Journal of Chromatography A, 1499 (2017) 78–89

Contents lists available at ScienceDirect

Journal of Chromatography A

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

Strategy for comparative untargeted metabolomics reveals honey markers of different floral and geographic origins using ultrahigh-performance liquid chromatography-hybrid quadrupole-orbitrap mass spectrometry

Yi Li[a,b,c,1], Yue Jin[a,b,1], Shupeng Yang[a,b,d], Wenwen Zhang[a], Jinzhen Zhang[a,b,e], Wen Zhao[a,b,e], Lanzhen Chen[a,b], Yaqin Wen[a], Yongxin Zhang[f], Kaizhi Lu[g], Yaping Zhang[g], Jinhui Zhou[a,b,d,e,*], Shuming Yang[c,**]

[a] Institute of Apicultural Research, Chinese Academy of Agricultural Sciences, Beijing 100093, PR China
[b] Laboratory of Risk Assessment for Quality and Safety of Bee Products, Ministry of Agriculture, Beijing 100093, PR China
[c] Institute of Quality Standard and Testing Technology for Agro-Products, Chinese Academy of Agricultural Sciences, Beijing 100081, PR China
[d] Key Laboratory of Bee Products for Quality and Safety Control, Ministry of Agriculture, Beijing 100093, PR China
[e] Bee Product Quality Supervision and Testing Centre, Ministry of Agriculture, Beijing 100093, PR China
[f] Institute of Chinese Materia Medica, China Academy of Chinese Medical Sciences, Beijing 100700, PR China
[g] Thermo Fisher Scientific Co., Ltd., Shanghai 201206, PR China

ARTICLE INFO

Article history:
Received 7 December 2016
Received in revised form 1 March 2017
Accepted 25 March 2017
Available online 28 March 2017

Keywords:
Comparative untargeted metabolomics
Honey
Markers
Floral origin
Geographic origin

ABSTRACT

Honey discrimination based on floral and geographic origins is limited by the ability to determine reliable markers because developing hypothetical substances in advance considerably limits the throughput of metabolomics studies. Here, we present a novel approach to screen and elucidate honey markers based on comparative untargeted metabolomics using ultrahigh-performance liquid chromatography-hybrid quadrupole-orbitrap mass spectrometry (UHPLC-Q-Orbitrap). To reduce metabolite information losses during sample preparation, the honey samples were dissolved in water and centrifuged to remove insoluble particles prior to UHPLC-Q-Orbitrap analysis in positive and negative electrospray ionization modes. The data were pretreated using background subtraction, chromatographic peak extraction, normalization, transformation and scaling to remove interferences from unwanted biases and variance in the experimental data. The pretreated data were further processed using principal component analysis (PCA) and a three-stage approach (*t*-test, volcano plot and variable importance in projection (VIP) plot) to ensure marker authenticity. A correlation between the molecular and fragment ions with a mass accuracy of less than 1.0 ppm was used to annotate and elucidate the marker structures, and the marker responses in real samples were used to confirm the effectiveness of the honey discrimination. Moreover, we evaluated the data quality using blank and quality control (QC) samples based on PCA clustering, retention times, normalized levels and peak areas. This strategy will help guide standardized, comparative untargeted metabolomics studies of honey and other agro-products from different floral and geographic origins.

1. Introduction

Metabolomics, as a holistic approach, is principally used to evaluate metabolite changes in samples of different pathological or physiological conditions and to further evaluate the biological role or function of a biological system at a specific developmental stage. Thousands of metabolites can be analyzed in a biological sample using metabolomics technology [1,2], which has been used for marker exploration and to study diseases [3–8], environ-

* Corresponding author at: Institute of Apicultural Research, Chinese Academy of Agricultural Sciences, Beijing 100093, PR China.
** Corresponding author.
E-mail addresses: zhoujinhui@caas.cn (J. Zhou), yangshumingcaas@sina.com (S. Yang).
[1] Yi Li and Yue Jin contributed equally to this work.

http://dx.doi.org/10.1016/j.chroma.2017.03.071
0021-9673/

ment changes [9], food nutrition [10,11], plant metabolites [12], and clinical treatments [13]. However, information on the use of metabolomics for traceability discrimination of honey is limited.

In recent years, consumers have shown an increasing interest in honey because of its health-promoting effects [14]. Honey composition is closely associated with its floral and geographic origins, which can be influenced by the soil and climate characteristics of the production area [14,15]. Mono-floral or mono-geographic honey is usually more valuable than mixed honey because it provides a choice in the distinct flavor and quality attributes. In the supply and distribution chain, the quality and prices of honey are determined by its floral and geographic origins, and consumers focus more on true-to-label honey with specific origins [16,17]. Therefore, the discrimination of honey with different floral and geographic origins is of great importance for consumers.

