식물생명공학회지 제34권 제3호
J Plant Biotechnol
Vol. 34, No. 3, 201-205 (2007)

Genetic Discrimination of *Catharanthus roseus* Cultivars by Multivariate Analysis of Fourier Transform Infrared Spectroscopy Data

Suk Weon Kim[1], Soohwa Cho[3], Hoeil Chung[3], and Jang R. Liu[2*]

[1]*Biological Resource Center*

[2]*Plant Genome Research Center, Korea Research Institute of Bioscience and Biotechnology (KRIBB), 111 Gwahangno, Yuseong-gu, Daejeon 305-806, Korea*

[3]*Department of Chemistry, Hanyang University, 17 Haengdang-dong, Seongdong-gu, Seoul 133-791, Korea*

ABSTRACT To determine whether pattern recognition based on metabolite fingerprinting for whole cell extracts of higher plants is applied to discriminate plants genetically, leaf samples of eight cultivars of Catharanthus roseus were subjected to Fourier transform infrared spectroscopy (FT-IR). FT-IR fingerprint region data were analyzed by principal component analysis (PCA). Major peaks as biomarkers were identified as the most significant contributors to distinguish samples by using genetic programming. A hierarchical dendrogram based on the results from PCA separated the eight cultivars into two major groups in the same manner as the dendrograms based on genetic fingerprinting methods such as RAPD and AFLP. A slight difference between the dendrograms was found only in branching pattern within each subgroup. Therefore, we conclude that the hierarchical dendrogram based on PCA of the FT-IR data represents the most probable chemotaxonomical relationship between cultivars, which is in general agreement with the genetic relationship determined by conventional DNA fingerprinting methods.

Introduction

Fourier transform infrared (FT-IR) spectroscopy is a rapid, simple, high-resolution analytical means for whole cell extracts. FT-IR is based on vibrations of functional groups and highly polar bonds of the sample components. Thus, FT-IR provides extremely high-density data sets of metabolite containing overlapping signals of the majority of the compounds when whole cell samples are analyzed. Multivariate analysis enables recognition of these data sets as simple patterns (metabolite fingerprinting). Using metabolite fingerprinting, FT-IR analysis has been applied to discrimination of closely related microbial strains (Goodacre et al. 1998, Timmins et al. 1998, Wenning et al. 2002). FT-IR also has been used in plant biology in many studies including screening of cell-wall mutant plants (Stewart et al. 1997, Chen et al. 1998), and phylogenetic discrimination of higher plant species and varieties (Kim et al. 2004). We have already reported that principal component analysis of the aromatic regions of ^{1}H NMR spectroscopy data exhibited possible relationships between cultivars, which were in general agreement with the genetic relationships determined by conventional DNA fingerprinting methods (Kim et al. 2007).

However, it remains to be determined whether pattern

*Corresponding author Tel 042-860-4430 Fax 042-860-4608
E-mail: jrliu@kribb.re.kr

recognition of FT-IR data of these approaches is valid in representing the genetic relationship between genotypes of a higher plant. We attempted to determine whether multivariate analysis of FT-IR data is able to discriminate eight cultivars of Catharanthus roseus.

Material and Methods

Plant Materials

Eight cultivars ('Apricot Delight', 'Cooler Grape', 'Cooler Peppermint', 'Equator Grape', 'Equator Rose', 'Equator White', 'Equator White Eye', and 'Little Bright Eye') of *Catharanthus roseus* (L.) G. Don were grown in a growth chamber (25°C, approximately 70 μmol $m^{-2}s^{-1}$ from cool-white fluorescent lamps with a 16-h photoperiod). Fully expanded leaves were excised from plants at the flowering stage and immediately plunged into liquid nitrogen before grinding with a mortar and pestle. Ground samples were then freeze-dried and stored at -70°C until used. Samples were run in triplicate using homogenized leaf samples from three different individual plants of each cultivar.

