Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy 140 (2015) 431–436

Contents lists available at ScienceDirect

Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy

journal homepage: www.elsevier.com/locate/saa

Discrimination of tomatoes bred by spaceflight mutagenesis using visible/near infrared spectroscopy and chemometrics

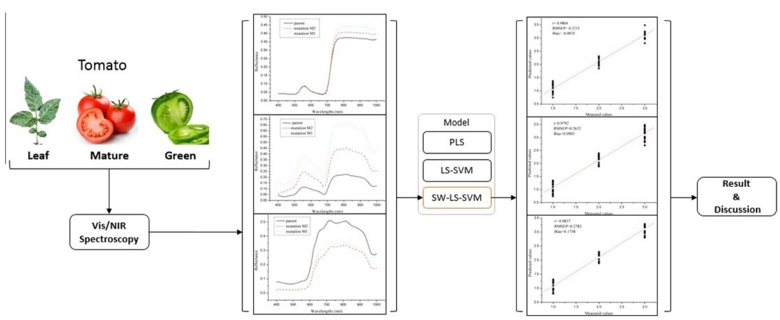
Yongni Shao [a], Chuanqi Xie [a], Linjun Jiang [a], Jiahui Shi [b], Jiajin Zhu [a], Yong He [a,*]

[a] College of Biosystems Engineering and Food Science, Zhejiang University, Hangzhou 310058, China
[b] Zhejiang Sports Science Research Institute, Hangzhou, China

HIGHLIGHTS

- Vis/NIR spectral was used to discriminate tomato breeds by spaceflight breeding from its leaf or fruit.
- Tomato breeds were divided into mutations M_1, mutations M_2 and its parent.
- The SW–LS-SVM were better than PLS and LV–LS-SVM models to predict the tomato breeds.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 13 October 2014
Received in revised form 5 January 2015
Accepted 11 January 2015
Available online 17 January 2015

Keywords:
Vis/near infrared spectroscopy
Tomato
Independent component analysis
Partial least squares analysis
Least squares-support vector machine

ABSTRACT

Visible/near infrared spectroscopy (Vis/NIR) based on sensitive wavelengths (SWs) and chemometrics was proposed to discriminate different tomatoes bred by spaceflight mutagenesis from their leafs or fruits (green or mature). The tomato breeds were mutant M_1, M_2 and their parent. Partial least squares (PLS) analysis and least squares-support vector machine (LS-SVM) were implemented for calibration models. PLS analysis was implemented for calibration models with different wavebands including the visible region (400–700 nm) and the near infrared region (700–1000 nm). The best PLS models were achieved in the visible region for the leaf and green fruit samples and in the near infrared region for the mature fruit samples. Furthermore, different latent variables (4–8 LVs for leafs, 5–9 LVs for green fruits, and 4–9 LVs for mature fruits) were used as inputs of LS-SVM to develop the LV–LS-SVM models with the grid search technique and radial basis function (RBF) kernel. The optimal LV–LS-SVM models were achieved with six LVs for the leaf samples, seven LVs for green fruits, and six LVs for mature fruits, respectively, and they outperformed the PLS models. Moreover, independent component analysis (ICA) was executed to select several SWs based on loading weights. The optimal LS-SVM model was achieved with SWs of 550–560 nm, 562–574 nm, 670–680 nm and 705–715 nm for the leaf samples; 548–556 nm, 559–564 nm, 678–685 nm and 962–974 nm for the green fruit samples; and 712–718 nm, 720–729 nm, 968–978 nm and 820–830 nm for the mature fruit samples. All of them had better performance than PLS and LV–LS-SVM, with the parameters of correlation coefficient (r_p), root mean square error of prediction (RMSEP) and bias of 0.9792, 0.2632 and 0.0901 based on leaf discrimination, 0.9837, 0.2783 and 0.1758 based on green fruit discrimination, 0.9804, 0.2215 and −0.0035 based on mature fruit discrimination, respectively. The overall results indicated that ICA was an effective way for the selection of SWs, and the Vis/NIR combined with LS-SVM models had the capability to predict the different breeds (mutant M_1, mutant M_2 and their parent) of tomatoes from leafs and fruits.

* Corresponding author. Tel./fax: +86 571 88982143.
E-mail address: yhe@zju.edu.cn (Y. He).

http://dx.doi.org/10.1016/j.saa.2015.01.018
1386-1425/

Introduction

Over the past 40 years, breeding by spaceflight mutagenesis has become a new trend of high technology agriculture. China is one of the three countries (China, USA and Russia) which master the spaceflight return technology. From the year 1987, China has successfully conducted spaceflight tests carrying plant seeds. By the effort of researchers, excellent breeds were obtained, demonstrating the obvious advantages of this technique. Spaceflight mutagenesis is a new breeding technology that combines spaceflight, biology and agriculture breeding. Tomato is a kind of vegetable which is also called "fruit". It is rich in nutrient, such as Vitamin C, organic acids, mineral, especially, the lycopene content. It is very popular with consumers because of its nutrient value. With the improvement of our living standards, the breeding and quality of tomato are attracted to many scientists to do the research on. Spaceflight mutagenesis technology has been applied to genetic improvement of process operations of varieties of crops and to realize the genetic improvement of high efficiency. At present, researches on the mechanism of spaceflight mutagenesis breeding are still in a preliminary stage.

