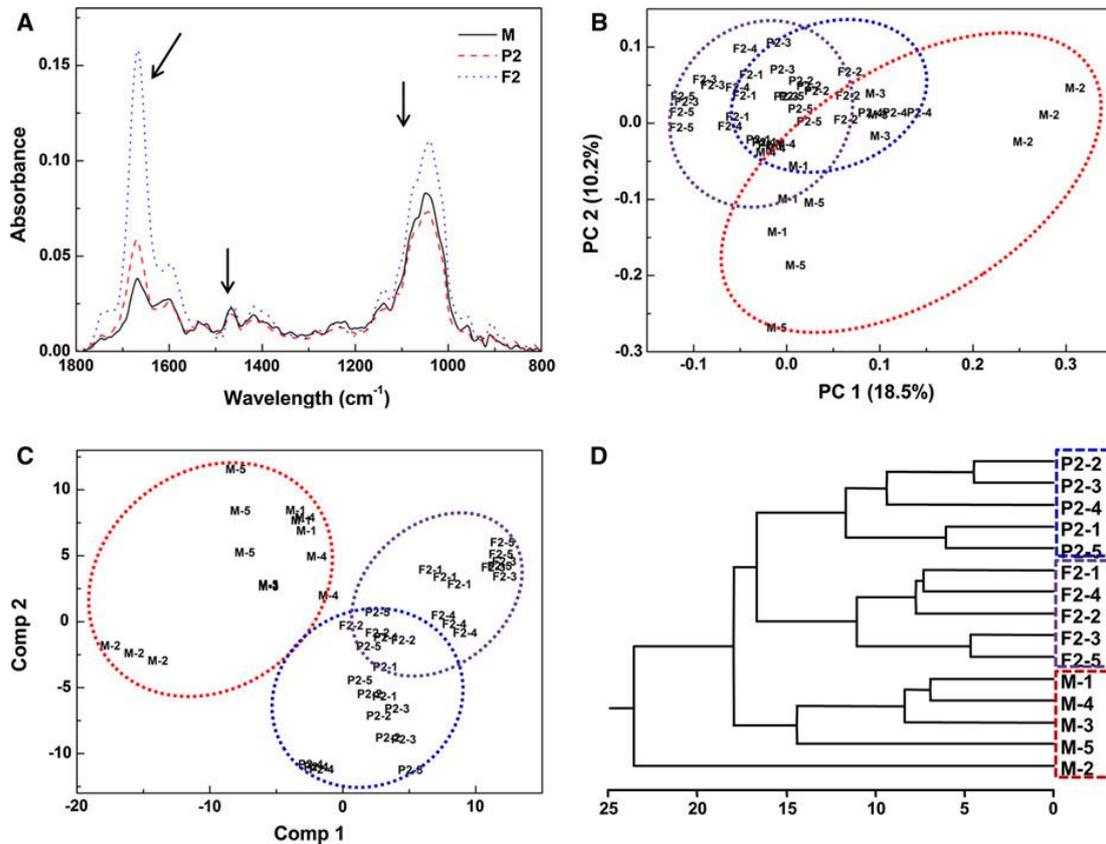
PLS-DA score plot는 PCA 분석 보다 더욱 확실하게 그룹 식별이 가능하다 (Fig. 23C). 부모 라인(P1과 M)과 F1 자손 라인(F1)은 집단 안에서 분산되어 나타났다. 모계 라인에서 F1 자손 라인이 전체적으로 분리되었다. PCA에 비해 동일한 라인의 경계가 좁아졌다. 이 결과는 각각의 모계(M), 부계(P1), 자손(F1)의 그룹 경계는 PCA에 비해 더욱 작아졌으며 동일 품종에 속하는 각 시료들의 반복구들도 각 그룹 내에 위치하여 PCA보다 식별능력이 향상되었음을 알 수 있었다(Fig. 23c).

FT-IR 스펙트럼 데이터의 PLS-DA로부터 HCA dendrogram을 확인하였다(Fig. 23d). HCA dendrogram에서 부모 라인(P1과 M)과 F1 자손 라인(F1)을 3개의 큰 분류로 나눠 나타났다. P1-5 샘플을 제외한 모든 부계 라인의 DNA 샘플이 첫 번째 분류로 존재했다. F1 자손 샘플 모두 두 번째 분류로 나타났다. M2 샘플을 제외하고 모든 모계 라인 샘플이 세 번째 분류로 나타났다. 이러한 결과는 하나의 부계 라인 샘플을 제대로 분류하지 못했지만 HCA dendrogram에서 부모 라인과 F1 자손을 식별 할 수 있음을 알 수 있었다.



**Fig. 23** Multivariate analysis of FT-IR spectral data from genomic DNA of *B. campestris* subsp. *pekinensis* parental inbred lines (P1 and M) and their F1 progeny lines (F1). A: Representative FT-IR spectra of parental inbred lines (P1 and M) and their F1 progeny lines (F1). B: PCA score plot of FT-IR spectral data. C: PLS-DA score plot of FT-IR spectral data. D: HCA dendrogram of PLS-DA score. Arrows indicate FT-IR spectra variables which showing significant variation between parental and their F1 hybrid lines (A). Dotted circles represent each boundary for group discrimination (B, C). Capitals and numbers represent sample name and replicates, respectively. M: a maternal inbred line of *B. campestris* subsp. *pekinensis*; P1: a paternal inbred line of *B. campestris* subsp. *pekinensis*; F1: F1 hybrid lines crossing with P1 and M inbred lines.

*B. campestris* subsp. *pekinensis* 의 F1 자손 식별을 적용하기 위해 또 다른 라인 genomic DNA의 FT-IR 스펙트럼을 확인하였다. 종자 산업에 대한 F1 자손 식별 시스템의 정확도를 높이기 위해, 유사한 유전적 배경을 가진 교배 라인의 FT-IR 스펙트럼 분석을 수행하였다. 같은 모계 라인(M), 서로 다른 부계 라인(P2) 그리고 자손 라인(F2)으로부터 genomic DNA의 FT-IR 스펙트럼 분석을 실시하였다. FT-IR 스펙트럼 데이터의 다변량 분석을 통해 부모 라인으로부터 F1 자손(F2)을 성공적으로 식별할 수 있었다(Fig. 24). *B. campestris* subsp. *pekinensis* 의 부모 교배 라인(P2와 M)과 F1 자손 라인(F2)으로부터 genomic DNA의 중요한 FT-IR 스펙트럼을 확인하였다(Fig. 24a). 처음 라인 실험과 유사하게 1,750–1,600, 1,550–1,400, 1,150–950cm<sup>-1</sup> 부위에서 부모 라인(P2와 M)과 F1 자손 라인(F2) 간에 특징적인 스펙트럼 변화를 볼 수 있었다(Fig. 24a). 게다가, PCA 상에서도 F1 자손 라인(F2)으로부터 부모 교배 라인(P2와 M)을 완벽하게 나눌 수 없었다(Fig. 24b). PLS-DA에서는 PCA보다 더욱 확실하게 그룹 식별이 가능하다(Fig. 24c). 모계 라인으로부터 F1 자손 라인(F2) 완벽하게 나뉘었다. 그러나, F1 자손 라인 (F2) 과 부계 라인이 약간의 중첩이 있음을 PLS-DA에서 관찰할 수 있었다. FT-IR 스펙트럼 데이터의 PLS-DA로부터 HCA dendrogram을 확인하였다(Fig. 24d). HCA dendrogram에서 부모 라인(P2와 M)과 F1 자손 라인(F2)을 3개의 큰 분류로 나눠 나타났다. 모든 부계 라인의 DNA 샘플이 첫 번째 분류로 존재했다. F1 자손 샘플(F2) 모두 두 번째 분류로 나타났다. M2 샘플을 제외하고 모든 모계 라인 샘플이 세 번째 분류로 나타났다. 이러한 결과는 HCA dendrogram에서 부모 라인 과 F1 자손을 식별 할 수 있음을 알 수 있었다. 다변량 분석과 FT-IR 스펙트럼 이 결합되어 사용된 식물의 genomic DNA를 통해 F1 자손 식별 시스템은 첫 표본이다. 두 개의 다른 라인(Fig. 23d, 24d)의 결과를 보면 FT-IR 스펙트럼 분석이 유성 생식으로 인해 변형된 고유의 DNA 특성을 가진 F1 자손 식별을 위한 새로운 도구로 적용 할 수 있는 것이 판단된다. 또한, 본 연구의 F1 자손 식별 시스템은 전체 genome 수준에서 DNA 이성종의 예비 평가와 예측에 이용될 수 있다. 하지만, 정확하게 genomic DNA(base composition, DNA backbone, secondary structure, etc.)의 특성이 F1 자손 식별에 중요한 역할을 할 것이다.



**Fig. 24** Multivariate analysis of FT-IR spectral data from genomic DNA of *B. campestris* subsp. *pekinensis* parental inbred lines (P2 and M) and their F1 progeny lines (F2). A: Representative FT-IR spectra of parental inbred lines (P2 and M) and their F1 progeny lines (F2). B: PCA score plot of FT-IR spectral data. C: PLS-DA score plot of FT-IR spectral data. D: HCA dendrogram of PLS-DA score. Arrows indicate FT-IR spectra variables which showing significant variation between parental and their F1 hybrid lines (A). Dotted circles represent each boundary for group discrimination (B, C). Capitals and numbers represent sample name and replicates, respectively. M: a maternal inbred line of *B. campestris* subsp. *pekinensis*; P2: a paternal inbred line of *B. campestris* subsp. *pekinensis*; F2: F1 hybrid lines crossing with M and P2 inbred lines.

#### 7. 4. 2. FT-IR 스펙트럼 데이터의 다변량 분석을 통한 두 개의 부모 라인으로부터 F1 자손 식별

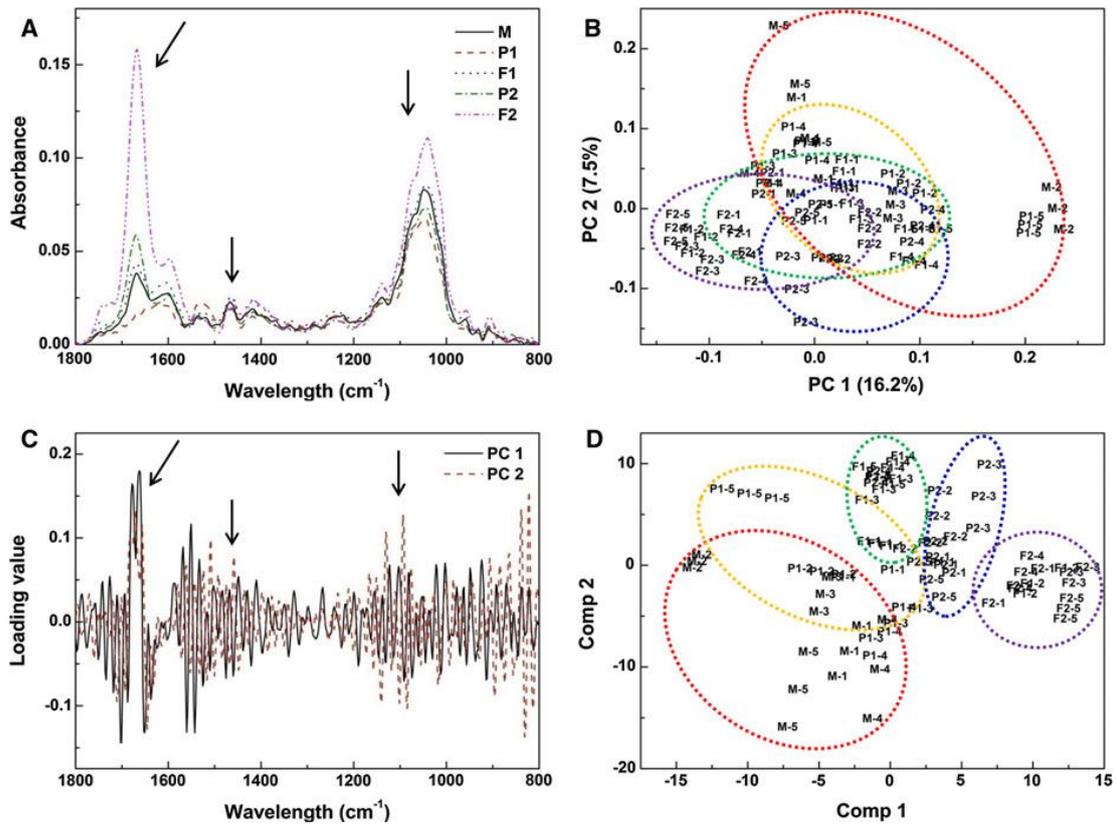
동일한 모계와 서로 다른 부계 사이에서 두 개의 서로 다른 F1 자손 라인을 식별 할 수 있었다(Fig. 25). *B. campestris* subsp. *pekinensis* 의 부모 교배 라인 (P1, P2, M)과 F1 자손 라인(F1과 F2)으로부터 genomic DNA의 중요한 FT-IR 스펙트럼을 확인하였다(Fig. 25a). 부모 라인과 F1 자손 라인에 genomic DNA의 FT-IR 스펙트럼을 보면 1,750–1,600, 1,550–1,400, 1,150–950 $\text{cm}^{-1}$  부위에서 특징적인 스펙트럼 변화를 볼 수 있었다(Fig. 25a). 매우 중요한 IR 스펙트럼의 변화가 단일 라인 실험의 결과와 매우 유사하게 나타났다. 부모 교배 라인(P1, P2, M)과 F1 자손 라인(F1, F2)이 PCA score plot 상에서 상당한 부분이 서로 중첩되어 나타났다(Fig. 25b). 이러한 결과는 F1 자손 라인(F1, F2)으로부터 부모 교배 라인 (P1, P2, M)이 완전하게 분리되지 못하고 비슷하게 나타났다(Fig. 25b).