FT-IR Analysis

Five milligrams of freeze-dried, ground leaves was mixed with 80 mg KBr and ground again to reduce the particle size to less than 5 mm in diameter, as large particles scatter the infrared beam and cause a slope baseline of spectrum. The finely ground mixture was then crushed in a pellet press in order to form a pellet through which the beam of the spectrometer was able to pass. Infrared analysis was performed using a BOMEM FTIR spectrophotometer, Model DA 3.2, equipped with liquid nitrogen cooled MCT detector and a KBr beam-splitter. To improve the signal-to-noise ratio, 32 interferograms were co-added and averaged with the analytical results. Infrared spectra were obtained by subtraction of the plate spectra (background) used for deposition of the samples. Spectral resolution was 4 cm^{-1} and spectra collected over the wave number ranged from 8000 cm^{-1} to 400 cm^{-1}. Spectra were processed using the GRAMS/386 program (Galactic Industries Corporation, Salem, NH). FT-IR spectra were collected from three separate analyses.

Spectral Data Processing and Statistical Analysis of FT-IR Spectra

For FT-IR data, procedures were implemented to minimize problems arising from baseline shifts. Spectra were first subjected to path-length correction, and then the spectra were baseline corrected so that the smallest absorbance (2000 cm^{-1}) was equal to 0. The smoothed second derivatives of these normalized spectra were calculated using the Savitzky-Golay algorithm (Savitzky and Golay, 1964) with 5-point smoothing. The generated ASCII file was imported into Matlab (version 6.5) for PCA analysis according to NIPALS algorithm (Wold 1966). Following this processing, PCA and hierarchical clustering analysis were conducted. Hierarchical dendrograms were constructed from PCA of FT-IR data by the unweighted pair group method with arithmetic mean analysis (UPGMA) method using the euclidean distance as the similarity measure with MVSP (version 3.13; Kovach Computing Services).

Genetic programming (GP) analysis was used to determine biomarker variables that discriminated plants at each hierarchical level of the dendrogram. Gmax-bio software (Aber Genomic Computing) was used with the default parameters of a population size of 1,000; a maximum program length of 44 nodes; fitness based on tournament selection/Gmax (v); a crossover operator used 80% of the time, and terminals of the mutations were selected 20% of the time. The operators used were the default numeric (0.1, 1, 3, 5, and rand) and arithmetic (1, 2, /, and *) operators.

Results and Discussion

Genetic Discrimination of *C. roseus* Cultivars by Multivariate Analysis of FT-IR Spectral Data

Quantitative FT-IR data for each sample were obtained (Fig. 1A). PCA of FT-IR data are displayed in a two-dimensional plot using the first two principal components (Figure 1B). Re-