Wang et al. analyzed tomato seeds by space flight mutagenesis using FT-IR. The results showed that the space environment could be used to apply the stretching vibration of C—O on the structure of carbohydrates in the space tomato seeds [1]. Guo et al. used X-ray fluorescence analysis to measure and analyze the elements of the fourth generation radix isatidis with space flight mutagenesis breeding [2]. In our early study, we used an analysis method of visible/near infrared spectroscopy (Vis/NIRS) combined with chemometrics to discriminate breeds of tomato by spaceflight mutagenesis from the green fruits [3]. In this study, we extend our samples to tomato leafs and fruits (green or mature), and try to use Vis/NIR to find the variation characteristics of spectral, and compare the discrimination capacity between them.

Near infrared spectroscopy (NIR), utilizing the spectral range from 780 to 2500 nm, has served for the past 30 years as a method to predict the quality of different foods and agricultural products due to its speedy analysis, little sample preparation requirement and low cost [4]. In our previous study, we used Vis/NIRS to analyze the physiological properties of tomatoes, including firmness, soluble solids content and the pH value [5]. Xu et al. used near infrared spectroscopy in detecting minor damages to tomato leafs, and showed that the sensitive bands of 1450 and 1900 nm modeled with severity level provided the highest correlation coefficient [6]. Pedro and Ferreira adopted near infrared spectroscopy to estimate solids and carotenoids in tomatoes, and the best models presented prediction results with RMSEP and *r* value of 0.4157, 0.9998 for total solids, 0.6333, 0.9996 for soluble solids, 21.5779, 0.9996 for lycopene, and 0.7296, 0.9981 for β-carotene [7]. Kamil et al. also evaluated tomato products (lycopene, β-carotene, starch, allura red pigment and paprika) based on the Fourier transformer infrared spectroscopy [8]. Sirisomboon et al. used near infrared spectroscopy to classify the maturity level and to predict textural properties of tomato variety "Momotaro". It showed that for the tomato maturity classification, the distinguish ability of 100.00% was obtained for red and pink tomatoes, and 96.85% for mature green tomatoes [9].

Various calibration methods have been used to relate near-infrared spectra with measured properties of materials. Principal components regression (PCR), partial least squares (PLS), multiple linear regression (MLR) and artificial neural networks (ANN) are the most used multivariate calibration techniques for NIRS [10]. PLS was usually considered for a large number of applications in food analysis and was widely used in multivariate calibration. Because it takes the advantage of the correlation relationship that already exists between the spectral data and the constituent concentrations. However PLS is based on linear models and unsatisfactory results may be obtained when non-linearity is present.

Least-squares support vector machine (LS-SVM) could handle the linear and nonlinear relationships between the spectra and response chemical constituents [11,12]. Therefore, a new combination of ICA with LS-SVM was proposed as a nonlinear calibration model for quantitative analysis using spectroscopic techniques. In our early study, we had already used this method to analyze irradiated rice with different irradiation doses [13].

In this study, we try to use Vis/NIR to find the variation characteristics of spectral for the tomato bred by spaceflight mutagenesis. The objectives of this paper were (1) to study the feasibility of using Vis/NIR spectroscopy to predict the different breeds (mutant M_1, mutant M_2 and their parent) of spaceflight tomatoes from its leafs or fruits (green or mature); (2) to compare the prediction precision of using different latent variables (4–9 LVs) for least squares-support vector machine (LS-SVM); and (3) to select the optimal sensitive wavelengths (SWs) for the development of portable instruments and online monitoring for commercial discrimination of different breeds of tomatoes.

Experiment

Vis/NIR analysis

Spectra were collected using a Vis/NIR scanning spectroradiometer (ASD Handheld FieldSpec, Boulder, USA) in reflectance mode. Measurements were made at ambient temperature (18–20 °C) over the wavelength range of 325–1075 nm at intervals of 1.5 nm. Reflection measurements were taken for each tomato plant with its leafs, green and mature fruits. A Lowell pro-lamp interior light source (Assembly/128930) with the Lowell pro-lamp 14.5 V Bulb (128690 tungsten halogen made in China), which could be used both in the visible and near infrared regions, was placed at a distance of 300 mm from the sample surface. For the leaf measurement, the reflectance spectra were taken at the center position of each leaf, and the scan number for each spectrum was set to 30. For the fruit measurement, the reflectance spectra were taken at the position around the equator for each fruit, and the scan number for each spectrum was also set to 30. The signals were preprocessed using View-Spec Pro V2.14 (Analytical Spectral Device, Inc., Boulder, CO 80301). Due to the scattering noises of the collection system, only spectral data from the wavelength of 400–1000 nm were used. In the following regression, the wavelengths were divided into the visible region (400–700 nm) and the near infrared region (700–1000 nm). In the trial for tomato leaf samples, a total of 150 samples were prepared, which were breeds of mutant M_1, mutant M_2 and their parent, with 50 samples for each breed. When building the model, they were randomly divided into calibration sets of 105 samples (35 samples for each breed) and prediction sets of 45 samples (15 samples for each breed). The same sample numbers and testing scheme were applied for the green and mature fruits of tomatoes.

Spectral preprocessing

The spectra data was preprocessed before the calibration stage. The Savitzky–Golay smoothing was used to reduce the noise [14,15], with a window width of 7 (3-1-3) points. The multiplicative scatter correction (MSC) was used to correct additive and multiplicative effects in the spectra [16].

Partial least squares analysis

In the development of the PLS model, calibration models were built between the spectra and the tomato leafs or fruits, full