PC 1, 2의 loading values를 실행하여 F1 자손 라인(F1, F2)으로부터 부모 교배 라인(P1, P2, M)의 식별에 중요한 IR 스펙트럼 부위를 관찰하였다. 부모 교배 라인(P1, P2, M)과 F1 자손 라인(F1과 F2)을 식별하는 PC 1, 2 scores의 가장 중요한 차이를 FT-IR 스펙트럼 변수에서 관찰하였다(Fig. 25c). PC 1, 2 loading values는 FT-IR 스펙트럼의 1,750–1600, 1,550–1,400, 1,200–1,000 $\text{cm}^{-1}$ 에서 중요한 차이를 관찰할 수 있었다(Fig. 25c). 이 결과는 특정 FT-IR 스펙트럼 차이와 동일하게 나타났다(Fig. 25a). Genomic DNA의 FT-IR 스펙트럼 분석에 관한 이전 보고에 따르면, thymine와 cytosine의 carbonyl stretching vibration 흡수되어 1,692 $\text{cm}^{-1}$  영역에 중첩되어 나타난다고 보고되었다(Liquier et al. 1991; Geinguenaud et al. 2000; Lindqvist et al. 2000). 또한, thymine, guanine, C=N and C=C vibrations의 carbonyl stretching이 1,650 $\text{cm}^{-1}$  영역에서 나타나고, NH<sub>2</sub> scissoring과 N–H bending이 대체적으로 1,605–1,560 $\text{cm}^{-1}$  영역에서 각각 나타난다(Zhou-Sun et al. 1997; Brewer et al. 2002). 추가적으로 purine와 pyrimidine (DNA bases) ring 형태의 파장이 흡수되어 1,466–1,488 $\text{cm}^{-1}$  영역에서 나타난다(Lindqvist et al. 2000). IR peak에 대한 연구들을 고려해보면, FT-IR 스펙트럼의 1,750–1,600와 1,550–1,400 $\text{cm}^{-1}$  영역에 중요한 스펙트럼 변화는 기본 구조의 변

형과 기본 조성의 조합에 의해 유래된 것을 알 수 있다. Lindqvist et al. (2000) 보고에 따르면, phosphate backbone 진동으로 인한 흡수 파장 영역인  $1,250-1,200\text{cm}^{-1}$  영역에서는  $\text{PO}_2^-$ 의 비대칭(ASM) 운동 구조가 나타나고,  $1,090-1,070\text{cm}^{-1}$  영역에서는  $\text{PO}_2^-$ 의 대칭(SM) 운동 구조가 나타난다.  $1,145\text{cm}^{-1}$  IR 흡수 파장 영역은 C-O-P stretching 진동이 나타나다(Dovbeshko et al. 2000; Gault et al. 2004). 이러한 결과를 고려할 때,  $1,200-1,000\text{cm}^{-1}$ 에서 중요한 스펙트럼의 변화는 DNA phosphate backbone의 형태 변화가 직접적으로 관련이 있을 것으로 판단된다.

Muntean et al. (2009)의 보고에서도  $1,000-1,200\text{cm}^{-1}$ 에서 스펙트럼 차이의 유사성은 고등 식물의 분류학적 관계를 가지고 있음을 보고하였다. 현재까지의 연구에서는  $1,750-1,600$ ,  $1,550-1,400$ ,  $1,200-1,000\text{cm}^{-1}$  영역은 *B. campestris* subsp. *pekinensis*의 F1 자손을 식별하는 중요한 영역임을 확인하였다. 그 영역들은 DNA의 기본 구조와 phosphodiester backbone을 나타낸다. 따라서, genomic DNA의 구조와 구성의 변화는 *B. campestris* subsp. *pekinensis*의 F1 자손에 전반적인 식별에 대해 중요하게 작용함을 알 수 있었다.

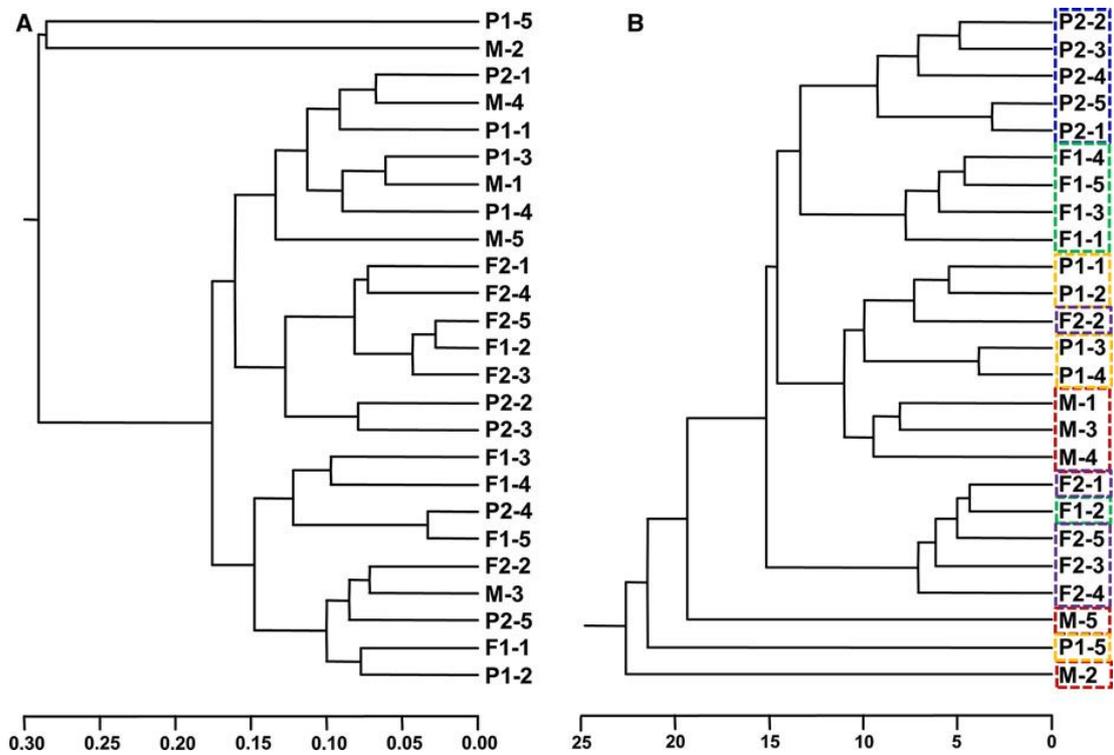
PLS-DA score plot는 PCA 분석 보다 더욱 확실하게 그룹 식별이 가능하다 (Fig. 25d). 각각의 부모 라인(P2와 M)과 F1 자손(F1과 F2)이 모계(M)와 자손(F1)에 중첩된 부계 라인(P1)을 제외하면 각각이 그룹화 되어 분포하였다. 두 F1 자손 라인(F1과 F2)은 모계 라인으로부터 완전히 분리되었다. PCA와 비교하면, 각각의 라인이 바깥쪽으로 그룹을 형성하였다. 이러한 결과는 같은 라인에 속하는 각각의 샘플이 PCA에서 보다 더 밀접하게 개별적인 그룹화를 나타냈고, PCA와 비교해 F1 자손 라인(F1과 F2)으로부터 부모 라인(P1, P2, M)을 식별하는데 있어서 PLSDA가 더 나은 것으로 나타났다(Fig. 25d).



**Fig. 25** Multivariate analysis of FT-IR spectral data on genomic DNA from *B. campestris* subsp. *pekinensis* parental inbred lines (P1, P2, and M) and their F1 progeny lines (F1 and F2). a Representative FT-IR spectra of each line. b PCA score plot of FT-IR spectral data. c Loading values of PC 1 and PC 2. d PLSDA score plot of FT-IR spectral data. Arrows indicate FT-IR spectra variables showing significant variation between parental and F1 progeny lines (a, c). Dotted circles represent each boundary for group discrimination (b, d). Capitals and numbers represent sample names and their replicates, respectively. M: a maternal inbred line of *B. campestris* subsp. *pekinensis*; P1 and P2: each paternal inbred line of *B. campestris* subsp. *pekinensis* F1 and F2: F1 hybrid lines crossed with P1, P2, and M inbred lines.

FT-IR 스펙트럼 데이터의 PCA(Fig. 26a)와 PLS-DA(Fig. 26b)로부터 HCA dendrograms을 나타냈다(Fig. 26) PCA의 HCA dendrogram으로부터 패턴 분류는 그룹화 되어 나타나지 않았다. 이 결과에서는 PCA로부터 HCA dendrogram은 모든 부모 라인으로부터 F1 자손 라인 식별이 불가능한 것으로 나타났다(Fig. 26a). 하지만, PLS-DA로부터 HCA dendrogram은 비록 F1-2, F2-2, P1-5, M1, M5 시료들이 잘못 분류되었지만, 각각의 패턴별로 분류되었다(Fig. 26b). 전체의 5개 샘플이 잘못 분류가 되었고, classification 오류율이 20%로 나타났다. HCA dendrogram에서 몇 개의 잘못 분류된 샘플들을 고려하지 않는다면, 부모 라인(P1, P2, M)과 F1 자손 라인(F1과 F2)을 3개의 큰 분류로 나눠 나타났다. 부계 라인(P2)과 자손 라인(F1)의 DNA 샘플들이 첫 번째 분류에 나타났다.

이러한 결과는 HCA dendrogram에서 부모 라인과 F1 자손을 식별 할 수 있음을 알 수 있었다. 첫 번째 분류에서는 모든 P2 샘플과 4개의 F1 샘플로 크게 두 개의 분류로 나뉘어 나타났다. 첫 번째 분류와 비슷하게 두 번째 분류에서도 4개의 P1 샘플과 3개의 M 샘플로 크게 두 개의 분류로 나뉘어 나타났다. 세 번째 분류에서는 모든 F2 자손 샘플들이 분류되었다. 이 결과들은 PLS-DA의 HCA dendrogram은 80% classification 정확도로 부모 라인으로부터 두 개의 F1 자손 라인 식별이 가능하였다. 단일 조합(Figs. 23d, 24d)의 Classification 정확도는 거의 완벽하였지만, 두 개의 다른 부모 조합에서는 정확도가 80%로 감소하였다(Fig. 26b). 이러한 결과는 정확도가 부모의 조합에 따라 달라질 수 있다는 것을 나타낸다.



**Fig. 26** HCA dendrograms from PCA (a) and PLS-DA (b) of *B. campestris* subsp. *pekinensis* parental inbred lines (P1, P2, and M) and their F1 progeny lines (F1 and F2). Capitals and numbers represent sample names and their replicates, respectively. M: a maternal inbred line of *B. campestris* subsp. *pekinensis* P1 and P2: each paternal inbred line of *B. campestris* subsp. *pekinensis* F1 and F2: F1 hybrid lines crossed with P1, P2, and M inbred lines.