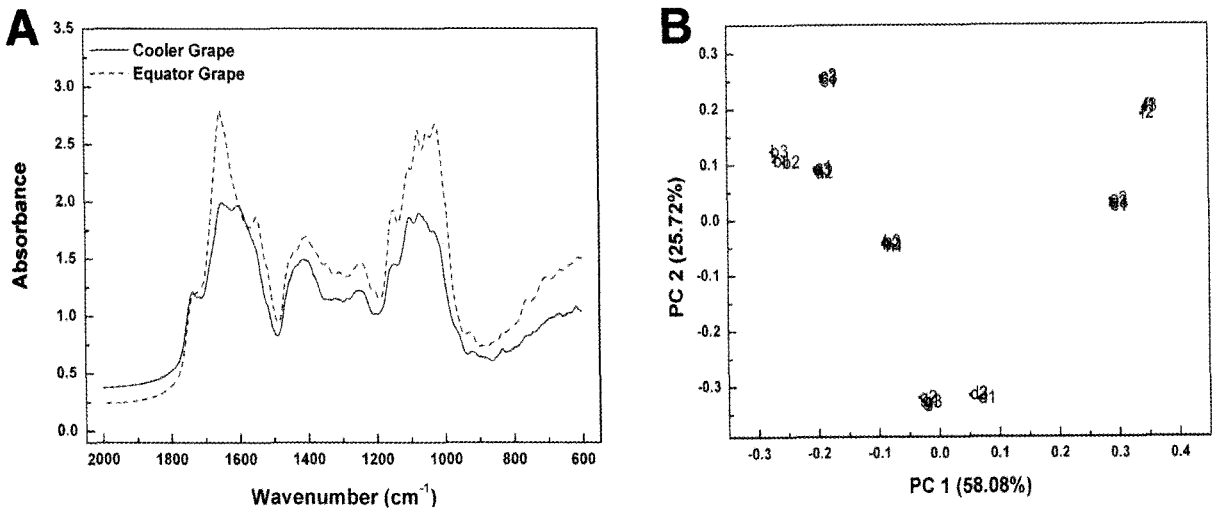

Figure 1. Representative FT-IR spectra (A) and two-dimensional PCA score plot (B) of FT-IR data from *C. roseus* cultivars and dendrogram based on PCA of FT-IR data of eight cultivars of *C. roseus*. A: Representative FTIR spectra of 'Cooler Grape' and 'Equator Grape'. B: Two-dimensional PCA plot of FT-IR data of eight cultivars of *C. roseus*. The first two principal components are displayed, accounting for 58.08 and 25.72% (83.80% total) of the total variation, respectively. a: 'Cooler Grape'; b: 'Cooler Peppermint'; c: 'Little Bright Eye'; d: 'Apricot Delight'; e: 'Equator Grape'; f: 'Equator Rose'; g: 'Equator White Eye'; h: 'Equator White'. The numbers (1, 2, and 3) following each letter indicate three replicates of each cultivar.

plicate samples of each cultivar were grouped in discrete clusters, indicating that PCA was able to discern one cultivar from another. A dendrogram based on hierarchical clustering analysis of FT-IR spectral data was constructed (Figure 2). The dendrogram separated the eight cultivars into two major groups (Figure 2) in the same manner as the dendrograms based on genetic fingerprinting methods such as RAPD and AFLP (Kim et al. 2007). The first group based on FT-IR data was consisting of 'Cooler Grape', 'Cooler Peppermint', and 'Little Bright Eye'. 'Cooler Peppermint' was clustered with 'Cooler Grape' whereas 'Cooler Peppermint was clustered with 'Little Bright Eye' in the dendrograms based on conventional DNA fingerprints (Kim et al. 2007). Within the second group based on FT-IR data, 'Apricot Delight' was clustered with 'Equator White' and 'Equator White Eye' whereas 'Apricot Delight' was clustered with 'Equator Rose' or 'Equator Grape' depending on the dendrogram based on DNA fingerprinting methods (Kim et al. 2007). Thus, the dendrogram based on FT-IR was the same as the dendrograms based on conventional DNA fingerprints in dividing the eight cultivars into two main groups, but the dendrograms based on FT-IR data showed a slight different branching order within each subgroup in

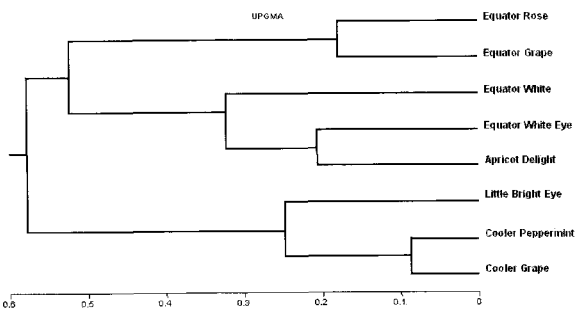

Figure 2. Hierarchical clustering analysis of FT-IR data of eight cultivars of *C. roseus*.

comparison with the dendrogram based on RAPD and AFLP. Considering these results, hierarchical dendrograms based on principal component analysis of the FT-IR spectral data exhibited possible relationships between C. roseus cultivars, which were in general agreement with the genetic relationships determined by conventional DNA fingerprinting methods. PCA that enables display of the natural relationship between the samples without prior knowledge was used to construct a dendrogram that discriminated 'Cooler' cultivars from 'Equator' cultivars in the main groups. In this study, samples for FT-IR contained the entire compounds including polymers and low molecular weight metabolites. 'Cooler' and 'Equator' cultivars

