Non-destructive Analysis of Volatile Compounds for Geographical Identification of Rice

Shengying Hu
Heilongjiang University
Hongbo Ren
Heilongjiang Academy of Agricultural Sciences
Song Yong
Heilongjiang University
Siyuan Gao
Heilongjiang University
Li Meng (mengli198026@126.com)
Heilongjiang University https://orcid.org/0000-0003-0724-4094

Original article

Keywords: Rice, Volatile compounds, GC-MS, Geographical identification

Posted Date: April 2nd, 2020

DOI: https://doi.org/10.21203/rs.3.rs-20291/v1

Abstract

Background In recent years, high-quality rice adulteration has become a serious problem. It is essential to prevent false origin labels and dishonest transactions. However, there is still a lack of rapid identification methods for discriminating rice from different sources. In this study, we developed a method to profile volatile organic compounds (VOCs) using headspace solid phase microextraction (HS-SPME) combined with gas chromatography mass spectrometry (GC-MS). In addition, the identification efficiency of the biomarkers was determined using several multivariate analysis methods.

Results Based on the t-test, fold changes and volcano plots, eight typical biomarkers were used based their differential levels. Among them, 2-acetyl-1-pyrroline (2-AP) is the most important source of aroma in rice flavor. Unsupervised analyses, including principal component analysis (PCA) and Cluster analysis, demonstrated the potential for geographic classification of rice between Wuchang and other regions. In addition, partial least squares discriminant analysis (PLS-DA) yielded a goodness of fit of 0.900, a goodness of prediction of 0.853, and a probability of substitution test of 0.012. Random forest (RF) algorithm further strengthened the discriminating ability of volatile compounds.

Conclusion In short, the current method can quickly distinguish rice from Wu Chang and other regions, and the research method can facilitate controlling the authenticity and quality of rice.

Background

Rice (Oryza sativa L) is the staple food for approximately 3.5 billion people. It is estimated that 480 million metric tons of milled rice is produced annually (Reig-Valiente et al., 2016). People consume rice not just because of the need for food but also predilection. In countries where rice is consumed, grain quality determines its market value. Hence, consumer preference, competitive markets, diverse rice origins and price differences can provide the motivation for deliberate mislabeling and adulteration motivated by economic gain, which not only threatens the credit of traders, but also infringes the right of consumers (Wang, 2017). To prevent the dishonest distribution, a high-throughput discrimination method of high sensitivity and precision is urgently needed.

Rice grain quality depends not only on its genetic background, but also on soil and climatic conditions, as well as agronomic treatments during rice growth and development (Suzuki et al., 2008). The key to the establishment of rice origin is to identify typical and specific biomarkers. Isotopic composition, mineral element content and chemical composition related to the climate, soil and temperature were used as biomarkers in rice tracing (Li et al., 2017; Vemireddy et al., 2015; Pramai et al., 2018). Various factors can be combined to accurately determine the origin and quality of rice. Several stable isotope ratios such as $\delta^{13}C$, $\delta^{15}N$, $\delta^{2}H$, $\delta^{18}O$, $^{87/86}S$ and multielemental concentrations were used to differentiate various rice origins (Emwas et al., 2015). Spectroscopy methods (e.g., near-infrared reflectance spectroscopy (NIR), nuclear magnetic resonance spectroscopy (NMR), Raman spectroscopy (RS)) were also used for the comprehensive comparison of components between different rice cultivars (Nazife et al., 2019). Sensitivity of the spectroscopy methods is relatively low and needs vast data to establish the standard. It is difficult to determine the relationship of elements or spectra with the rice traits that attract the consumers most.

