Food Chemistry 141 (2013) 259–265

Contents lists available at SciVerse ScienceDirect

Food Chemistry

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

Analytical Methods

Discrimination of oolong tea (*Camellia sinensis*) varieties based on feature extraction and selection from aromatic profiles analysed by HS-SPME/GC–MS

Jie Lin, Pan Zhang, Zhiqiang Pan, Hairong Xu, Yaoping Luo, Xiaochang Wang *

Institute of Tea Science, Zhejiang University, 866 Yuhangtang Road, Hangzhou, Zhejiang 310058, PR China

A R T I C L E I N F O

Article history:
Received 29 January 2012
Received in revised form 15 February 2013
Accepted 23 February 2013
Available online 14 March 2013

Keywords:
Oolong tea
Economic adulteration
Variety
Volatile compounds
HS-SPME/GC–MS
Chemometric analysis

A B S T R A C T

This study aimed to develop an objective and accurate analytical method to discriminate oolong tea varieties that easily causing adulteration by potential volatile compounds. A total of 75 oolong tea samples of five similar varieties (Tieguanyin, Benshan, Maoxie, Huangjingui and Jinguanyin) were analysed by headspace solid phase microextraction (HS-SPME) coupled with gas chromatography–mass spectrometry (GC–MS). The relative content of 26 major volatile compounds varied significantly according to variety, combined with the results of hierarchical cluster analysis (HCA), indicating that the varietal differences of aromatic profile remain significant for tea cultivars with very close origin. Principal component analysis (PCA) of the aromatic profiles showed that the feature of variety dominated over the other features (like producing region and quality). By stepwise linear discriminant analysis (S-LDA), 18 volatiles with the best discriminating capacity were selected, and 4 discriminant functions (DFs) enabled simultaneously discrimination of the five oolong varieties with 100% correct rate.

1. Introduction

Oolong tea is one kind of partially fermented teas and can be produced with young green shoots of different tea varieties. Under the same processing procedures (including plucking, sun withering, cooling, tossing, withering, fixation, rolling and drying), some oolong varieties may have quite similar appearances and flavour characteristics, which make them quite difficult to differentiate. For this reason, these similar varieties can easily cause economic adulteration, and expensive teas are substituted with cheaper or inferior teas so as to defraud the consumers and get a higher profit.

Economic adulteration includes use of tea from other geographical origins, lower quality grades, reduced or extended ageing, substitutes and similar but less valued varieties. Various papers have reported the authentications of geographical origin (Borse, Jagan Mohan Rao, Nagalakshmi, & Krishnamurthy, 2002; Fernández, Pablos, Martín, & González, 2002; Fernández-Cáceres, Martín, Pablos, & González, 2001; He et al., 2009; Kim, Scotter, Voyiagis, & Hall, 1997; Pligrim, Watling, & Grice, 2010; Seetohul, Islam, O'Hare, & Ali, 2006), quality grade (Gall, Colquhoun, & Defernez, 2004; He et al., 2009; Pongsuwan et al., 2007) and ageing times (Yu, Wang, & Wang, 2009) of tea. Besides, classifications of black, green, oolong, white, jasmine and Pu-erh teas, have also been investigated by various researchers (Alcázar et al., 2007; Chen et al., 2007; Fernández-Cáceres et al., 2001; Liu, van Espen, Adams, Yan, & Vanbelle, 1987; Togari, Kobayashi, & Aishima, 1995;Wang et al., 2008); however, from a practical point of view, these tea categories can be easily differentiated by their appearances (colour or shapes), which limits the use of these methods. We know of no relevant research to detect economic adulteration caused by tea varieties based on their chemical composition.

Variety authentication is important for oolong tea because some fine oolong varieties (e.g., Tieguanyin, Dahongpao, whose geographical origins are protected by government regulations) are sold at a premium price in the market compared to other inferior varieties. Some oolong varieties are so similar in appearances and flavour that correct differentiation is only possible for experts or experienced tea tasters. Conventional chemical methods are not sufficient for identification because oolong varieties have close varietal origins. Thus, there is significant interest in developing accurate methods to discriminate these varieties.

For this purpose, we are striving to develop an accurate and robust technique based on HS-SPME/GC–MS and chemometrics to discriminate oolong varieties by potential volatile compounds. HS-SPME has been proved to be a fast, simple and convenient sampling method (Augusto & Zini, 2002; Lv et al., 2012), and also has been successfully applied to differentiation studies on wine varieties (Zhang et al., 2010), wine types (Pizarro, Esteban-Díez, Sáenz-González, & González-Sáiz, 2008), apple categories (Abrodo

* Corresponding author. Tel./fax: +86 571 88982380.
E-mail address: xcwang@zju.edu.cn (X. Wang).

0308-8146/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.foodchem.2013.02.128

et al., 2010) and different maturation and ageing of Balsamic Vinegars of Modena (Cirlini, Caligiani, Pall, & Palla, 2011). Chemometric methods, like One-Way ANOVA, HCA, PCA and stepwise LDA, were applied to significance testing, data visualisation and feature extraction & selection, so as to extract meaningful attributes from aromatic profiles and construct discriminant model for oolong varieties.