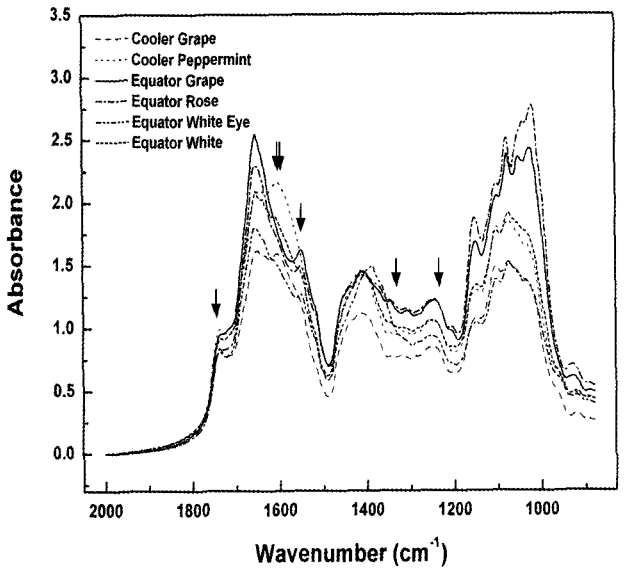

Figure 3. Biomarkers in FT-IR spectra of two Cooler cultivars and four Equator cultivars that discriminated Cooler cultivars from Equator cultivars identified by GP analysis. Arrows indicate six major peaks (biomarkers) of V371 (1593 cm^{-1}), V359 (1569 cm^{-1}), V393 (1226 cm^{-1}), V232 (1325 cm^{-1}), V372 (1596 cm^{-1}), and V449 (1743 cm^{-1}).

showed prominent differences in the FT-IR polypeptide region (Figure 3). We did not elucidate that FT-IR region (1750-1550 cm^{-1}) had a prominent role in classification of C. roseus cultivars in this study. However, we suggest that FT-IR is an excellent alternative method for investigating the genetic relationship between genotypes of a higher plant species because of its rapidity and simplicity.

Biomarkers for Genetic Discrimination of *C. roseus* Cultivars by Genetic Programming Analysis

Genetic programming (GP) analysis of FT-IR data revealed that biomarkers (variables) contributed most to discrimination of 'Cooler' cultivars from 'Equator' cultivars. GP ranked the top six biomarkers (variables) in order for discrimination of the two kinds of the cultivars: V371 (1593 cm^{-1}), V359 (1569 cm^{-1}), V393 (1226cm^{-1}), V232 (1325 cm^{-1}), V372 (1596 cm^{-1}), and V449 (1743 cm^{-1}) (Fig. 3). There have been a few studies that assigned these peaks. Three peaks were in the region between 1596 and 1569 cm^{-1}, which are assigned to the amide II vibration (1600 -1500 cm^{-1}) of proteins (Kota et al. 2002). The peaks at 1580–1600 cm^{-1} reflect the aromatic C=C stretch in simple and polycyclic aromatics, including quinones (Breton et al. 1994; Breton 1997). A peak was also the region at around 1350 cm^{-1}, which is assigned to the amide III (1400-1200 cm^{-1}, corresponding to NH vibration modes) (Krimm and Bandekar 1986). Considering these results, amide II and III vibrations were deconvoluted from the variables as the most possible compounds for discrimination of 'Cooler' cultivars from 'Equator' cultivars. Thus, differences these regions of FT-IR spectrum can provide the basis for genetic discrimination of C. roseus cultivars. We have previously reported that pyrolysis mass spectrometry (PyMS) discriminates plants phylogenetically in a similar manner (Kim et al. 2003). However, biomarkers determined by GP of FT-IR data cannot be deconvoluted to pure chemical compounds in this study. Also, it remains to be confirmed by other approaches whether this convolution is valid. Fingerprints based on FT-IR of whole cell extracts of plant samples represent relatedness between sums of chemical compounds comprising the samples. A hierarchical dendrogram dichotomically visualizes the relatedness. Therefore, we conclude that the hierarchical dendrogram based on PCA of the FT-IR data represents the most probable chemotaxonomical relationship between cultivars, which is in general agreement with the genetic relationship determined by conventional DNA fingerprinting methods.