Specific markers can be used to simply and conveniently classify honey traceability. Previously developed methods involved in the discrimination of honey were based on target metabolites identified as potential markers. Endogenous compounds such as flavonoids [17–19], phenolic compounds [20,21], carbohydrates [22,23], minerals and trace elements [24–26], volatile compounds [27–29], organic acids [30,31], and amino acids [32,33] as well as other specific compounds, including abscisic acid [21,34–37], hesperetin [38], methyl anthranilate [38] and 3-aminoacetophenone [39], have been used to discriminate the geographic and botanical origin of honey with multiple-instrument technologies. However, honey is a complex natural product produced by honeybees, and it comprises both large amounts and varieties of endogenous compounds depending on the species of nectariferous plants and the geographic area [40,41]. Therefore, finding reliable markers from a class of hypothetical substances that represent different floral and geographic origins using a target metabolite analysis is difficult.

Exploring specific markers using comprehensive and global untargeted metabolomics can address this problem. Untargeted metabolomics yields information on as many metabolites as possible in a complex sample via analysis of all the information in the dataset. Untargeted metabolite profiling has been widely used to identify markers via analysis of a large amount of data from different samples under different conditions using multivariate statistics and library matching [42,43].

Some approaches, including nuclear magnetic resonance (NMR) spectroscopy and high-resolution mass spectrometry (HRMS), have been reported as the main analytical techniques used in untargeted metabolic profiling to evaluate the changes in metabolite levels [1,44–46]. The main advantage of NMR is that it yields a massive amount of data through nondestructive sample measurements [1]. More recently, liquid chromatography (LC) or gas chromatography (GC) coupled with MS has been widely used to identify metabolites in untargeted analyses because the techniques can resolve individual chemical components with complex compositions into separate peaks in the front-end separation systems and yield accurate mass information in the rear-end MS system [47].

Herein, we describe a promising protocol to efficiently analyze comparative untargeted metabolomics data obtained from honey samples using an UHPLC-Q-Orbitrap, which presents an analytical challenge for metabolic profiling. Honey samples with different floral and geographic origins were subjected to simple sample preparation, UHPLC separation and Q-Orbitrap analysis, and the obtained data underwent pretreatment, processing, high-throughput identification of metabolite markers, structural elucidation and marker validation. This method can be extended to screen and identify markers in massive amounts of data to discriminate honey samples and other agro-products with different floral and geographic origins.

2. Material and methods

2.1. Reagents and chemicals

LC–MS-grade acetonitrile and formic acid were purchased from Fisher Scientific (Pittsburgh, PA, USA). Purified water from a Milli-Q purification system (Millipore, Bedford, MA, USA) was used throughout the experiments.

2.2. Sample collection

Selecting typical and authentic honey samples is essential to obtain effective data for marker screening using untargeted metabolomics. Honey samples with specific floral and geographic origins were collected during the blossom period, and no other nectar plants existed within the range of the honeybee flight path. Two hundred and ten raw honey samples were collected from apiaries in different provinces of China between March and September 2014 by our researchers to ensure the authenticity of the botanical and geographic origin. Of these samples, 150 Litchi honey samples were collected from the main producing areas in the provinces of Yunnan (28 samples), Hainan (29 samples), Fujian (29 samples), Guangdong (33 samples) and Guangxi (31 samples), and 60 acacia honey samples were collected from the main producing areas in the provinces of Liaoning (27 samples) and Shaanxi (33 samples). Voucher specimens were deposited at the Institute of Apicultural Research, and samples were stored at 20 °C in a well-ventilated and dark room to avoid further metabolite degradation. If the honey crystallized prior to analysis, it was heated to homogeneity in a water bath of less than 50 °C. To improve the homogeneity of the honey samples, 3–6 samples with the same floral and geographic origins were mixed into a single sample that was representative of the typical characteristics of a specific origin. In total, there were 5 typical samples with the same floral origin for each province. The moisture content of the honey was analyzed prior to analysis to ensure an unbiased evaluation of the endogenous substances in the different honey samples.

2.3. Sample preparation

Portions of 1.0 g of the honey samples were weighed into 10 mL polypropylene centrifuge tubes and mixed with 5.0 mL of pure water via shaking in a mixer for 5 min until the samples were completely fluid. The fluid samples were then centrifuged at 10,000 rpm for 10 min to remove any impurity particles. The final solutions were transferred to vials for further analysis via the UHPLC-Q-Orbitrap after filtering the samples through filter membranes with a 0.22 μm pore size.

2.4. Blank control and QC samples

In this study, the pure water used in the sample preparation served as the blank control to verify the cleanliness of the UHPLC-Q-Orbitrap system, and the blank response was subtracted from the response for each sample as the background correction in the data processing. The QC sample was prepared by mixing 1.0 g of a honey sample from each geographic area and each floral origin to homogeneity and subjecting it to the sample preparation procedure described above. The QC sample was used to evaluate the repeatability and robustness of this protocol as an unbiased representation of the whole sample set.

2.5. Instrumentation

The untargeted metabolites were separated in the UHPLC system equipped with a Thermo Scientific Hypersil GOLD C-18 column