#### 7. 4. 3. *B. campestris* subsp. *pekinensis*의 서로 다른 두 개의 부모 조합으로부터 F1 자손의 Classification, prediction, cross-validation

부모 조합으로부터 F1 자손의 PLS-DA 식별이 실제 적용 가능성을 확인하기 위해 bootstrap test를 실시하였다(Table 4). Bootstrap test의 정확도는 92% 이었고, P 값(P=0)은 통계적으로 유의하였다(Table 4). 이러한 결과는 명확하게 교배종에 대한 조합이 PLS 모델링을 통해 높은 정확도로 구별 할 수 있다는 것을 나타냈다. 본 연구에서는 전체 9개의 F1 자손 라인 스펙트럼 중에서 2개의 스펙트럼을 정확하게 예측하지 못했다. 그러나, 다른 스펙트럼은 정확하게 예측하고 자신의 그룹으로 분류 되었다. 많은 실험 요인이 FT-IR 스펙트럼의 재현성에 영향을 미칠 수 있다. Mello and Vidal (2012) 보고에 따르면, 특정한 nucleo base 그룹의 상대적으로 강렬한 진동이 비록 변하기 쉬운 구조의 DNA이지만 다른 FT-IR 특성을 일으켜 polypeptide-DNA 복합체가 나타내는 정보를 막을 수 있다 (Mello and Vidal 2012). 따라서 F1 자손의 예측 정확도는 샘플이 증가되면 다양한 통계 연산을 사용하여 FT-IR 스펙트럼의 재현성을 높이며 예측 정확도를 높일 수 있다.

본 연구에서는 전체 genome 수준에서 유사성을 관찰할 수 없었다. Genomic DNA의 유사한 스펙트럼 차이는 고등식물에서 분류학적 관계를 나타낼 수 있었다(Muntean et al. 2009). 따라서, FT-IR 스펙트럼의 유사성을 기준으로 F1 자손의 식별은 genomic의 유사성으로 설명 될 수 있다. 결론적으로, FT-IR이 성공적으로 자신의 교배조합에서 교배종을 식별하고 선택하는데 사용될 수 있을 것이다. 또한, 이 기술은 교배육종에 빠르고 정확한 선택을 위해 이용될 수 있을 것이다.

**Table 4.** Summary of the PLS-DA classification results from FT-IR spectral data of genomic DNA. The discriminant functions were determined using the first seven principal components determined by PCA. In bootstrapping, each case was classified by the functions derived from all other cases. P1 and P2: each paternal inbred line of *B. campestris* subsp. *pekinensis*; F1 and F2: F1 hybrid lines crossing with P1, P2 and M inbred lines, respectively.

		Prediction					Total	
		M	P1	F1	P2	F2		
Cross validate d	Count	M	5				5	
		P1		5			5	
		F1			4	1	5	
		P2				5	5	
		F2				1	4	5
Total			5	5	4	6	4	25

$R^2=0.958$

$P < 0.05$

Probability =  $1.8E \times 10^{-14}$

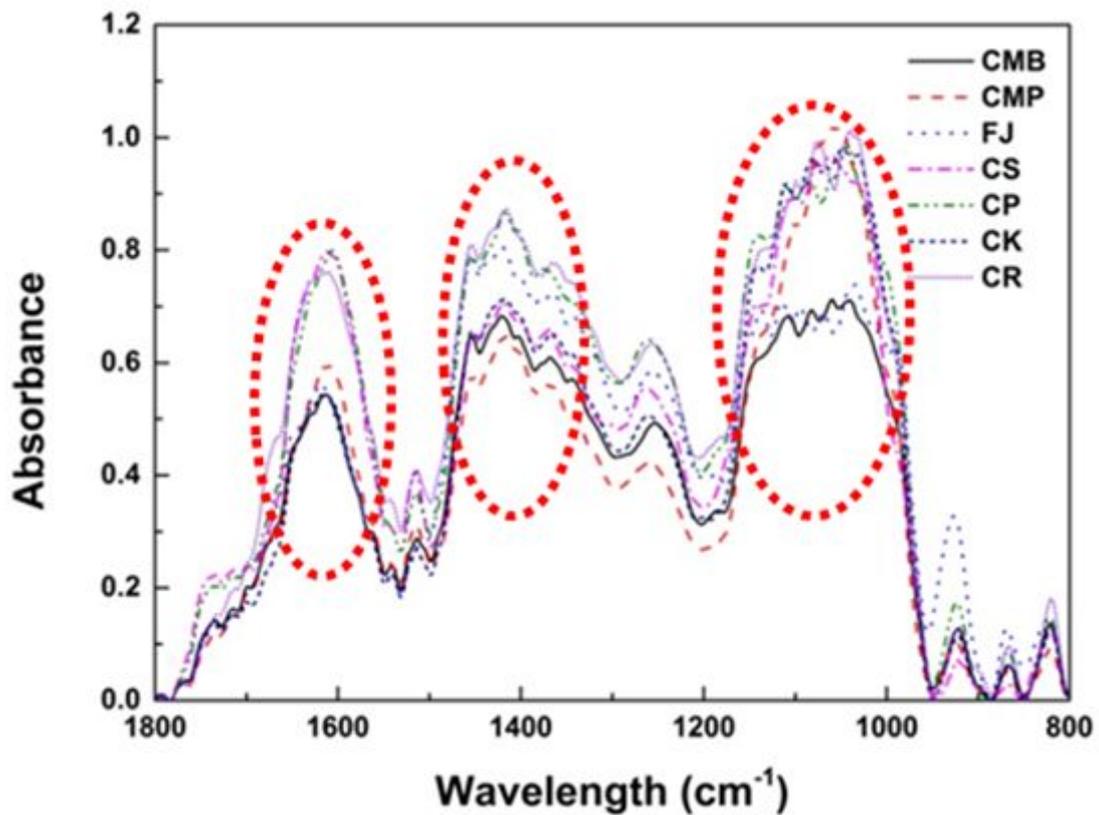
## 8. Discussion

본 연구는 적외선 분광분석 데이터의 다변량 통계분석 기법을 활용하여 다양한 식물종의 신속하고 간편한 품종 식별 시스템을 확립하고, 더 나아가 기능성 성분의 함량이 우수한 라인을 비교적 간편한 방법으로 식별할 수 있는 선발체계를 확립함으로써 우수 품종 육성의 기간을 단축하여 품종 개발을 더욱 가속화시키기 위한 수단으로 활용 체계를 확립하기 위함이었다. 식물이나 작물의 품종을 식별하기 위해서는 RAPD (random amplification of polymorphic DNA), RFLP (restriction fragment length polymorphism), SSR (simple sequence repeat), AFLP (amplified fragment length polymorphism)와 같은 유전자지문분석방법 즉 분자마커를 이용한 식별 방법들이 개발되어 왔지만, 이들 방법은 대상 작물에 대한 유전적 정보가 제공되어야 한다는 점과 분석 비용이 많이 들어가고, 장시간 분석을 해야 하는 단점들이 있다. 이런 유전자지문분석방법들의 대안으로 대사체 수준에서 지문분석 방법이 대두되고 있다. 본 연구에 사용한 적외선분광분석 (FT-IR)은 매우 간단하게 식물체에 존재하는 대사산물을 추출한 다음 빠른 속도로 분석을 할 수 있는 장점을 가지고 있다. 또한 적외선 스펙트럼의 부위들은 각각의 대사물질이나 핵산 성분들에 대한 특이적인 화학적 정보를 포함하고 있다. 이런 정보들을 기반으로 다변량통계분석 기법을 활용하여 식물의 각 품종간의 스펙트럼 차이를 분석하고 주요성분이나 유전적인 정보 차이를 예측할 수 있다.

Schulz and Baranska (2007)에 의하면 FT-IR 스펙트럼에 반영된 peak 정보들을 proteins과 amino acids, Lipids와 fatty acids, Carbohydrates로 나눌 수 있다. 식물 조직에서 amino acids과 Proteins이 IR 스펙트럼 해석 유용하게 작용하는데, 특히 CONH 그룹과 amide A, amide B, amide I-VII는 연관성이 높다. Amide I은  $1,680-1,600\text{cm}^{-1}$ 에서 측정이 되는데 주로 C=O의 진동이 나타나고, amide II와 amide III은 각각  $1,580-1,480\text{cm}^{-1}$ ,  $1,300-1,230\text{cm}^{-1}$ 에서 peptide 그룹의 C-N와 N-H 진동으로 나타난다. Lipids와 fatty acids는  $700-1,500\text{cm}^{-1}$  부위에서 측정이 되는데 주로 포화 fatty acid와 단일 불포화 fatty acid acyl 그룹이 측정된다.  $1,711\text{cm}^{-1}$  부위는 작은 분자의 자유 fatty acid이 나타나고,  $1,660-1,670\text{cm}^{-1}$  부

위에서는 단일 이성질체 성분과 트랜스 이성질체 성분이 측정된다. 미세한 C=C 진동 운동의 파장( $1,600\text{cm}^{-1}$ )과  $\text{CH}_2$ 의 절단 방식( $1,444\text{cm}^{-1}$ )을 이용하여 triglycerol의 iodine values과 불포화 식물성 기름을 예측할 수 있다고 나타났다. IR 스펙트럼은 acylglycerol 분자의 배열 측정에 활용될 수 있다. FT-IR 스펙트럼은 mono-, di-, oligo-의 polysaccharides에 대한 특이적인 파장을 식별하기 위해 사용되었다. Sucrose의 스펙트럼은  $847\text{cm}^{-1}$ ,  $868\text{cm}^{-1}$ ,  $898\text{cm}^{-1}$ ,  $885\text{cm}^{-1}$  부위에서 각각 aglucose,  $\alpha$ -fructose,  $\alpha$ -glucose과  $\beta$ -glucose,  $\beta$ -anomer들이 특징적으로 나타난다고 보고되었다. 따라서 FT-IR 스펙트럼을 이용한 품종 식별이 가능할 것으로 판단된다(Schulz and Baranska 2007).

본 연구에서 적외선 분광분석 데이터의 다변량 통계분석 기법을 활용하여 두과작물의 품종간 식별, 배추 교배종의 F1 식별, yam에 존재하는 기능성 성분, 감귤의 carotenoid 계열 기능성 성분 함량 예측이 가능함을 확인하였다. 또한 두과작물에서 적용하였던 방법들을 기반으로 감귤의 품종식별이 가능함을 조사하였다(Fig. 27). 감귤은 제주도내에서 가장 많이 재배되며 시중에 판매되고 있는 7종의 감귤을 사용하였다. 실험에 사용된 감귤은 비가림 재배 궁천조생(*Citrus unshiu* Marc. cv. Miyagawa (under rain shelter)), 노지 재배 궁천조생(*C. unshiu* Marc. cv. Miyagawa (on the field)), 금귤(*Fortunella japonica*), 세토카(*C. hybrid* 'Setoka'), 감평(*C. hybrid* 'Kanpei'), 청견(*C. hybrid* 'Kiyomi') 그리고 부지화(*C. hybrid* 'Shiranuhi')이며 각 종별로 3개를 대사체 분석에 사용하였다. FT-IR 스펙트럼상의  $1,500\text{-}1,700$ ,  $1,300\text{-}1,500$ ,  $950\text{-}1,100\text{cm}^{-1}$  부위에서 대사체의 양적, 질적 패턴 변화가 두과작물과 유사하게 나타났다(Fig. 3, 27). 앞서 설명한 것과 같이 스펙트럼이 나타내는 부위는 각각 amide I 과 II을 포함하는 아미노산 및 단백질 계열의 화합물, phosphodiester group을 포함한 핵산 및 인지질 그리고 단당류나 복합 다당류를 포함하는 carbohydrates 계열의 화합물들의 질적, 양적 정보를 반영하는 부위이다(Parker 1983; Dumas and Miller 2003; Wolkers et al. 2004; Yee et al. 2004; D'Souza et al. 2008; Lopez-Sanchez et al. 2010). 이러한 물질들이 품종간에 차이를 뚜렷하게 나타냄을 알 수 있었다. 따라서 FT-IR 스펙트럼 분석은 감귤의 품종별 주요 대사체의 질적, 양적 변화를 예측하고 함량이 높은 계통의 선발 수단으로 신속한 활용이 가능할 것으로 기대된다.