Volatile organic compounds (VOCs) are a source of aroma in rice. The aroma has minute differences in rice cultivars with unique and attractive taste. It is difficult to describe the aroma by words. Therefore, the volatile profiles in rice were used to distinguish rice samples from different sources (Bryant et al., 2011). Volatile substances can be extracted by headspace solid phase microextraction (HS-SPME), which is a micro-extraction technique that does not require sample destruction. Since 1990s, the SPME fibers have been used to analyze volatiles in foods, and their advantages include simplicity, high sensitivity and reproducibility (Li et al., 2018; Lee et al., 2014). Untargeted metabolomics allows unbiased and detailed analysis of compounds, and has been well established using liquid chromatography (LC) or gas chromatography (GC) with mass spectrometry (MS) (Kim et al., 2015). GC-MS is suitable for the separation and detection of primary metabolites with a high repeatability retention index. MS can determine components with high sensitivity by mass-to-mass ratio (m/z) and fragments of specific components. In addition, this method allows screening of compounds directly in a reference database, including the National Institute of Standards and Technology (NIST) library.

Although MS-based non-targeted metabolomics has excellent sensitivity and accuracy in differential analysis of biomarkers (Raro et al., 2007), there are few studies on profiling VOCs for geographical identification of rice. Wu Chang Rice grown in Wu Chang county, Heilongjiang province in the Northeast of China has a high reputation for its attractive aroma, which earned its higher market price than cultivars from other production areas in China. It was rated as a China National Geographical Indication Product. Driven by the

interests, deliberate mislabeling “Wu Chang Rice” and adulteration motivated by economic gain take place frequently. Despite the large differences in the quality and composition of rice, it seems challenging to distinguish the Wu Chang rice aroma and to authenticate the rice origin (Suzuki et al., 2008).

This study aims to address this challenge by developing a validating/authenticating stagey of the Wu Chang Rice based on the VOCs data got by HS-SPME/GC-MS method. We first extracted volatiles from rice using SPME, followed by identification of typical biomarkers using multivariate analysis to evaluate their identification efficiency. The results showed that the HS-SPME/GC-MS method can classify Wu Chang rice from others origin and eight aroma biomarkers of Wu Chang rice were obtained. This strategy is also promising for determining other premium high-value rice from other region.

Results And Discussion

Measured compounds

Wu Chang is famous for its rice quality throughout China. Since Wuyoudao1 was generated successfully through breeding by Tianyongtai and was grown dominantly since 1990s in Wu Chang, the reputation of Wu Chang rice has become more popular than ever. The long growing period and critical climate requirement resulted in high labor cost. Together with the restriction to Wu Chang district, the high quality rice has gained high market value. To prevent mislabeling and illegal distribution of fake Wu Chang rice, a precision method to differentiate Wu Chang rice from other rice cultivars is highly demanded. Our strategy here is to analyze the VOCs in the rice because Wu Chang rice has unique attractive flavor. SPME is an important extraction technique developed by Pawliszyn and coworkers in the early 90 s (Lashgari et al., 2019), and it is particularly suitable for the analysis of volatiles in food and drinks such as seafood and coffee beans. But the VOCs analysis in rice has rarely been reported. In this work, CAR/PDMS extraction head was used. The peaks were shown in the chromatogram in Fig. 1. Twenty-two VOCs were identified in the rice samples, including eight aldehydes, five alcohols, two ketones, two heterocyclic compounds, and five hydrocarbon compounds. The details can be found in Table S1. In addition, some siloxane derivatives accounted for some peaks in the chromatogram due to the PDMS in the SPME fibers (Grimm et al., 2001; Laguerre et al., 2007). Table S2 listed the siloxane derivatives detected in this study.

Chemometric analysis of biomarkers for rice screening

To ensure the accuracy of the results, relative standard deviations (RSD) greater than 30% were removed from all the analyses. After data pre-processing and normalization, single factor and multivariate analyses were used to guide the biomarker selection. T-test, volcano plot, fold change (FC), Empirical Bayesian Analysis of Microarray (EBAM) analyses showed that the concentrations of eight features were significantly different between samples ($p < 0.05$, FDR < 0.1, FC < 0.8 or FC > 1.2) (Fig. S1). The data were then subjected to 10 times cross-validation and 1000 alignment tests to determine whether the different samples could be differentiated. To detect potential markers, we obtained a VIP score to evaluate the importance of the features in the recognition model. In the univariate analysis, the feature with a VIP score greater than 0.8 was regarded as significantly different in the univariate analysis and was considered a candidate biomarker (Kim et al., 2015).

