Classification of 1-methylcyclopropene treated apples by fluorescence fingerprint using partial least squares discriminant analysis with stepwise selectivity ratio variable selection method

Vipavee Trivittayasil [a,b,1], Mizuki Tsuta [a,*], Satoshi Kasai [c], Yosuke Matsuo [a], Yasuyo Sekiyama [a], Toshihiko Shoji [d], Ryoko Aiyama [a], Mito Kokawa [e], Junichi Sugiyama [a]

[a] *Food Research Institute, National Agriculture and Food Research Organization, Ibaraki, 305-8642, Japan*
[b] *International Research Fellow of the Japan Society for the Promotion of Science, Tokyo, 102-0083, Japan*
[c] *Apple Research Institute, Aomori Prefectural Industrial Technology Research Center, Aomori, 036-0332, Japan*
[d] *Institute of Fruit Tree Science, National Agriculture and Food Research Organization, Ibaraki, 305-8605, Japan*
[e] *Faculty of Life and Environmental Sciences, University of Tsukuba, Ibaraki, 305-8572, Japan*

ARTICLE INFO

ABSTRACT

In this study, we investigated the potential of using fluorescence fingerprint (FF) for nondestructive identification of apples treated with 1-methylcyclopropene (1-MCP). In total, 442 apples of two cultivars (Fuji and Orin) and different storage times (0, 4, 5, 6, and 8 months) were assessed. The classification model used in this study was built using partial least squares discriminant analysis (PLSDA) with the stepwise selectivity ratio (SR) method. The stepwise SR method is a recursive variable selection method proposed in this study. FF was capable of classifying 1-MCP-treated apples with accuracies of 91.23%, 89.74%, and 90.17% for calibration, cross-validation, and validation results, respectively. PLSDA with the stepwise SR method could identify four aggregations of wavelength conditions, which are important to the classification. In addition, a non-targeted approach was taken to screen the metabolites characterizing 1-MCP-treated and control apples by liquid chromatography-mass spectrometry (LC/MS) and nuclear magnetic resonance (NMR) spectroscopy. The observed difference in metabolic profiles may contribute to the difference in the fluorescence profiles of 1-MCP treated and control apples.

1. Introduction

1-Methylcyclopropene (1-MCP) has been extensively studied in the past years in terms of its potential to prolong the freshness of agricultural produces [1,2]. Since its patent approval in 1996 [3], the use of 1-MCP has been attracting interest and has already been approved for use in more than 40 countries by 2011 [4]. One of the most successful applications of 1-MCP is in the apple industry [5]. As a postharvest treatment, 1-MCP has been reported to prevent softening and acidity loss [6], and reduce storage disorders such as senescent breakdown and superficial scald in apples [5]. Harvested apples treated with 1-MCP were reported to retain their characteristics of harvested fruits even after storing for 6 days at 20 °C. As the use of 1-MCP had become widespread with even more growth expected [7], there is a need for a third party to be able to distinguish treated apples from untreated ones. One of the reasons for such need is that 1-MCP is not allowed to be used for organic produces

according to USDA's National List of Allowed and Prohibited Substances [8] thus a screening test would be needed. Another reason is to determine whether the apples are suitable for prolonged storage or not. The currently available method for the determination of 1-MCP treatment is chromatography [9], which can be expensive and time-consuming for screening test. A nondestructive method capable of classifying 1-MCP treated apples for screening purpose at a lower cost and shorter time is therefore desired.

In this study, the capability of fluorescence fingerprint (FF) to nondestructively distinguish 1-MCP-treated apples from untreated ones was investigated. FF, also known as fluorescence excitation–emission matrix, was chosen because of its high sensitivity compared with absorption spectroscopic methods [10]. FF could capture the overall chemical profile, specifically fluorophores, of a measured sample nondestructively. There are several fluorophores present in apples such as chlorophyll [11], procyanidins [12], flavonoids [13], and other

---

* Corresponding author. 2-1-12 Kan-nondai, Tsukuba, Ibaraki 305-8642, Japan.
E-mail address: mizukit@affrc.go.jp (M. Tsuta).
[1] Present address: Cyberagent, Inc., Tokyo, Japan 101-0021.

phenolic compounds [14], giving rise to fluorescence properties in apples. Fluorescence properties have been used as follows. For apples, they are used to monitor changes during maturation, ripening, and senescence [11,15], sort apples by quality [16,17], detect mealiness [18], and assess the flesh firmness of apples [11]. For apple juice, fluorescence properties are used to assess its quality during processing [19], determine the degree of naturalness [20], analyze its browning [21], and detect pesticide residues [22]. As the metabolic profiles of 1-MCP-treated and control apple fruits were reported to diverge from 1 day after the treatment [23], the fluorescence profiles of 1-MCP-treated and control apple fruits were expected to be different. Therefore, there is a potential use of FF for classifying 1-MCP-treated and control apple fruits. The metabolic profiles of the apple fruits used in this study were analyzed by liquid chromatography-mass spectrometry (LC/MS) and nuclear magnetic resonance (NMR) spectroscopy.