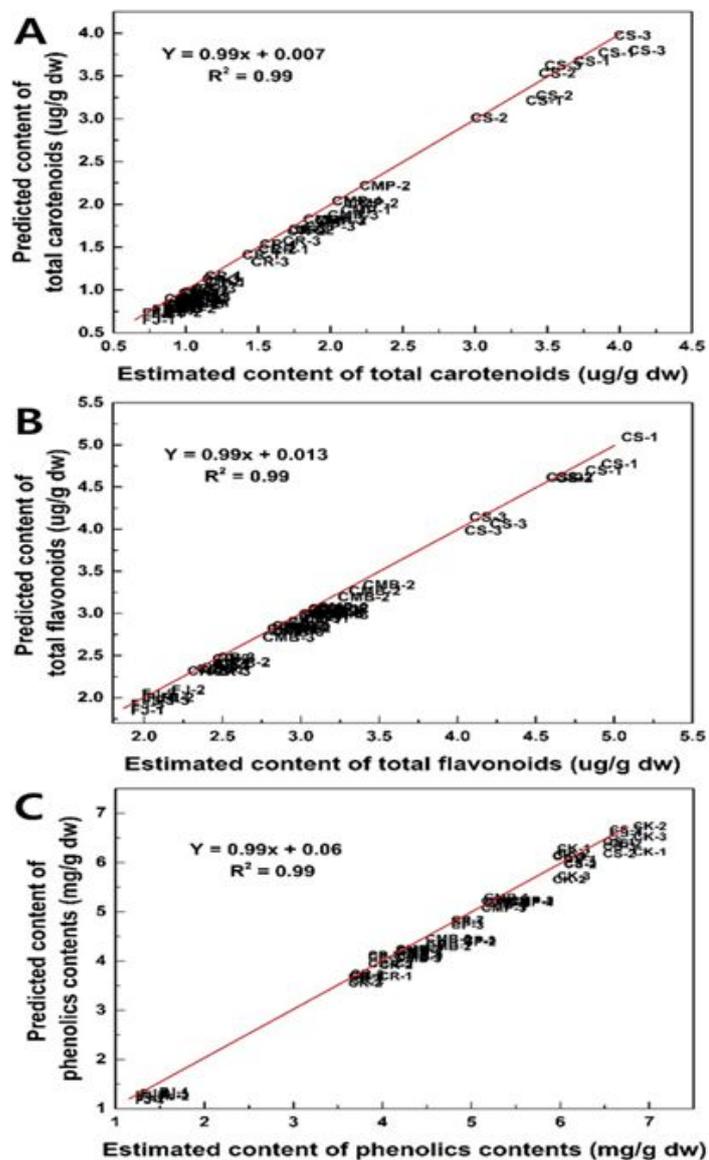
**Fig. 27** Representative FT-IR spectra from Citrus fruits. (A) CMB : *Citrus unshiu* Marc. cv. Miyagawa (under rain shelter). (B) CMP : *Citrus unshiu* Marc. cv. Miyagawa (the bare ground). (C) FJ : *Fortunella japonica*. (D) CS : *C. hybrid* 'Setoka'. (E) CP : *C. hybrid* 'Kanpei'. (F) CK : *C. hybrid* 'Kiyomi'. (G) CR : *C. hybrid* 'Shiranuhi'.

아프리카 양에서 확립한 기능성 성분 함량 예측 PLS 모델 방법을 감귤에 적용하여 감귤에서도 유사한 성분 함량 예측 체계를 확립하였다. 먼저 감귤에 존재하는 total carotenoid, flavonoid, phenolics compounds 함량을 측정하였다(Table 5). 그 결과 HPLC로 분석한 carotenoid 함량과는 차이가 있지만(Fig. 15), HPLC 분석보다 빠르고, 저 비용으로 분석이 가능했다. FT-IR 스펙트럼 데이터와 기능성 성분 분석 데이터를 이용하여 PLS 모델링을 실시하였다(Fig. 28). Total carotenoids의 경우, PLS 모델링을 통해 FT-IR 스펙트럼 데이터로부터 예측된 함량 값과 동일 시료의 실측 함량 값을 이용하여 회귀분석을 수행한 결과 상관계수가  $R^2 = 0.99$ 로 높게 나타났다(Fig. 28A). Total flavonoids와 total phenolics의 함량 예측 모델링을 통한 선형회귀분석 상관계수는 각각  $R^2 = 0.99$ (Fig. 28B) 및  $0.99$ (Fig. 28C)로 나타났다. 이들 결과는 감귤 시료의 FT-IR 스펙트럼 데이터만으로도 이들 시료 내에 존재하는 total carotenoids, flavonoids 및 phenolics의 함량을 90%이상의 정확성으로 함량을 예측할 수 있는 결과라 사료된다. 이 결과들은, 아프리카 양(Fig. 11)보다 높은 예측 율을 보였으며, HPLC 분석을 이용한 carotenoid 함량 예측 모델(Fig. 17)보다 높게 나타났다. 하지만 HPLC 분석은 다양한 carotenoid 성분을 분석한 것이고 전체적인 carotenoid 성분을 분석한 것과는 확인이 다를 수 있다. 따라서, 분석 시료가 증가하고 다양한 다변량통계 기법을 적용 한다면 예측 정확도는 높아 질 수 있을 것으로 판단된다. 아울러 본 연구에서 확립된 감귤의 기능성 성분 함량 예측 모델을 이용하여 모델의 정확도 검증 연구를 통하여 감귤의 기능성 성분 예측 수단으로 활용이 가능할 것으로 기대된다. 이상의 결과를 볼 때, 감귤의 carotenoid 함량 예측 모델은 타 작물에 적용이 가능하고, 육종의 직접적인 소재로 활용이 가능하며 기능성 품종개발 가속화에 기여 할 수 있을 것으로 예상된다.

**Table 1.** Quantitative analysis of total carotenoids, flavonoids and phenolics contents from *citrus* using UV-VIS spectrophotometric analysis. All samples were run in nine replicates. CMB : *Citrus unshiu* Marc. cv. Miyagawa (under rain shelter), CBP : *C. unshiu* Marc. cv. Miyagawa (on the field), FJ : *Fortunella japonica*, CS : *C. hybrid* 'Setoka', CP : *C. hybrid* 'Kanpei', CK : *C. hybrid* 'Kiyomi', CR : *C. hybrid* 'Shiranuhi'.

Species	line numbers	Compound contents		
		Total carotenoids (ug·g <sup>-1</sup> dry wt)	Total flavonoids (ug·g <sup>-1</sup> dry wt)	phenolics compounds (mg·g <sup>-1</sup> dry wt)
FJ	9	0.75±0.09	1.97±0.09	1.26±0.12
CK	9	0.90±0.06	2.85±0.07	6.16±0.44
CMP	9	1.93±0.13	3.02±0.05	5.16±0.15
CMB	9	0.85±0.04	3.01±0.23	4.20±0.17
CP	9	1.02±0.07	2.73±0.24	4.44±0.50
CS	9	3.53±0.34	4.50±0.36	6.29±0.24
CR	9	1.54±0.13	2.35±0.06	3.67±0.17

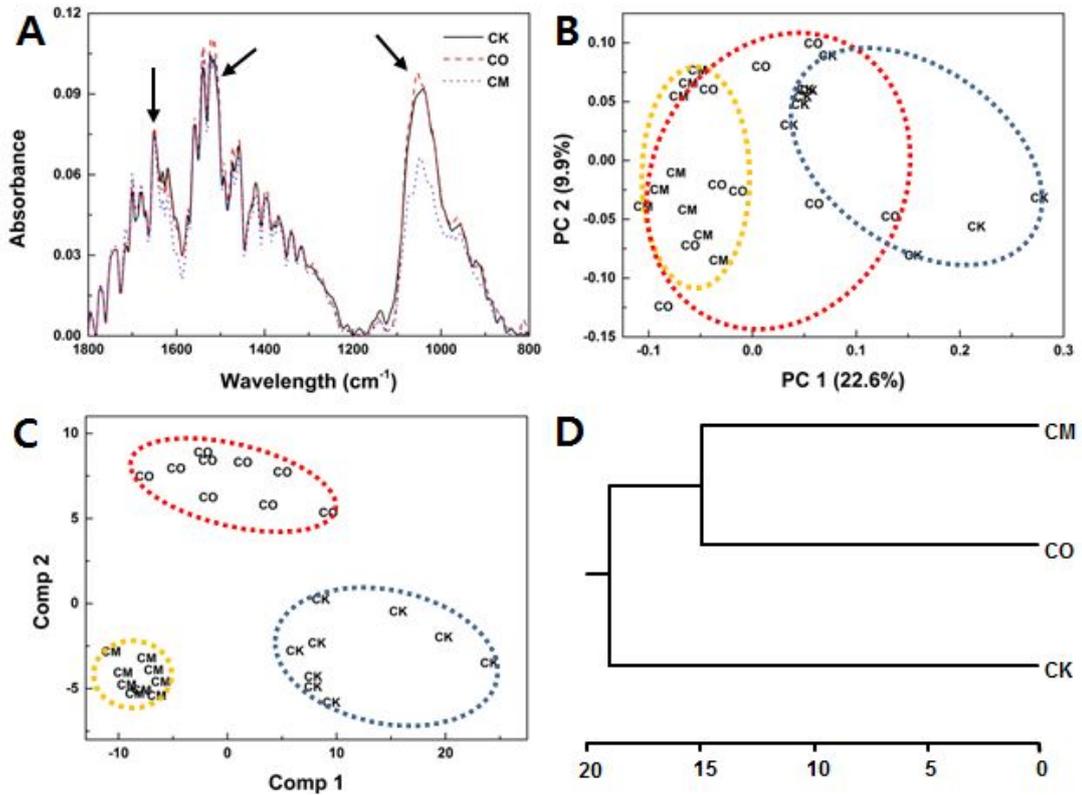
Data represent the mean±SD of measurements.



**Fig. 28** Linear regression analysis between estimated and predicted values of total carotenoids (A), flavonoids (B), and phenolics contents (C) by the PLS regression model from FT-IR spectral data. Regression coefficient values ( $R^2$ ) are 0.99, respectively.

*B. campestris* subsp. *pekinensis* 의 부모 교배 라인과 F1 자손 라인으로부터 genomic DNA를 이용한 FT-IR 스펙트럼 식별이 감귤에도 적용 가능성을 확인하였다. 6 품종과 교배 육종에 사용된 4품종의 감귤인 궁천조생(*Citrus unshiu* Marc. cv. Miyagawa), 금귤(*Fortunella japonica*), 세토카(*C. hybrid* 'Setoka'), 감평(*C. hybrid* 'Kanpei'), 청견(*C. hybrid* 'Kiyomi'), 부지화(*C. hybrid* 'Shiranuhi'), 폰칸(*Citrus reticulata* 'Ponkan'), 베니마돈나(*C. hybrid* 'Ehimo Kashi 28 gou'), 서지향(*C. hybrid* 'Nishinokaori') 그리고 네이블오렌지(*Citrus sinensis* (L.) Osbeck)이며 각 품종별로 3개를 분석에 사용하였다. 각 감귤품종인 청견, 부지화, 감평 DNA 시료의 FT-IR 스펙트럼을 비교한 결과 감귤 DNA 시료들은 FT-IR 스펙트럼상의 1,750-1,600, 1,550-1,400, 1,150-950cm<sup>-1</sup> 부위에서 DNA의 양적, 질적 패턴 변화가 나타났고(Fig. 29A, 30A, 31A), 이는 *B. campestris* subsp. *pekinensis* 의 부모 교배 라인과 F1 자손 라인 비슷하게 나타남을 알 수 있었다(Fig. 23A, 24A). 이 중요한 스펙트럼의 부위들은 backbone과 sugar pucker 사이의 base pairing, base stacking 그리고 glycosidic bond 회전이 민감하게 작용한다고 보고되어 있다(Brewer et al. 2002; Gonzalez-Ruiz et al. 2011). 이 결과, FT-IR 스펙트럼의 특징적 차이는 genomic DNA가 교배되더라도 교배 라인과 자손의 고유한 genomic DNA 특성에 나타날 수 있다는 것을 의미한다. 이는 유성 생식을 통해 DNA 재조합이 일어나도 교배 라인과 자손에서 genomic DNA의 고유한 스펙트럼 변화가 발생하더라도 DNA 배열을 확인할 수 있음을 나타낸다. 따라서, FT-IR 스펙트럼 분석은 식물 종의 유성 생식을 통해 혼성 DNA의 변화를 조사할 수 있는 새로운 도구로 적용 할 수 있는 것으로 판단된다.

