Food Research International 100 (2017) 814–821

Contents lists available at ScienceDirect

Food Research International

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

Combination of mass spectrometry-based targeted lipidomics and supervised machine learning algorithms in detecting adulterated admixtures of white rice

Dong Kyu Lim[a,1], Nguyen Phuoc Long[a,1], Changyeun Mo[b], Ziyuan Dong[a], Lingmei Cui[a], Giyoung Kim[b], Sung Won Kwon[a,c,*]

[a] Research Institute of Pharmaceutical Sciences and College of Pharmacy, Seoul National University, Seoul 08826, Republic of Korea
[b] National Institute of Agricultural Sciences, Rural Development Administration, Jeonju 54875, Republic of Korea
[c] Plant Genomics and Breeding Institute, Seoul National University, Seoul 08826, Republic of Korea

ARTICLE INFO

Keywords:
White rice
Adulteration
Discrimination
Targeted lipidomics
Lysophospholipids
Machine learning

ABSTRACT

The mixing of extraneous ingredients with original products is a common adulteration practice in food and herbal medicines. In particular, authenticity of white rice and its corresponding blended products has become a key issue in food industry. Accordingly, our current study aimed to develop and evaluate a novel discrimination method by combining targeted lipidomics with powerful supervised learning methods, and eventually introduce a platform to verify the authenticity of white rice. A total of 30 cultivars were collected, and 330 representative samples of white rice from Korea and China as well as seven mixing ratios were examined. Random forests (RF), support vector machines (SVM) with a radial basis function kernel, C5.0, model averaged neural network, and k-nearest neighbor classifiers were used for the classification. We achieved desired results, and the classifiers effectively differentiated white rice from Korea to blended samples with high prediction accuracy for the contamination ratio as low as five percent. In addition, RF and SVM classifiers were generally superior to and more robust than the other techniques. Our approach demonstrated that the relative differences in lysoGPLs can be successfully utilized to detect the adulterated mixing of white rice originating from different countries. In conclusion, the present study introduces a novel and high-throughput platform that can be applied to authenticate adulterated admixtures from original white rice samples.

1. Introduction

In food and herbal medicines, chemical profiles of agricultural products are dissimilar due to the diversity of the cultivation and production processes (Akula & Ravishankar, 2011). Differences in the constituents have a significant impact on the practice of grading and determining the market values of plant products, which have a profound influence on agricultural, pharmaceutical, and food policies (Nguyen et al., 2016). Malpractice of agricultural product distribution has been a major problem in food industry (Spink & Moyer, 2011). The problem is even more serious when considering blended products. Particularly, the intentional mixing of adulterated or extraneous ingredients with original product is a classical malpractice, but it is difficult to detect owing to the similarities in terms of appearance, taste, and fragrance. Thus, many illegal food distributions are composed of mixed products (Jabeur et al., 2014; Nguyen et al., 2016).

Rice (*Oryza sativa* L.) is a promising target of adulteration resulting from its current global oversupply and declining demand trend (Abdullah, Ito, & Adhana, 2006). In representative Northeast Asian countries, such as Korea, China, and Japan, *japonica* rice is the most commonly consumed rice type. Despite several differences in the genomes originating from numerous genetic modifications to improve the productivity and quality of rice, *japonica*-type rice cultivated in different countries shares remarkably similar phenotypes. This morphological similarity has begun to emerge in illegal distribution, especially in Korea where a new rice tariff policy was promulgated in 2015. Since then, adulterated admixtures of white rice have started to cause serious problems in the marketplace. More seriously, Korean and Chinese frequently consume rice in powdered form (i.e., rice flour), which may facilitate illegal distribution of mixed rice powder. Subjective expert inspection, in this case, does not guarantee acceptable reproducibility. Consequently, a reliable and highly reproducible technique that can

* Corresponding author at: Research Institute of Pharmaceutical Sciences and College of Pharmacy, Seoul National University, Seoul 08826, Republic of Korea.
E-mail address: swkwon@snu.ac.kr (S.W. Kwon).
[1] These authors contributed equally to this work.

http://dx.doi.org/10.1016/j.foodres.2017.08.006
Received 2 May 2017; Received in revised form 2 August 2017; Accepted 2 August 2017
Available online 04 August 2017

effectively authenticate blending products with excellent sensitivity and accuracy is urgently needed.

In recent years, metabolomics and its subfield of lipidomics have evolved rapidly in different research fields and become popular large-scale analytic methods for small molecules (Cubero-Leon, Peñalver, & Maquet, 2014; Wood, 2014). The applications of metabolomics include, but not limited to, disease diagnosis and prognosis, drug discovery, nutrition, and plant biology (Alonso, Marsal, & Julià, 2015). In addition, the remarkable advancements in metabolomics-based food authenticity, contaminant screening, and geographical discrimination have improved global food policies (Ellis, Muhamadali, Allen, Elliott, & Goodacre, 2016). Among the representative methods for identifying and quantifying the metabolites in biological samples, targeted lipidomics specifically profiles and measures different lipid species in a high-throughput manner (Hyötyläinen, Bondia-Pons, & Orešič, 2013). Electrospray ionization-mass spectrometry (ESI-MS) combined with liquid chromatography is frequently used for targeted lipidomics owing to its excellent sensitivity, selectivity, and wide coverage of lipids in plant materials (Cajka & Fiehn, 2015; De Vos et al., 2007). Despite their excellent sensitivity and accuracy, discrimination methods that use a separation step are disadvantageous because they reduce the throughput due to a fairly long analysis time. Direct infusion mass spectrometry (DI-MS), on the other hand, fulfills the need for high-throughput analysis by providing a quick analysis method with excellent sensitivity, selectivity, and precision (Kim et al., 2015).