Acknowledgements

This work was supported by a grant to JRL from the Crop Functional Genomics Center of the 21st Century Frontier Research Program, a grant to JRL from the Korea Science and Engineering Foundation through the Plant Metabolism Research Center of the Kyung Hee University funded by the Korea Ministry of Science and Technology, a grant to JRL from Studies on Molecular Genomics of Marine and Extreme Organisms funded by the Korean Ministry of Marine Affairs and Fisheries, a grant to JRL from the BioGreen 21 Program funded by the Korean Rural Development Agency, and a grant to JRL from KRIBB Research Initiative Program.

References

Breton J, Burie JR, Boullais C, Berger G, Nabedryk E (1994) Binding sites of quinones in photosynthetic bacterial reaction centers investigated by light-induced FTIR difference spectroscopy: binding of chainless symmetrical quinones to the QA site of Rhodobacter sphaeroides. Biochemistry 33: 12405-12415

Breton J (1997) Efficient exchange of the primary quinone acceptor Q(A)in isolated reaction centers of Rhodopseudomonas viridis. Proc Natl Acad Sci 94: 11318-11323

Chen L, Carpita NC, Reiter WD, Wilson RH, Jeffries C, McCann MC (1998) A rapid method to screen for cell-wall mutants using discriminant analysis of fourier transformation infrared spectra. Plant J 16: 385-392

Goodacre R, Timmins M, Burton R, Kaderbhai N, Woodward AM, Kell DB, Rooney PJ (1998) Rapid identification of urinary tract infection bacteria using hyperspectral whole-organism fingerprinting and artificial neural networks. Microbiology 144: 1157-1170

Kim SW, Ban SH, Chung H, Cho SH, Chung HJ, Choi PS, Yoo OJ, Liu JR (2004) Taxonomic discrimination of higher plants by multivariate analysis of Fourier transform infrared spectroscopy data. Plant Cell Rep 23: 246-250

Kim SW, Ban SH, Jeong SC, Chung HJ, Ko S, Yoo OJ, Liu JR (2007) Genetic discrimination between Catharanthus roseus cultivars by metabolite fingerprinting using ^{1}H-NMR spectra of aromatic compounds. Biotech Bioprocess Eng (submitted)

Kota Z, Horvath LI, Droppa M, Horvath G, Farkas T, Pali T (2002) Protein assembly and heat stability in developing thylakoid membranes during greening. Proc Natl Acad Sci 99: 12149-12154

Krimm S, Bandekar J (1986) Vibrational spectroscopy and conformation of peptides, polypeptides and proteins. Advances in Protein Chemistry 38: 181-363

Savitzky A, Golay MJE (1964) Smoothing and differentiation of data by simplified least squares procedures. Analytical Chemistry 36: 1627-1639

Stewart D, Yahiaoui N, McDougall GJ, Myton K, Marque C, Boudet AM, Haigh J (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

Timmins EM, Howell SA, Alsberg BK, Noble WC, Goodacre R (1998) Rapid differentiation of closely related Candida species and strains by pyrolysis-mass spectrometry and Fourier transform-infrared spectroscopy. J Clinical Microbiology 36: 367-374

Wenning M, Seiler H, Scherer S (2002) Fourier-transform infrared microspectroscopy, a novel and rapid tool for identification of yeasts. Applied and Environmental Microbiology 68: 4717-4721

Wold H (1966) Estimation of principal components and related models by iterative least squares. In: Krishnaiah, K. R. (Ed.), Multivariate Analysis, Academic Press, New York, pp. 391-420.

(Received July 27, 2007; Accepted August 20, 2007)