감귤 genomic DNA의 적외선 분광분석 데이터의 PCA 분석을 통해 교배종과 F1종간에 상관관계를 분석했다. 청견의 경우, 네이블오렌지와 중첩이 가장 크게 나타났으며, 궁천조생과는 중첩이 이뤄지지 않았다(Fig. 29B). 이는 청견의 형질이 궁천조생보다 네이블오렌지에 가까운 형질을 보이는 것으로 판단된다. 부지화는 청견과 폰칸의 중간 위치에 그룹을 이루면서 별도의 그룹을 형성하고 있다(Fig. 30B). 부지화는 청견의 형질과 폰칸의 형질을 동시에 나타내고 있는 것으로 판단된다. 또한, 감평은 부지화와 많은 중첩을 나타냈고, 서지향은 오른쪽에 별도의 그룹을 형성하고 있다(Fig. 31B). 감평의 그룹이 부지화의 그룹과 중첩되고 있

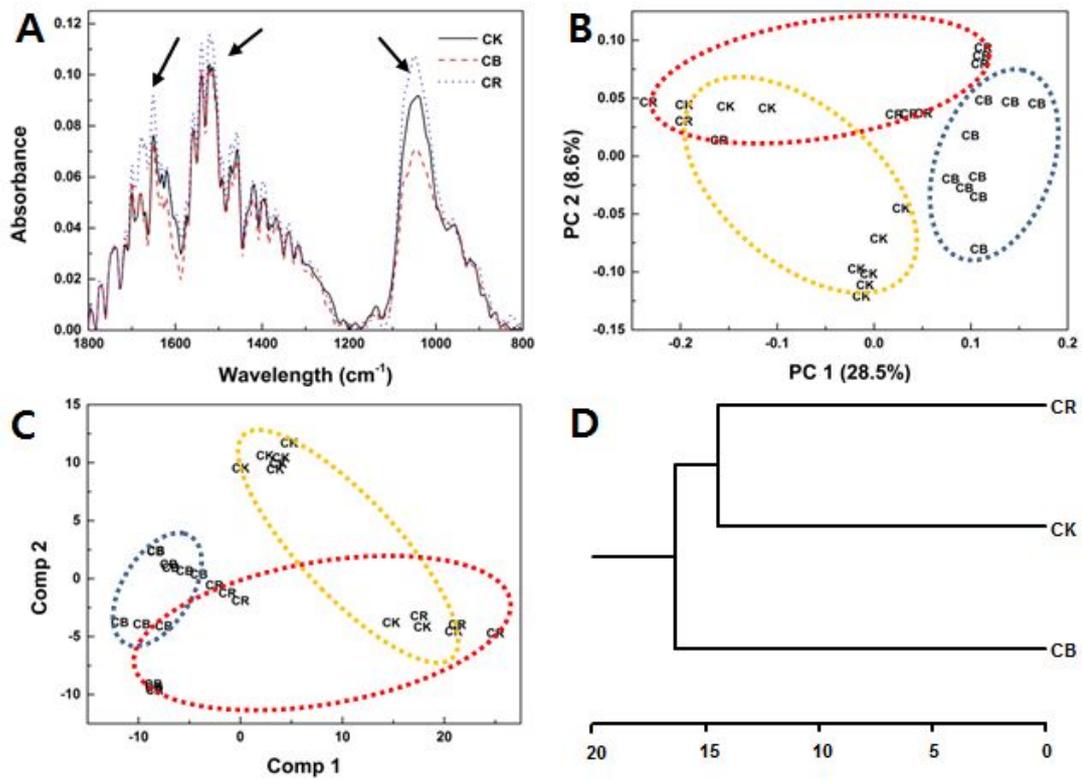


**Fig. 29** Multivariate analysis of FT-IR spectral data on genomic DNA from *C. hybrid* 'Kiyomi' cross breeding lines (CM and CO) and their progeny lines (CK). A: Representative FT-IR spectra of cross breeding lines (CM and CO) and their progeny lines (CK). B: PCA score plot of FT-IR spectral data. C: PLS-DA score plot of FT-IR spectral data. D: HCA dendrogram of PLS-DA score. Arrows indicate FT-IR spectra variables showing significant variation between cross breeding lines and hybrid lines (A). Dotted circles represent each boundary for group discrimination (B, C). Capitals and numbers represent sample names and replicates, respectively. CK : *C. hybrid* 'Kiyomi', CO : *Citrus sinensis* (L.) Osbeck, CM : *Citrus unshiu* Marc. cv. Miyagawa.

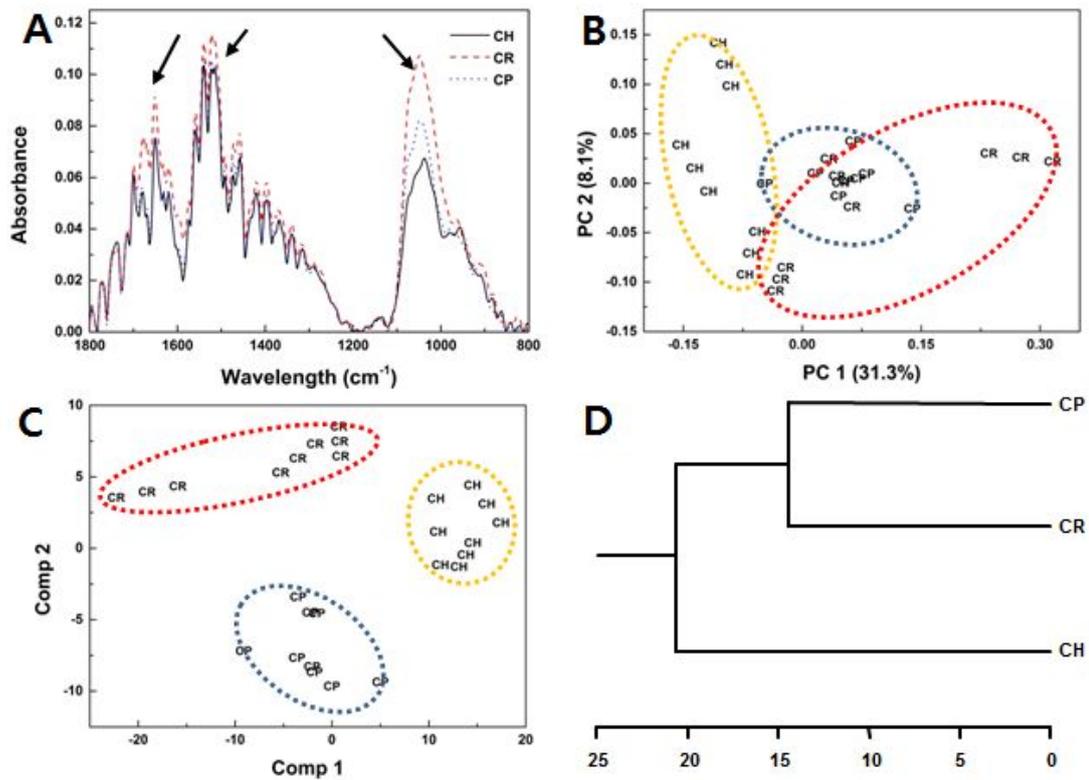
으나, 서지향 그룹과 부지화 그룹 사이에 분포하며 그룹을 형성하고 있어 두 품종의 형질을 동시에 나타내는 것으로 판단된다. 이러한 결과는 동일한 라인에서 스펙트럼 변화가 동일 라인 집단 사이에서 매우 많이 나타난 것으로 *B. campestris* subsp. *pekinensis* 의 부모 교배 라인과 F1 자손 라인에 확인되었다 (Fig. 23B, 24B). 본 연구에서는 1,750–1,600, 1,550–1,400, 1,200–1,000 $\text{cm}^{-1}$  영역은 *B. campestris* subsp. *pekinensis*의 F1 자손을 식별과 마찬가지로 감귤에서도 중요한 영역임을 확인하였다. 그 영역들은 DNA의 기본 구조와 phosphodiester backbone을 나타낸다. 따라서, genomic DNA의 구조와 구성의 변화는 *B. campestris* subsp. *pekinensis*의 F1 자손에서와 동일하게 감귤 교배종에서도 전반적인 식별에 대해 중요하게 작용함을 알 수 있었다.

PLS-DA 분석은 *B. campestris* subsp. *pekinensis* 의 분석(Fig. 23C, 24C)과 동일하게 PCA 분석 보다 더욱 확실하게 감귤 시료들의 그룹 식별이 가능하였다(Fig. 29C, 30C, 31C). 각각의 감귤 시료들의 그룹 경계는 PCA에 비해 더욱 작아졌으며 동일 품종에 속하는 각 시료들의 반복구들도 각 그룹 내에 위치하여 PCA보다 감귤의 식별능력이 향상되었음을 알 수 있었다. 청견의 경우 궁천조생이 좌측 하단에 그룹을 형성했고, 네이블오렌지는 상단에 그룹을 형성했다. 3개의 그룹이 각각 다른 위치에 그룹화 되면서 청견의 형질이 다른 2 품종과 관계가 있음을 보여주고 있다(Fig. 29C). 부지화는 하단에 그룹을 형성하고, 폰칸보다는 우측 상단에 그룹을 형성한 청견과 중첩되는 모습을 볼 수 있었다. 이처럼 부지화는 폰칸의 형질 보다는 청견의 형질을 더 많이 나타내는 것으로 판단된다(Fig. 30C). 또한, 감평은 하단에 그룹을 형성하고 있고, 좌측 상단에는 부지화가 그룹을 형성하고, 우측 상단에는 서지향이 그룹을 형성하고 있다(Fig. 31C).

HCA dendrogram 분석 결과를 보면 PLS-DA dendrogram에서 또한 감귤 교배품종 간에 유연관계를 확연히 나타냈다(Fig. 29D, 30D, 31D). 청견의 유연관계를 보면, 궁천조생과 네이블오렌지가 유연관계가 높았고, 두 품종과 청



**Fig. 30** Multivariate analysis of FT-IR spectral data on genomic DNA from *C. hybrid* 'Shiranuhi' cross breeding lines (CK and CB) and their progeny lines (CR). A: Representative FT-IR spectra of cross breeding lines (CK and CB) and their progeny lines (CR). B: PCA score plot of FT-IR spectral data. C: PLS-DA score plot of FT-IR spectral data. D: HCA dendrogram of PLS-DA score. Arrows indicate FT-IR spectra variables showing significant variation between cross breeding lines and hybrid lines (A). Dotted circles represent each boundary for group discrimination (B, C). Capitals and numbers represent sample names and replicates, respectively. CK : *C. hybrid* 'Kiyomi', CB : *Citrus reticulata* 'Ponkan', CR : *C. hybrid* 'Shiranuhi'.



**Fig. 31** Multivariate analysis of FT-IR spectral data on genomic DNA from *C. hybrid* 'Shiranuhi' cross breeding lines (CR and CH) and their progeny lines (CP). A: Representative FT-IR spectra of cross breeding lines (CR and CH) and their progeny lines (CP). B: PCA score plot of FT-IR spectral data. C: PLS-DA score plot of FT-IR spectral data. D: HCA dendrogram of PLS-DA score. Arrows indicate FT-IR spectra variables showing significant variation between cross breeding lines and hybrid lines (A). Dotted circles represent each boundary for group discrimination (B, C). Capitals and numbers represent sample names and replicates, respectively. CR : *C. hybrid* 'Shiranuhi', CH : *C. hybrid* 'Nishinokaori', CP : *C. hybrid* 'Kanpei'.

견이 유연관계가 있음을 나타냈다(Fig. 29D). 이것은 궁천조생과 네이블오렌지의 유전적인 형질이 청견에서 동시에 나타나면서 유전적으로 두 품종과 높은 유연관계를 나타내는 것으로 판단된다. 부지화의 경우는 부지화와 청견이 유연관계가 높았고, 폰칸은 부지화와 유연관계가 낮았다(Fig. 30D). 형태적인 표현이 부지화가 청견의 크기, 모양이 유사한 것을 볼 수 있고, 유전적으로 청견에 형질이 많이 나타났음을 알 수 있었다. 또한, 감평의 경우 감평과 부지화가 높은 유연관계를 나타냈고, 서지향과는 감평이 낮은 유연관계를 나타냈다(Fig. 31D). 이러한 결과는 HCA dendrogram에서 *B. campestris* subsp. *pekinensis* 의 부모 라인과 F1 자손과 동일한 식별 내용을 알 수 있었다(Fig. 23D, 24D). 따라서, FT-IR 스펙트럼 분석이 유성 생식으로 인해 변형된 고유의 DNA 특성을 가진 F1 자손 식별을 위한 새로운 도구로 적용 할 수 있는 것이 판단된다. 또한, 본 연구의 F1 자손 식별 시스템은 전체 genome 수준에서 DNA 이성종의 예비 평가와 예측에 이용될 수 있다. 하지만, 정확하게 genomic DNA(base composition, DNA backbone, secondary structure, etc.)의 특성이 F1 자손 식별에 중요한 역할을 할 것이다.