Eight VOCs were screened as potential identification biomarkers for the rice samples from Wu Chang and other regions. These compounds were three aldehydes (heptanal, octanal, (E)-2-decenal), 3 alcohols (1-heptanol, (E)-2-decen-1-ol, 3,7,11-Trimethyl-3-dodecanol), one ketone (3-octene-2-one), and one heterocyclic compound (2-acetyl-1-pyrroline). Details of the selected biomarkers can be found in Table 1. The box plot of the eight discriminatory biomarkers was showed in Fig. 2a. Among them, the concentrations of 2-acetyl-1-pyrroline, heptanal, (E)-2-decen-1-ol, (E)-2-decenal, 3,7,11-trimethyl-3-dodecanol, octanal were significant higher in Wu Chang rice ($P < 0.05$), while the concentrations of 1-heptanol and 3-octene-2-one were significantly lower ($P < 0.05$) in Wu Chang rice compared with rice from other origins. The amount of 2-acetyl-1-pyrroline was significantly higher in Wu Chang rice, and it was reported to be an important chemical trait to differentiate fragrant rice from non-fragrant rice.

Table 1
Characteristics of the 8 typical biomarkers and the statistical analyses

Retention time(min)	Compound name	formula	NIST match(%)	Retention index (RI)		RSD (%)	VIP Score	T-test		FC
				Experiment	Reference			P-value	FDR	
14.06	Heptanal	$C_7H_{14}O$	93	839	844	28.93	2.33	0.032	0.035	2.33
14.76	2- Acetyl-1-pyrroline	C_6H_9NO	96	897	901	5.22	3.41	0.0002	0.004	8.69
16.53	1-Heptanol	$C_7H_{16}O$	92	739	754	27.97	0.99	0.028	0.017	0.76
17.98	Octanal	$C_8H_{16}O$	93	907	941	5.81	0.89	0.043	0.037	1.25
19.35	3-Octen-2-one	$C_8H_{14}O$	92	800	810	27.96	0.98	0.049	0.034	0.59
20.38	(E)-2-Decen-1-ol	$C_{10}H_{20}O$	90	778	790	10.36	1.34	0.016	0.013	1.47
25.42	(E)-2-Decenal	$C_{10}H_{18}O$	91	804	856	28.35	0.88	0.028	0.024	1.24
25.82	3-Dodecanol,3,7,11-trimethyl-	$C_{15}H_{32}O$	90	780	817	19.15	0.89	0.036	0.035	1.22

The role of biomarkers in rice flavor

Flavor is a very important trait of rice, especially in fragrant rice (Huang, 2005). The general aroma group has an oxygen-containing group (carbonyl group, hydroxyl group, aldehyde group, ketone group, ether group, phenoxy group, or ester group) in addition to a nitrogen-containing group, aromatic groups. The compounds contain atoms such as sulfur, phosphorus.

In addition to certain types of aromatic groups, the odor of substances is also related to the carbon chain structure of these volatile components. Unsaturated compounds have stronger aroma than saturated compounds, while double bonds can enhance the intensity of odor, the triple bond has an enhancement and even a pungent odor. Moreover, the branching of carbon bonds, especially the presence of tertiary and secondary carbon atoms, has a significant effect on the aroma, and the number of carbon atoms also contributes to the odor of the substance (Kang et al., 2007). Studies have shown that of the eight selected biomarkers, 2-acetyl-1-pyrroline (2-AP) is the most important source of flavor in rice, followed by aldols and ketones.

The flavor of 2-acetyl-1-pyrroline in rice

2-AP was first discovered by Buttery et al., who ranked the levels of 2-AP in different rice varieties. When fragrant rice is compared with non-fragrance rice, it is easy to distinguish them. Obviously, 2-AP is the most important aroma compound in rice, especially aromatic rice (Buttery et al., 1982). In this study, 0.05 ppm 2-AP aroma was described as popcorn, its threshold in water was 0.1 nL/L, and Schieberle studied its odor threshold in air at 0.02 ng/L (Engel et al., 1999). This extremely low threshold makes it become an important source of food aroma.