The conventional classification method for FF data is partial least squares discriminant analysis (PLSDA). PLSDA can handle both multi-collinearity problem and datasets with the number of variables higher than that of samples. Examples of PLSDA applications to FF data include classification of white wines according to grape variety [24] and characterization and classification of honey [25]. Variables are often selected depending on the purpose such as identifying important variables and improving the model performance [26]. Reducing the variables is especially important for light spectroscopy, because as a shorter measurement time and lower instrumental cost could be achieved by using fewer variables. There are several indices for variable importance developed for partial least squares method: variable of importance (VIP) [27], selectivity ratio (SR) [28], and significant multivariate correlation (sMC) [29]. SR does not have recommended cutoff value when identifying important variables. The threshold value is usually chosen arbitrarily by trial and error. A new approach using a more objective criterion is proposed in this study.

2. Materials and methods

2.1. Samples

Apples of two cultivars, Fuji and Orin, were harvested from an orchard in Aomori prefecture, Japan at the commercial ripening stage on 29 October and 4 November 2014, respectively. A total of 442 apples were used in this study. Within a day after the harvest, approximately half of all the apples of each cultivar were treated with 1 μL/L 1-MCP for 24 h at room temperature by using SmartFresh™ (Agrofresh Inc. Philadelphia, PA, USA) according to the manufacturer's recommendations, and the remaining were treated as control samples.

They were stored at 0 °C until the FF experiment was carried out. Zero degree Celsius was used to simulate the actual storage condition. Immediately after the start of the storage experiment, 80 apples were taken out for FF measurement and were immediately put back into a cooling room after the measurement. This round of initial measurement was labeled as the measurement at 0 month. All apples were separated into groups and the ones belonging to each group were measured for FF after storing for a certain period (4, 5, 6, or 8 months; See Table S1 for the distribution of apples). After the FF measurement, they were immediately frozen in liquid nitrogen and stored at −80 °C until lyophilization. Frozen samples were lyophilized using a vacuum freeze drier (FDU-2110, EYELA, Tokyo, Japan) for 5 days. Lyophilized whole fruit including the skin and pulp was immediately ground in a mechanical mill (Waring blender 7011HS, Osaka Chemical Co.Ltd. Osaka, Japan), and the resulting fine powder was stored at −30 °C until the LC/MS and NMR experiment.

2.2. Fluorescence fingerprint (FF) spectrophotometer

The FF measurement was performed using a fluorescence spectro-photometer (FP-8500; JASCO, Japan) equipped with an epi-fluorescence

unit (EFA-833; JASCO, Japan). The unit enabled the placement of a solid sample such as apples directly on the measurement window. Before the experiment, the apples were brought out of the cooling room to allow their temperature to return to room temperature. This was performed to prevent any unwanted effect on spectra introduced by unstable temperature. The left and right sides of the equator of each apple were measured. An apple was placed such that the desired measurement point covers the measuring window. The experiment was performed inside a small dark room. The excitation wavelength was in the 200–650 nm range and the emission wavelength was 230–750 nm range with 10 nm interval for both excitation and emission. The band slit for both sides was 10 nm. The photomultiplier was set at 310 V with a scanning speed of 60000 nm/min. The visualization or any preprocessing of FF data was performed using EEM package v1.1.1 [30] in R software v3.3.1.

2.3. PLSDA with stepwise variable selection using SR and VIP

PLSDA was carried out to construct a classification model. FF, LC/MS, or NMR data were used as predictor variables and the treatment group (2 classes: control and 1-MCP-treated) were used as response variables. Variable importance was calculated using SR [31] and VIP [26]. In this paper, we proposed using a recursive variable selection method, which is an iterative process, for an objective variable selection.

In the proposed method, a cross–validation (CV) is performed using PLSDA. The variables are subsequently selected on the basis of the variable of importance index, SR or VIP. The methods for identifying important variables based on SR and VIP differed because SR does not have a clearly defined threshold like VIP. For SR, the variables were ranked according to their SR, that is, their importance to the model. After the ranking, a predefined number of variables were retained in each round. The model and subsequent SR were recalculated using the remaining variables. This variable selection process is repeated recursively as many times as defined. The number of iterations depends on how many variables are set to be retained in each round, which is determined manually by the considering the number of total variables. For stepwise VIP, variables with a VIP score >1 are known to be important to the model [26,32]; thus, variables with a VIP score higher than 1 are retained in each round. The number of rounds is determined manually.

2.4. Classification of 1-MCP-treated apples by FF

In this study, 1044 FF data (522 apple FF data × 2 measurement locations) were available. For each sample there were 2438 variables obtained from 46 excitation wavelength × 53 emission wavelength. Three-dimensional FF spectral data were unfolded before PLSDA. This method was previously referred to as unfolded-PLSDA [24]. The data were divided into calibration (60%) and validation (40%) sets. To ensure that the samples of different storage times are equally distributed in both sets, the data were first divided according to storage time before 60% from each subgroup was randomly selected to make up the calibration set. The final calibration group contained the combination of the calibration group of each storage month, while the remaining samples were grouped into the validation group.

The classification model was developed using the calibration group (Fig. 1). CV of 10 random subsets with 20 iterations was employed to determine the parameter, in this case, the number of latent variable (LV) components. The number of LV components was determined by locating a *knee* drop in the scree plot of CV [33]. Several models were built to compare various preprocessing methods, as listed in Table 1, and recursive variable selection methods using VIP and SR. For SR, the numbers of variables retained in each round were predefined as 1000, decreasing by 5 until 200, and decreasing by 1 until 10. For VIP, a total of 6 rounds were calculated. The optimal model was chosen based on the lowest CV error rate. The classification model was built using the PLS Toolbox v8.2 (Eigenvector Inc. Wenatchee, WA, USA) in MATLAB R2015b software (MathWorks Inc. Natick, MA, USA).