따라서 본 연구는 적외선 분광분석 데이터의 다변량 통계분석 기법을 활용하여 다양한 식물종의 신속하고 간편한 품종 식별 시스템을 확립하고, 더 나아가 기능성 성분의 함량이 우수한 라인을 비교적 간편한 방법으로 식별할 수 있는 선별체계를 확립함으로써 우수 품종 육성의 기간을 단축하여 품종 개발을 더욱 가속화시키기 위한 수단으로 활용이 가능함을 확인하였다. 아울러 본 기술이 감귤의 경우에도 품종식별이나 기능성 성분 함량 예측 수단으로 활용이 가능함을 확인할 수 있었다. 본 연구결과를 종합해 보면 적외선 분광분석 데이터의 다변량 통계분석 기법은 교배육성중인 계통의 품종 식별, 유용 기능성 물질의 함량이 높은 라인 신속 선별 수단으로 활용이 가능할 것으로 기대되며 육종을 통한 기능성 품종개발 가속화에 기여 할 수 있을 것으로 예상된다. 또한, 이 기술은 교배 육종에 빠르고 정확한 선택을 위해 이용될 수 있을 것이며, 기능성 품종 육성에도 효과적인 수단으로 활용이 가능할 것으로 기대된다.

## 9. REFERENCE

- 전익수, 김병률, 한석호, 민자혜. 2013. 한·중 FTA가 제주 감귤 생산액과 관련 전후방 산업에 미치는 파급효과 분석 농업경제 제36권 제1호 1-24.
- 김창명. 2006. 제주 감귤 산업의 발전 전략 한국식품저장유통학회 학술대회 39-56.
- Argyri, A.A., R.M. Jarvis, D. Wedge, Y. Xu, E.Z. Panagou, R. Goodacre, and G.J.E. Nychas. 2013. A comparison of Raman and FT-IR spectroscopy for the prediction of meat spoilage. Food Control 29:461-470.
- Argov, S., J. Ramesh, A. Salman, I. Sinelnikov, J. Goldstein, H. Guterman, and S. Mordechai. 2002. Diagnostic potential of Fourier transform infrared microspectroscopy and advanced computational methods in colon cancer patients. J. Biomed. Opt. 7:248-254.
- Armstrong, G.A. and J.E. Hearst. 1996. Carotenoids 2: Genetics and molecular biology of carotenoid pigment biosynthesis. Faseb. J. 10:228-37.
- Aktumsek, A., G. Zengin, G.O. Guler, Y.S. Cakmak, and A. Duran. 2013. Assessment of the antioxidant potential and fatty acid composition of four *Centaurea L. taxa* from Turkey. Food Chem. 141:91-97.
- Banyay, M., M. Sarkar, and A. Graslund. 2003. A library of IR bands of nucleic acids in solution. Biophys. Chem. 104:477-488.

- Baranska, M., W. Schütze, and H. Schulz. 2007. Determination of lycopene and beta-carotene content in tomato fruits and related products: Comparison of FT-Raman, ATR-IR, and NIR spectroscopy. *Analytical Chem.* 78:8456-8461.
- Bastiena, P., V.E. Vinzi, and M. Tenenhaus. 2005. PLS generalised linear regression. *Computational Stat. Data Analysis.* 48:17-46.
- Brewer, S.H., S.J. Anthireya, S.E. Lappi, D.L. Drapcho, and S. Franzen. 2002. Detection of DNA Hybridization on Gold Surfaces by Polarization Modulation Infrared Reflection Absorption Spectroscopy. *Langmuir.* 18:4460-4464.
- Bronner, W.E. and G.R. Beecher. 1995. Extraction and measurement of prominent flavonoids in orange and grapefruit juice concentrates. *J. Chromatography A.* 705:247-256.
- Cercós, M., G. Soler, D.J. Iglesias, J. Gadea, J. Forment, and M. Talon. 2006. Global analysis of gene expression during development and ripening of citrus fruit flesh. A proposed mechanism for citric acid utilization. *Plant Mol. Biol.* 62:513-527.
- Chen, Y., M. Xie, H. Zhang, Y. Wang, S. Nie, and C. Li. 2012. Quantification of total polysaccharides and triterpenoids in *Ganoderma lucidum* and *Ganoderma atrum* by near infrared spectroscopy and chemometrics. *Food Chem.* 135:268-275.
- Chen, E.H., D.D. Wei, G.M. Shen, G.R. Yuan, P.P. Bai and J.J. Wang. 2013. De novo characterization of the *Dialeurodes citri* transcriptome: Mining genes

- involved in stress resistance and simple sequence repeats (SSRs) discovery. *Insect Mol. Biology.* 23:52-66.
- Chen, L., N.C. Carpita, W.D. Reiter, R.H. Wilson, C. Jeffries and M.C. McCann. 1998. A rapid method to screen for cell-wall mutants using discriminant analysis of Fourier transform infrared spectra. *Plant J.* 16: 385–392.
- Colthup, N.B., L.H. Daly, and S.E. Wiberly. 1990. *Introduction to Infrared and Raman Spectroscopy.* 3d ed. Boston:Academic Press.
- Cunningham, F.X., and E. Gantt. 1998. Genes and enzymes of carotenoid biosynthesis in plants. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 49:557-583.
- Dovbeshko, G.I., N.Y. Gridina, E.B. Kruglova, and O.P. Pashchuk. 2000. FTIR spectroscopy studies of nucleic acid damage. *Talanta.* 53:233–246.
- D'Souza, L., P. Devi, M.P.D. Shridhar, and C.G. Naik. 2008. Use of Fourier Transform Infrared (FTIR) spectroscopy to study cadmium-Induced changes in *Padina Tetrastromatica* (Hauck) *Anal. Chem. Insights* 3:135-143.
- Duarte, I.F., A. Barros, I. Delgadillo, C. Almeida, and A.M. Gil, 2002. Application of FTIR spectroscopy for the quantification of sugars in mango juice as a function of ripening. *J. Agric. Food Chem.* 50:3104-3111.
- Dumas, P. and L. Miller. 2003. The use of synchrotron infrared microspectroscopy in biological and biomedical investigations. *Vib. Spec.* 32:3-21.
- Dukor, R.K. 2002. *Vibrational spectroscopy in the detection of cancer.* In:

- Chalmers JM, Griffiths PR (eds) Handbook of vibrational spectroscopy 5. Wiley, London, pp 3335–3361.
- Emura, K., S. Yamanaka, H. Isoda, and K.N. Watanabe. 2006. Estimation for different genotypes of plants based on DNA analysis using near-infrared (NIR) and fourier-transform infrared (FT-IR) spectroscopy. *Breed Sci.* 56:399-403.
- EsteveAgelet. L., A.A. Gowen, C.R. Hurburgh Jr., and C.P. O'Donell. 2012. Feasibility of conventional and Roundup Ready soybeans discrimination by different near infrared reflectance technologies. *Food Chem.* 134:1165-1172.
- Falk, M., K.A. Hartman, and R.C. Lord. 1963. Hydration of deoxyribonucleic acid. III. A spectroscopic study of the effect of hydration on the structure of deoxyribonucleic acid. *J. Am. Chem. Soc.* 85:391–394.
- Fanciullino, A.L., C. Dhuique-Mayer, F. Luro, J. Casanova, R. Morillon, and P. Ollitrault. 2006. Carotenoid diversity in cultivated *Citrus* is highly influenced by genetic factors. *J. Agric. Food Chem.* 54:4397-4406.
- Fiehn, O., J. Kopka, P. Drmann, T. Altmann, R. Trethewey and L. Willmitzer. 2000. Metabolite profiling for plant functional genomics. *Nat. Biotechnol.* 18:1157-1161.
- Fischer, G., S. Braun, R. Thissen, and W. Dott. 2006. FT-IR spectroscopy as a tool for rapid identification and intra-species characterization of airborne filamentous fungi. *J. Microbiol. Methods* 64:63-77.
- Gallardo-Velázquez, T., G. Osorio-Revilla, M. Zuñiga de Loa, and Y.

- Rivera-Espinoza. 2009. Application of FTIR-HATR spectroscopy and multivariate analysis to the quantification of adulterants in Mexican honeys. *Food Res. Intl.* 42:313–318.
- Gault, N., J.L. Poncy, and J.L. Lefaix. 2004. Radiation-induced apoptosis: a new approach by means of infrared microspectroscopy. *Can J. Physiol. Pharmacol.* 82:38-49.
- Geinguenaud, F., J. Liquier, M.G. Brevnov, O.V. Petrauskene, Y.I. Alexeev, E.S. Gromova, and E. Taillandier. 2000. Parallel self-associated structures formed by T,C-rich sequences at acidic pH. *Biochemistry* 39:12650-12658.
- Giovannucci, E., E.B. Rimm, Y. Liu, M.J. Stampfer, and W.C. Willett. 2003. A prospective study of cruciferous vegetables and prostate cancer. *Cancer Epidemiol Biomarkers Prev* 12:1403–1409.
- Gonzalez-Ruiz, V., A.I. Olives, M.A. Martin, P. Ribelles, M.T. Ramos, and J.C. Menendez. 2011. An overview of analytical techniques employed to evidence drug-DNA interactions. Applications to the design of genosensors. In: *Biomedical engineering, trends, research and technologies*, M. A. Komorowska and S. Olszynska-Janus (eds.), In Tech, pp 65-90.
- Goodacre, R., É.M. Timmins, R. Burton, N. Kaderbhai, D.B. Kell and P.J. Rooney. 1998. Rapid identification of urinary tract infection bacteria using hyperspectral whole-organism fingerprinting and artificial neural networks. *Microbiology.* 144:1157-1170.
- Goodner, K.L., R.L. Rouseff, and H.J. Hofsommer. 2001. Orange, mandarin, and hybrid classification using multivariate statistics based on carotenoid

- profiles. *J. Agric. Food Chem.* 49:1146-1150.
- Gross, J. 1987. In *Carotenoids: Pigments in Fruits*, Academic Press, London, UK.
- Griffiths, P. A., and J. A. de Haseth. 1986. *Fourier Transform Infrared Spectrometry*. New York: Wiley-Interscience.
- Harbaum, B., E.M. Hubbermann, C. Wolff, R. Herges, Z. Zhu, and K. Schwarz. 2007. Identification of flavonoids and hydroxycinnamic acids in pakchoi varieties (*Brassica campestris* L. ssp. *chinensis* var. *communis*) by HPLCESI-MS and NMR and their quantification by HPLCDAD. *J. Agric. Food Chem.* 55:8251-8260.
- Hirai, M.Y., M. Yano, D.B. Goodenowe, S. Kanaya, T. Kimura, M. Awazuhara, M. Arita, T. Fujiwara, and K. Saito. 2004. Integration of transcriptomics and metabolomics for understanding of global responses to nutritional stresses in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci.* 101:10205-10210.
- Höskuldsson, A. 1988. PLS regression methods. *J. Chemometrics* 2:211-228.
- Jin, I.D., Y.I. Kwon, H.Y. Lee, and I.J. Kim. 2014. Microscopic Observation of Leaves in Citrus Mutants. *J. Asian Agric. Biotechnol.* 30:11-15.
- Jun, I.S. and J.Y. Yoon. 2013. Analysis of the Competitiveness Factors of the Citrus Industry of Jeju. *Kor. J. Food Preservation.* 30:23-42.
- Kato, M., Y. Ikoma, H. Matsumoto, M. Sugiura, H. Hyodo, and M. Yano. 2004. Accumulation of carotenoids and expression of carotenoid biosynthetic genes during maturation in citrus fruit. *Plant Physiol.*

134:824-837.

Khan, M.K., Z.E. Huma, and O. Dangles. 2014. A comprehensive review on flavanones, the major citrus polyphenols. *J. Food Composition and Analysis*. 33:85–104.

Kim, D.O., O.I. Padilla-Zakour, and P.D. Griffiths. 2004. Flavonoids and antioxidant capacity of various cabbage genotypes at juvenile stage. *J. Food Sci.* 69:685-689.

Kim, S.W., Y.K. Kwon, J.M. Seo, T.H. Woo, and J.R. Liu. 2011. Prediction and discrimination of taxonomic relationship within *Orostachys* species using FT-IR spectroscopy combined by multivariate analysis. *J. Plant Biotechnol.* 38:9-14.

Kim, S.W., S.H. Ban, H. Chung, S.H. Cho, H.J. Chung, P.S. Choi, O.J. Yoo, and J.R. Liu. 2004. Taxonomic discrimination of higher plants by multivariate analysis of Fourier transform infrared spectroscopy data. *Plant Cell Rep.* 23: 246-250.