Regardless of the obvious advantages, the current metabolomics approaches, especially in food and herbal medicines, depend heavily on a limited number of chemometric tools (Putri et al., 2013). Indeed, the most popular chemometric method for metabolomics-based discrimination study is partial least squares-discriminant analysis (PLS-DA) (Gromski et al., 2015; Liland, 2011). Although PLS-DA is well-developed, it is prone to overfitting and sensitive to outliers. PLS-DA also requires a strict validation strategy that is often neglected. Other effective supervised learning methods, such as k-nearest neighbors (kNN), support vector machine (SVM), random forest (RF), and neural network (NNet), however, have not received suitable consideration or been applied (Berrueta, Alonso-Salces, & Héberger, 2007; Chen et al., 2013; Gromski et al., 2015). In recent past, Fernández-Delgado *et al.* conducted an exhaustive examination of the classification performances of 179 classifiers and concluded that RF, SVM with a Gaussian kernel, C5.0, and model averaged NNet were among the most powerful classifiers in terms of their average overall accuracy (Fernández-Delgado, Cernadas, Barro, & Amorim, 2014). Collectively, the greater integration of machine learning classifications and metabolomics as well as lipidomics is necessary to further progress their applications (Blekherman et al., 2011).

We successfully discriminated the geographical origins of different *japonica* rice cultivars cultivated in Korea, China, and Japan using 17 available lysoglycerophospholipids (lysoGPLs) in white rice: 6 lysophosphatidylcholines (lysoPCs), 7 lysophosphatidylethanolamines (lysoPEs), and 4 lysophosphatidylglycerols (lysoPGs) (Lim, Mo, Long, Kim, & Kwon, 2017). We hypothesized that the relative differences in lysoGPLs can be utilized to detect the adulterated mixing of white rice originating from different countries, for instance, white rice from Korea that is blended with white rice from China and *vice versa*. In this study, we combined DI-MS-based targeted lipidomics with some of the most powerful pattern recognition algorithms to detect the adulterated mixed white rice from the original white rice. White rice grown in different years (2014, 2015, 2016) was collected and used in different analyses to demonstrate that there is no significant effect of abiotic factors on the detection of adulterated mixed white rice. Finally, we aimed to focus on the authenticity of white rice from Korea; hence, white rice from Korea was considered to be the original sample and was compared to mixed white rice or white rice from China.

2. Materials and methods

2.1. Materials and reagents

Commercial white rice cultivated in 2014, 2015, and 2016 was purchased from local markets in Korea and China. Upon arrival, the collected samples were freeze-dried in the dark for two days and stored at -70°C until being used to avoid metabolite alterations. Caffeine was purchased from Sigma-Aldrich (St. Louis, MO, USA). Polytetrafluoroethylene (PTFE) syringe filters with a pore size of 0.2 μm were purchased from Advantec (Tokyo, Japan). HPLC grade acetonitrile, isopropanol, and water were purchased from J.T. Baker (Phillsburg, NJ, USA).

2.2. White rice information and preparation

In this study, we examined adulterated white rice that originated from the mixed white rice from China and Korea by artificially preparing mixed rice samples. For convenience, K and C stand for white rice from Korea and white rice from China, respectively. For 2014 white rice (batch A), we initially marked the samples as K1, K2, …, to K30; and C1, C2, …, to C30. Then, a set of three different individual samples was mixed in a random manner to prepare a representative sample, with a total of 30 samples for each country (100% K and 100% C), using our in-house, R-generated, random sequence. In addition, three predefined ratios were composed, and each contained 30 samples: 75% K mixed with 25% C (K/25% C), 50% K mixed with 50% C (K/50% C), and 25% K mixed with 75% C (K/75% C). The predefined ratio mixed samples were prepared by blending one representative sample of white rice from Korea with a representative sample of white rice from China. A total of 100 samples of white rice cultivated in 2015 from each country were prepared by an identical method and predefined ratios (batch B). Furthermore, to achieve a lower limit of detection for K mixed with C, a total of 180 samples from 2016 white rice were also prepared using an identical method with the following additional ratios (each ratio contained 30 samples, batch C): 100% K, 95% K mixed with 5% C (K/5% C), 90% K mixed with 10% C (K/10% C), 85% K mixed with 15% C (K/15% C), 80% K mixed with 20% C (K/20% C), and 75% K mixed with 25% C (K/25% C). Detailed information on 150 samples from batch A, 100 samples from batch B, and 180 samples from batch C is provided in **Table S1** in the Supplementary data.

2.3. Sample preparation

The optimized extraction method for lysoGPLs extraction in white rice was applied as previously described (Lim, Long, et al., 2017; Liu et al., 2014). In brief, the dried white rice samples were pulverized, precisely weighed (150 mg), and mixed with 1 mg of caffeine. Caffeine can be considered as an internal standard to evaluate the repeativity of DI-multiple reaction monitoring (MRM) results during analysis. The mixture was extracted with 6 mL of 75% isopropanol and sonicated for 2 h at 100 °C. Afterward, the crude extract was centrifuged at 12,000 rpm for 5 min at room temperature. Finally, the extract was filtered through a 0.2 μm PTFE filter, and the supernatant was collected for the analysis. The sequence of the process was randomized to minimize the bias caused by differences in the order the sample.

2.4. Instrument parameters

An Agilent triple-quadrupole MS 6460 system (Agilent, CA, USA) with the ESI ion source was employed for the experiments. The analyses were conducted in positive ion mode for lysoPCs and negative ion mode for lysoPEs and lysoPGs. The settings were the same as those used in our previous paper. The DI-MRM-MS analysis time was approximately one minute for each sample with two ion modes. In brief, a 50% acetonitrile flow of 0.2 mL/min was maintained to reduce the contamination of the