2. Materials and methods

2.1. Tea samples

A total of 75 oolong tea samples produced in the autumn of 2010, and sorted by five varieties (Tieguanyin, Benshan, Maoxie, Huangjingui and Jinguanyin) were collected from reliable sources of Anxi County or Longquan County, as described in Table 1. For each variety, 15 samples with different prices and from 3–5 producing regions or sources were analysed. All Jinguanyin samples were collect from the 3 largest tea companies in Longquan County of Zhejiang Province. The five oolong varieties all have loose ball shapes, and are particularly difficult to differentiate by appearances. Adulteration normally happens when Benshan, Maoxie or Huangjingui are mixed in or sold as Tieguanyin, because Tieguanyin of good quality has a premium price in the market and its geographical origin is protected by government regulation as well. In this study, 12 Tieguanyin samples were collected at prices of RMB 500–1400/kg, whereas all Benshan, Maoxie and Huangjingui samples were collected at prices only of RMB 200–400/kg (Table 1). All the tea samples were kept in aluminum foil bag and stored in dark at 4 °C until analysis.

2.2. Headspace solid-phase microextraction (HS-SPME) analysis

Ten grams of dry tea sample was transferred to a 100 ml glass septum flask, and SPME fibre coated with 65 μm polydimethylsiloxane/divinylbenzene (PDMS/DVB) was rapidly inserted into the headspace of the flask. The commercially available SPME fibre and the manual SPME holder (SAAB-57330U) were from Supelco (Bellefonte, PA, USA). The extraction was kept in a water bath at 50 °C for 40 min. The PDMS/DVB fibre was preconditioned for 5 min in the injection port of the GC at 220 °C before each extraction, and the experiment was carried out in duplicate for each tea sample.

2.3. Gas chromatography–mass spectrometry (GC–MS) analysis

An Agilent 6890 gas chromatography coupled with an Agilent 5973N mass spectrometry was used to perform the analysis. A HP-INNOWax capillary column (30 m × 0.25 mm × 0.25 μm; Agilent Technologies, USA) was equipped, with purified Helium as the carrier gas, at a constant flow rate of 1 ml/min. The oven temperature was held at 50 °C for 5 min and then increased to 220 °C at a rate of 3 °C/min, and held at 220 °C for 5 min and then increased to 240 °C at a rate of 10 °C/min, and finally held at 240 °C for 5 min. The injector and ion source temperature were set at 220 and 200 °C respectively, and MS was scanned at 70 eV over 35–600 a.m.u. The mass spectrometer was operated in the full scan, and the area of each peak was determined by ChemStation software (Agilent Technologies).

2.4. Data analysis

The relative percentages of the detected peaks were obtained by peak-area normalisation, all relative response factors being taken as one. The alignment of peaks was performed by comparison of their retention time and mass fragments. Afterward, the peak selection for the subsequent statistical analysis was performed based on two criteria: commonly existing in all 75 samples and having an average relative content more than 0.5%. Aligned peaks (26) were finally selected and imported to form a data set. In order to get the Kovats' retention indices for each peak, 1 μl *n*-alkane mixture (C_8 to C_{20}; Sigma–Aldrich) was injected under the same GC conditions. Identification of the selected peaks was made by searching NIST98 MS data library (a match quality of 95% minimum was used as a criterion) and comparison of their Kovats indices (KI) with the literature (Christensen, Jakobsen, Kristiansen, & Møller, 1997; Demirci, Baser, Tabanca, & Wedge, 2006; Demirci, Demirci, & Başer, 2003; Kim, Cho, Ahn, Cho, & Cha, 2005; Petersen, Poll, & Larsen, 1998; Tabanca, Kirimer, Demirci, Demirci, & Başer, 2001; Tabanca et al., 2006; Vian, Fernandez, Visinoni, & Chemat, 2008).

The data set was processed by different pattern recognition methods, such as principal component analysis (PCA), hierarchical cluster analysis (HCA) and stepwise linear discriminant analysis (S-LDA). In order to eliminate the effect of the different size of the variables, the raw data set was automatic scaled before analysis. PCA was performed by SIMCA-P+ (version 12.0, Umetrics, Umeå, Sweden). HCA, S-LDA and One-Way ANOVA were processed by SPSS statistical package (Version 16.0 for Window, SPSS Inc. Chicage, IL, 2007).

PCA (Tipping & Bishop, 1999) was put forward in order to have a better visualisation of all the information contained in the data set. By projecting the objects of the data set into the space of the first few components, it is possible to visualise the differences among the various objects.

HCA (Romesburg, 1984) is a statistical method for finding relatively homogeneous clusters of cases based on measured characteristics. It allows visualisation of the intrinsic structure of the data set without a priori assumption about the origin of the samples. In our study, HCA was defined by Euclidean distances and Ward's linkage method using the whole data set.

S-LDA (Lachenbruch, 1979) was carried out in our study to extract best discriminant features which separate the five classes of oolong varieties. The program enters one feature or removes features in alternate steps by analysing their effects on the separations of the five groups based on the Wilks' lambda criterion. In this work, the *F* values to enter and to remove were set at 2.8 and 1.8, respectively. For a five-class problem, four canonical dis-

Table 1
Description of tea samples, including sample numbers, prices in Chinese yuan (RMB) and producing region.

Variety	Label	Number	Prices (RMB/kg)	Producing region
Tieguanyin	T	15	140–1400	Gande, Jiandou, Longjuan and Xianghua of Anxi County, Fujian Province
Benshan	B	15	200–400	Jingu, Shangqing, Xiping, Guanqiao and Huqiu of Anxi County, Fujian Province
Maoxie	M	15	200–300	Xiping, Huqiu, Daping and Longmen of Anxi County, Fujian Province
Huangjingui	H	15	200–340	Jingu, Xiping and Huqiu of Anxi County, Fujian Province
Jinguanyin	J	15	300–800	Longquan County, Zhejiang Province
Total		75		