Kim, S.W., S.H. Cho, H. Chung, and J.R. Liu. 2007. Genetic discrimination between *Catharanthus roseus* cultivars by multivariate analysis of fourier transform infrared spectroscopy data. *J. Plant Biotechnol.* 34:201-205.

Kim, S.W., S.R. Min, J.H. Kim, S.K. Park, T.I. Kim, and J.R. Liu. 2009. Rapid discrimination of commercial strawberry cultivars using fourier transform infrared spectroscopy data combined by multivariate analysis. *Plant Biotechnol. Rep.* 3:87-93.

- Kim, J.K., S.Y. Lee, S.M. Chu, S.H. Lim, S.C. Suh, Y.T. Lee, H.S. Cho, and S.H. Ha. 2010. Variation and correlation analysis of flavonoids and carotenoids in Korean pigmented rice (*Oryza sativa* L.) cultivars. *J. Agric. Food Chem.* 58:12804-12809.
- Kimura, M. and D.B. Rodriguez-Amaya. 2002. A scheme for obtaining standards and HPLC quantification of leafy vegetable carotenoids. *Food Chem.* 78:389-398.
- Kofalvi, S.A. and A. Nassuth. 1995. Influence of wheat streak mosaic virus infection on phenylpropanoid metabolism and the accumulation of phenolics and lignin in wheat. *Physiol. Mol. Plant Pathol.* 47:365-377.
- Krishnan, P., N.J. Kruger, and R.G. Ratcliffe. 2005. Metabolite fingerprinting and profiling in plants using NMR. *J. Exp. Bot.* 56:255-265.
- Kwon, Y.S., J.Y. Moon, Y.S. Kwon, D.Y. Park, W.M. Yoon, I.H. Song, and S.I. Yi. 2003. AFLP analysis on radish and Chinese cabbage were proved to be useful in defining and identification of varieties. *Kor. J. Breed* 35:319-328.
- Lee, D. and D. Chapman. 1986. Infrared spectroscopic studies of biomembranes and model membranes. *Biosci. Rep* 6:235-256.
- Lee, E.J., R. Shaykhutdinov, A.M. Weljie, H.J. Vogel, and P.J. Facchini. 2009. Quality assessment of ginseng by <sup>1</sup>H NMR metabolite fingerprinting and profiling analysis. *J. Agric. Food Chem.* 57:7513-7522.
- Lee, H.S. and W.S. Castle. 2001. Seasonal changes of carotenoid pigments and color in Hamlin, Earlygold, and Budd Blood orange juices. *J. Agric. Food*

Chem. 49:877-882.

Lee, J.H. and M.G. Choung. 2011. Nondestructive determination of herbicide-resistant genetically modified soybean seeds using near-infrared reflectance spectroscopy. *Food Chem.* 126:368-373.

Lee, S.H., S.J. Suh, K.H. Lee, J.B. Yang, S.U. Choi, and S.S. Park. 2013. Anti-Inflammatory Effect of Peel Extracts from Citrus Fruits. *J. Fd Hyg. Safety.* 28:342-348.

Lei, Z., D.V. Huhman, and L.W. Sumner. 2011. Mass spectrometry strategies in metabolomics. *J. Biol. Chem.* 286:25435-25442.

Leopold, L.F., N. Leopold, H.-A. Diehl, and C. Socaciu. 2011. Quantification of carbohydrates in fruit juices using FTIR spectroscopy and multivariate analysis. *Spectroscopy* 26:93-104.

Lichtenthaler, H.K. and C. Buschmann. 2001. Chlorophylls and carotenoids: measurement and characterization by UV-VIS spectroscopy. *Curr. Prot. Food Anal. Chem.* F4.3.1–F 4.3.8.

Lindqvist, M., M. Sarkar, A. Winqvist, E. Rozners, R. Stromberg, and A. Graslund. 2000. Optical spectroscopic study of the effects of a single deoxyribose substitution in a ribose backbone: Implications in RNA-RNA interaction. *Biochemistry* 39:1693-1701.

Liu, S.R., W.Y. Li, D. Long, C.G. Hu, and J.Z. Zhang. 2014. Development and Characterization of Genomic and Expressed SSRs in Citrus by Genome-Wide Analysis. *PLoS One.* 28:e75149.

- Liu, Y., Y. Ying, H. Yu, and X. Fu. 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.
- Liu, Q., J. Xu, Y.Z. Liu, X.L. Zhao, X.X. Deng, L.L. Guo, and J.Q. Gu. 2007. A novel bud mutation that confers abnormal patterns of lycopene accumulation in sweet orange fruit (*Citrus sinensis* L. Osbeck). *J. Exp. Bot.* 58:4161- 4171.
- Liquier, J., P. Coffinier, M. Firon, and E. Taillandier. 1991. Triple helical polynucleotidic structures: sugar conformations determined by FTIR spectroscopy. *J. Biomol. Struct. Dyn.* 9:437-445.
- Lopez-Sanchez, M., M.J. Ayora-Canada, and A. Molina-Diaz. 2010. Olive fruit growth and ripening as seen by vibrational spectroscopy. *J. Agric. Food Chem.* 58:82-87.
- Magwaza, L.S., U.L. Opara, P.J.R. Cronje, S. Landahl, H.H. Nieuwoudt, A.M. Mouazen, B.M. Nicolai, and L.A. Terry. 2014. Assessment of rind quality of 'Nules Clementine' mandarin fruit during postharvest storage: 2. Robust Vis/NIRS PLS models for prediction of physico-chemical attributes. *Scientia Horticulturae* 165:421–432.
- Malins, D.C., N.L. Polissar, and S.J. Gunselman. 1997. Models of DNA structure achieve almost perfect discrimination between normal prostate, benign prostatic hyperplasia (BPH), and adenocarcinoma and have a high potential for predicting BPH and prostate cancer. *Proc. Natl. Acad. Sci.* 94:259-264.

- Malins, D.C., P.M. Johnson, E.A. Barker, N.L. Polissar, T.M. Wheeler, and K.M. Anderson. 2003. Cancer-related changes in prostate DNA as men age and early identification of metastasis in primary prostate tumors. *Proc. Natl. Acad. Sci.* 100:5401-5406.
- Malins, D.C., N.K. Gilman, V.M. Green, T.M. Wheeler, E.A. Barker, and K.M. Anderson. 2005. A cancer DNA phenotype in healthy prostates, conserved in tumors and adjacent normal cells, implies a relationship to carcinogenesis. *Proc. Natl. Acad. Sci.* 102:19093-19096.
- Mammadov, J., R. Aggarwal, R. Buyyarapu, and S. Kumpatla. 2012. SNP markers and their impact on plant breeding. *Int. J. Plant Genomics* 2012, Article ID 728398. doi:10.1155/2012/728398.
- Marti, G., J. Boccard, F. Mehl, B. Debrus, L. Marcourt, P. Merle, E. Delort, L. Baroux, H. Sommer, S. Rudaz, and J.L. Wolfender. 2014. Comprehensive profiling and marker identification in non-volatile citrus oil residues by mass spectrometry and nuclear magnetic resonance. *Food Chem.* 150:235–245.
- Mas, S., A. Crescenti, P. Gasso', R. Deulofeu, R. Molina, A. Ballesta, T.W. Kensler, and A. Lafuente. 2007. Induction of apoptosis in HT-29 cells by extracts from isothiocyanates-rich varieties of *Brassica oleracea*. *Nutr. Cancer* 58:107-114.
- Matsumoto, H., Y. Ikoma, M. Kato, T. Kuniga, N. Nakajima, and T. Yoshida. 2007. Quantitative analysis of carotenoids in citrus fruit by LC-MS and comparison of patterns of seasonal changes for carotenoids among citrus varieties. *J. Agric. Food Chem.* 55:2356-2368.

- Mehouachi, J., D. Serna, S. Zaragoza, M. Agustí, M. Talon, and E. Primomillo. 1995. Defoliation increases fruit abscission and reduces carbohydrate levels in developing fruits and woody tissues of *Citrus unshiu*. *Plant Science* 107:189-197.
- Melendez-Martinez, A.J., I.M. Vicario, and F.J. Heredia. 2003. A routine high-performance liquid chromatography method for carotenoid determination in ultrafrozen orange juices. *J. Agric. Food Chem.* 51:4219-4224.
- Mello, M.L.S and B.C. Vidal. 2012. Changes in the infrared microspectroscopic characteristics of DNA caused by cationic elements, different base richness and single-Stranded form. *PLoS One* 7(8): e43169.
- Mevik, B.H. and R. Wehrens. 2007. The pls package: principal component and partial least squares regression in R. *J. Stat. Software* 18.
- Moran, N.A., and T. Jarvik. 2010. Lateral Transfer of Genes from Fungi Underlies Carotenoid Production in Aphids. *Science*. 328(5978):624-627.
- Mungur, R., A.D.M. Glass, D.B. Goodenow, and D.A. Lightfoot. 2005. Metabolite fingerprinting in transgenic *Nicotiana tabacum* altered by the *Escherichia coli* glutamate dehydrogenase Gene. *J. Biomed. Biotechnol* 2005:198-214.
- Muntean, C.M., A. Halmagyi, M.D. Puia, and I. Pavel. 2009. FT-Raman of genomic DNA from plant tissues. *Spectroscopy* 23:59-70.

- Nageswara-Rao, M., M. Irey, S.M. Garnsey, and S. Gowda. 2013. Candidate gene makers for *Candidatus Liberibacter asiaticus* for detecting citrus greening disease. *J. Biosci.* 38:229–237.
- Naumann, A., M. Navarro-Gonzalez, S. Peddireddi, U. Kues, and A. Polle. 2005. Fourier transform infrared microscopy and imaging: detection of fungi in wood. *Fungal Genet Biol.* 42:829-835.
- Naumann, D., D. Helm, and H. Labischinski. 1991. Microbiological characterizations by FT-IR spectroscopy. *Nature* 351:81-82.
- Ngo-Thi, N.A., C. Kirschner, and D. Naumann. 2003. Characterization and identification of microorganisms by FT-IR microspectrometry. *J. Mol. Struct.* 661-662:371-380.
- Olson, J.A. 1989 Provitamin A function of carotenoids: the conversion of  $\beta$ -carotene into vitamin. *A. J. Nutr.* 119:105-108.
- Pandey, K.K. 1999. A study of chemical structure of soft and hardwood and wood polymers by FTIR spectroscopy. *J. Appl. Polym. Sci.* 71:1969-1975.
- Park, H.G., S.H. Lee, H.Y. Kim, H.S. Jeong, E.Y. Kim, Y.W. Yun, S.Y. Nam, and B.J. Lee. 2011. Comparison in Antioxidant Effects of Four Citrus Fruits. *J. Fd Hyg. Safety.* 26:355-360.
- Parker, F.S. 1983. Applications of infrared, Raman and resonance Raman spectroscopy in biochemistry, Plenum Press, New York. 527.
- Páscoa, R.N.M.J., L.M. Magalhães, and J.A. Lopes. 2013. FT-NIR spectroscopy

as a tool for valorization of spent coffee grounds: Application to assessment of antioxidant properties. *Food Res. Intl.* 51:579-586.

Peterson, J.J., J.T. Dwyer, G.R. Beecher, S.A. Bhagwat, S.E. Gebhardt, D.B. Haytowitz, and J.M. Holden. 2006. Flavanones in oranges, tangerines (mandarins), tangors, and tangelos: a compilation and review of the data from the analytical literature. *J. Food Composition and Analysis.* 19:S66–S73.

Polat, I., Y.A. Kacar, T. Yesiloglu, A. Uzun, O. Tuzcu, O. Gulsen, M. Incesu, G. Kafa, E. Turgutoglu, and S. Anil. 2012. Molecular characterization of sour orange (*Citrus aurantium*) accessions and their relatives using SSR and SRAP markers. *Genet. Mol. Res.* 11:3267-3276.

Preisner, O., J.A. Lopes, R. Guiomar, J. MacHado, and J.C. Menezes. 2007. Fourier transform infrared (FT-IR) spectroscopy in bacteriology: Towards a reference method for bacteria discrimination. *Anal. Bioanal. Chem.* 387:1739-1748.

Ramchiary, N., D.V. Nguyen, X. Li, C.P. Hong, V. Dhandapani, S.R. Choi, G. Yu, Z.Y. Piao, and Y.P. Lim. 2011. Genic microsatellite markers in *Brassica rapa*: development, characterization, mapping, and their utility in other cultivated and wild *Brassica* relatives. *DNA Res.* 18:305-320.

Roessner, U., A. Luedemann, D. Brust, O. Fiehn, T. Linke, L. Willmitzer, and A. Fernie. 2001. Metabolic profiling allows comprehensive phenotyping of genetically or environmentally modified plant systems. *Plant Cell* 13:11-29.

Sahu, R.K., S. Mordechai, and E. Manor. 2008. Nucleic acids absorbance in

mid IR and its effect on diagnostic variates during cell division: A case study with lymphoblastic cells. *Biopolymers* 89: 993-1001.

Sambrook, J., E.F. Fritsch, and T. Maniatis. 1989. *Molecular cloning: A laboratory manual*, 2nd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Schulz, H., and M. Baranska. 2007. Identification and quantification of valuable plant substances by IR and Raman spectroscopy. *Vibrational Spectroscopy* 43:13–25.

Silverstein, R. M., F. C. Bassler, and T. C. Morrill. 1981. *Spectrometric Identification of Organic Compounds*. 4th ed. New York: John Wiley and Sons.

Smith, A. L. 1979. *Applied Infrared Spectroscopy: Fundamentals, Techniques, and Analytical Problem Solving*. Chemical Analysis series, vol. 54. New York: John Wiley and Sons.

Son, H.S., G.S. Hwang, K.M. Kim, E.Y. Kim, F. van den Berg, W.M. Park, C.H. Lee, and Y.S. Hong. 2009. <sup>1</sup>H NMR-based metabolomic approach for understanding the fermentation behaviors of wine yeast strains. *Anal. Chem.* 81:1137-1145.

Song, S.Y., E.Y. Jie, M.S. Ahn, D.J. Kim, I.J. Kim, and S.W. Kim. 2014. Discrimination of High Functional African Yams Using FT-IR Fingerprinting Combined by Multivariate Analysis and Quantitative Prediction of Functional Compounds by PLS Regression Modeling. *Kor. J. Hort. Sci. Technol.* 32:105-114.

- Song, S.Y., E.Y. Jie, M.S. Ahn, I.H. Lee, I.S. Nou, B.W. Min, and S.W. Kim. 2014. Fourier transform infrared (FT-IR) spectroscopy of genomic DNA to discriminate F1 progenies from their paternal lineage of Chinese cabbage (*Brassica rapa* subsp. *pekinensis*). *Mol. Breeding* 33:453–464.
- Song, S.Y., J.M. Heo, and I.J. Kim. 2014. Current State of Citrus Mutation Breeding (2013). *J. Asian Agri. Biotech.* 30:65-70.
- Song, S.Y., T.J. Ha, K. C. Jang, I.J. Kim, and S.W. Kim. 2012. Establishment of rapid discrimination system of leguminous plants at metabolic level using FT-IR spectroscopy with multivariate analysis. *J. Plant Biotechnol.* 39:121–126.
- Stadnik, M.J. and H. Buchenauer. 2000. Inhibition of phenylalanine ammonia-lyase suppresses the resistance induced by benzothiadiazole in wheat to *Blumeria graminis* f. sp. *tritici*. *Physiol. Mol. Plant Pathol.* 57:25-34.
- Stewart, D., N. Yahiaoui, G.J. McDougall, K. Myton, C. Marque, A.M. Boudet, and J. Haigh. 1997. Fourier-transform infrared and Raman spectroscopic evidence for the incorporation of cinnamaldehydes into the lignin of transgenic tobacco (*Nicotiana tabacum* L.) plants with reduced expression of cinnamyl alcohol dehydrogenase. *Planta.* 201:311-318.
- Sun, X., Q. Mu, D. Jiang, C. Wang, X.C. Wang, and J.G. Fang. 2012. A new strategy employed for identification of sweet orange cultivars with RAPD markers. *Genet. Mol. Res.* 11:2071-2080.

- Sundaram, J., B. Park, A. Hinton, S.C. Yoon, and K.C. Lawrence. 2012. Identification and characterization of Salmonella serotypes using DNA spectral characteristics by Fourier Transform Infrared (FT-IR) spectroscopy. Proceedings of SPIE- The International Society for Optical Engineering. pp 8369.
- Tadeo, F.R., M. Cercos, J.M. Colmenero-Flores, D.J. Iglesias, M.A. Naranjo, G. Rios, E. Carrera, O. Ruiz-Rivero, I. Lliso, R. Morillon, P. Ollitrault, and M. Talon. 2008. Molecular Physiology of Development and Quality of Citrus. Advances in Botanical Research. 47:147-223.
- Talon, M. and F.G. Gmitter Jr. 2008. Citrus Genomics. Int. J. Plant Genom. 2008:1-17.
- Taveira, M., D.M. Pereira, C. Sousa, F. Ferreres, P.B. Andrade, A. Martins, J.A. Pereira, and P. Valentão. 2009. In vitro cultures of Brassica oleracea L. var. costata DC: Potential plant bioreactor for antioxidant phenolic compounds. J. Agric. Food Chem. 57:1247-1252.
- Terhoeven-Urselmans, T., T.G. Vagen, O. Spaargaren, and K.D. Shepherd. 2010. Prediction of soil fertility properties from a globally distributed soil mid-infrared spectral library. Soil. Sci. Soc. Am. J. 74:1792-1799.
- Thormann, C.E., M.E. Ferreira, L.E.A. Camargo, J.G. Tivang, and T.C. Osborn. 1994. Comparison of RFLP and RAPD markers to estimating genetic relationships within and among cruciferous species. Theor. Appl. Genet 88:973-980.
- Timmins, É.M., S.A. Howell, B.K. Alsberg, W.C. Noble, and R. Goodacre. 1998.

- Rapid differentiation of closely related *Candida* species and strains by pyrolysis-mass spectrometry and Fourier transform-infrared spectroscopy. *J. Clin. Microbiol.* 36:367-374.
- Tripoli, E., M.L. Guardia, S. Giammanco, D.D. Majo, and M. Giammanco. 2007. Citrus flavonoids: Molecular structure, biological activity and nutritional properties: A review. *Food Chem.* 104:466–479.
- Trygg, J., E. Holmes, and T. Londstedt. 2007. Chemometrics in metabonomics. *J. Proteomes Res.* 6:467-479.
- van Poppel, G., D.T. Verhoeven, H. Verhagen, and R.A. Goldbohm. 1999. Brassica vegetables and cancer prevention. *Epidemiology and mechanisms. Adv. Exp. Med. Biol.* 472:159-168.
- Wenning, M., H. Seiler, and S. Scherer. 2002. Fourier-transform infrared microspectroscopy, a novel and rapid tool for identification of yeasts. *Appl. Environ. Microbiol.* 68:4717-4721.
- Williams, J.G.K., A.R. Kubelik, L.J. Livak, J.A. Rafalski, and S.V. Tingey. 1990. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucl. Acids Res.* 18:6531-6535.
- Wold, H. 1966. Estimation of principal components and related models by iterative least squares, p. 391-420. In: K. R. Krishnaiah (ed.). *Multivariate Analysis*. Academic Press, New York.
- Wold, S., M. Sjöström, and L. Eriksson. 2001. PLS-regression: a basic tool of chemometrics. *Chemometrics and Intelligent Lab. Systems* 58:109–130.

- Wolkers, W.F., A.E. Oliver, F. Tablin, and J.H. Crowe. 2004. A fourier transform infrared spectroscopy study of sugar glasses. *Carb. Res.* 339:1077-1085.
- Wu, C.H., H.N. Murthy, E.J. Hahn, and K.Y. Paek. 2007. Improved production of caffeic acid derivatives in suspension cultures of *Echinacea purpurea* by medium replenishment strategy. *Arch. Pharm. Res.* 30:945-949.
- Yang, C.M., K.W. Chang, M.H. Yin, and H.M. Huang. 1998. Methods for the determination of the chlorophylls and their derivatives. *Taiwania* 43:116-122.
- Yee, N., L.G. Benning, V.R. Phoenix, and F.G. Ferris. 2004. Characterization of metal-Cyanobacteria sorption reactions: A combined Macroscopic and infrared spectroscopic investigation. *Environ. Sci. Technol.* 38:775-82.
- Yesudas, C.R., R. Bashir, M.B. Geisler, and D.A. Lightfoot. 2013. Identification of germplasm with stacked QTL underlying seed traits in an inbred soybean population from cultivars Essex and Forrest. *Mol. Breeding* 31:693-703.
- Yuan, J., C. Wang, H. Chen, H. Zhou, and J. Ye. 2013. Prediction of fatty acid composition in *Camellia oleifera* oil by near infrared transmittance spectroscopy (NITS). *Food Chem.* 138:1657-1662.
- Zhishen, J., T. Mengcheng, and W. Jianming. 1999. The determination of flavonoid contents in mulberry and their scavenging effects on superoxide radicals. *Food Chem.* 64:555-559.
- Zhou-Sun, B., J. Sun, S.M. Gryaznov, J. Liquier, T. Garestier, C. Helene, and E.

Taillandier. 1997. A physico-chemical study of triple helix formation by an oligodeoxythymidylate with N34'-P5' phosphoramidate linkages. Nucl. Acids Res. 25:1782-1787.

Zude, M., L. Spinelli, and A. Torricelli. 2008. Approach for non-destructive pigment analysis in model liquids and carrots by means of time-of-flight and multi-wavelength remittance readings. Analytica Chimica Acta 623:204-212.

## ACKNOWLEDGMENT

박사과정을 시작한지도 4년이란 시간이 흘렀습니다. 4년이라는 시간동안 많은 일을 경험했습니다. 새로운 학문을 공부하게 되었고, 보다 넓은 세상을 보았습니다. 힘들 때도 있었고, 즐겁고, 슬플 때도 있었습니다. 모진 풍파를 경험하면서 더욱더 단단해졌습니다. 박사라는 호칭에 어울리는 사람이 되도록 많은 노력을 아끼지 않았습니다. 주위에 많은 질책과 칭찬 속에서 무럭무럭 자라 이제 성목이 될 준비를 마쳤습니다. 더 넓은 세상에서 과학도로서의 소양과 지식을 실천하겠습니다. 또한 저를 성숙이란 열매로 바꿔주신 모든 분에게 이렇게 지면으로나마 감사의 인사를 드리고자 합니다.

먼저 그동안 지도해주신 김인중 교수님께 진심으로 깊은 감사를 드립니다. 저에게 새로운 길을 열어주시고 이끌어주시고 가르쳐주신 김석원 박사님께 정말 진심으로 감사 또 감사를 드립니다. 그리고 한결같은 마음으로 걱정해주신 류기중 교수님, 언제나 좋은 말씀을 해주신 김찬식 교수님, 항상 아낌없는 격려를 주신 이효연 교수님, 그리고 박사과정 동안 저의 학문적 지식을 넓혀주신 학과 교수님들께 진심으로 감사드립니다.

한국생명공학연구원이라는 낯선 곳에서 엄마 같이 챙겨주신 박종미 선생님, 잘못된 점을 누구보다 먼저 말씀해 주신 지은이 선생님, 언제든지 물어보면 친절하게 알려준 동생이지만 친구 같은 안명숙, 처음 대전이란 곳에서 만나 때로는 친구처럼 때로는 동생처럼 도와준 장혜림이 정말로 고맙고 감사합니다. 박사과정 동안 많은 힘이 되었습니다. 혼자 괴로워하면서 고민하고 있을 때 다독여주신 고석민 박사님, 언제나 밝은 미소로 응원해 주신 선현진 박사님, 누나처럼 반겨주신 송인자 박사님, 부족한 저를 격려해주신 은창호 박사님께 감사드립니다.

4년이라는 시간동안 부모님께 많은 걱정을 끼쳐 정말 죄송합니다. 어머니 앞으로는 멋진 아들로 살아가겠습니다. 집에서 항상 응원해준 사랑하는 여수정, 제대로 놀아주지도 못하고 그런데도 건강하게 자라준 사랑하는 나의 딸 송아름, 그리고 나의 아들 송우준 정말 많이 사랑합니다. 앞으로는 좋은 남편, 훌륭한 아빠, 멋진 아들이 되겠습니다. 우리 가족 정말 많이 많이 사랑합니다. 그리고 고맙습

니다. 믿고 지금까지 기다려줘서 감사합니다.

박사과정 동안 힘들고 지칠 때 많은 분들의 성원과 응원이 없었다면 지금의 저의 모습은 상상할 수도 없었을 것입니다. 돌아보면 후회한 것도 많지만 앞으로 더욱 멋진 사람으로 많은 분들의 성원과 응원에 보답하겠습니다.