It was originally thought that 2-AP was only produced by the Maillard reaction during rice cooking. However, further studies showed that 2-AP was mostly produced by rice plants, although some 2-AP in rice was produced during cooking (Buttery et al., 1982; Engel et al., 1999; Sakthivel et al., 2009). Mahattanatawee et al. studied three fragrant rice varieties, comparing their main aroma active compounds by GC-O (olfactometry) and GC-MS. Hexanal, octanal, 2-AP, (E,E)-2,4-decanedialdehyde, (E)-2-nonanedialdehyde, 4-vinyl-2-methoxy -phenol and hydrazine were identified as aromatic compounds shared by these varieties (Mahattanatawee et al., 2014). In a study, a sensory analysis of 36 rice varieties (including aromatic and non-aromatic) was conducted to develop rice aroma and flavor vocabulary. Among the 18 attributes such as “popcorn”, “Starch”, “nut”, “wax”, etc., 2-AP-based “popcorn” was positively correlated with “butter” and “corn”, and negatively correlated with “earth” and “smoke”(Limpawattana et al., 2010). Mathure et al. used SPME to analyze both fragrant and non-fragrance rice samples and found that 2-AP was also detected in some non-fragrance varieties, but the average concentration was about 10 times higher in the fragrant varieties then non-fragrant rice (Mathure et al., 2014).

2-AP as a natural source of healthy fragrance in fragrant rice can increase appetite, improve human metabolism and improve human immunity. In this study, the correlation of 2-acetyl-1-pyrroline with other volatile materials was analyzed, as shown in Fig. 2c. The results showed that among the 22 VOCs, (E)-2-octenal, (Z)-heptenal, 1-octene-3-ol, dodecane, and heptadecane were negative. The rest of the VOCs were positively correlated. The different concentrations of 2-AP in different varieties also explain the significant differences in the contents of rice samples from Wu Chang and other regions. So it is a typical biomarker for identifying geographical origin.

The flavor of aldols in rice

Aldols are one of the key substances in rice. The carbonyl compound produces a native, rich aroma. Studies by Widjaja et al. showed that aldehydes were odorous compounds found in many plants and foods, especially in fragrant rice (Widjaja et al., 1996). In a study by Frankel et al., aldehyde compounds were obtained by oxidation of free fatty acids, wherein the decomposition products of linoleic acid hydrogen peroxide were hexanal, heptaldehyde and 2-heptanal (Frankel et al., 2012).

Concepcion et al. found that several saturated aldehydes such as valeraldehyde, hexanal, heptaldehyde, octanal, furfural and 2-heptanal were odor generating compounds in rice (Concepcion et al., 2018). Heptanal and octanal give rice a pleasant fresh grassy scent and a light fruit scent. However, if the content is too high, it produces a disgusting rancid taste. Unsaturated aldehydes such as (E)-2-nonenal have a pleasant orange aroma when the concentration is extremely low. The higher concentrations of these aldehyde compounds in aromatic rice indicate their important role in the aroma characteristics of rice.

Alcohol compounds are known by-products of unsaturated fatty acid (UFA) oxidation and are further decomposed by aldehyde compounds. Volatile alcohols produce a milder odor, e.g., the turf tast of 1-heptanol. Most alcohol compounds have higher thresholds, and only higher contents or a non-saturated alcohol such as (E)-2-nonenol and 3,7,11-trimethyl-3-dodecyl alcohol will have a great influence on the flavor.

Flavor of ketone compounds in rice

The source of ketones is mainly from oxidative or thermal degradation of polyunsaturated fatty acids, amino acid degradation or microbial oxidation, As Wilkens et al., reported 2-heptanone as a product of linoleic acid oxidation, Yalayan et al., found that 2,3-butanedione was a product of glucose degradation with a strong buttery aroma (Wilkens et al., 1970; Yaylayan et al., 1999). Similar to 2-AP, ketone compounds give rice a pleasant aroma. For example, Daygon et al., reported that compounds such as 2-heptanone, 2-hexanone, and 3-octene-2-one have fruity aroma and floral aroma in rice, and contribute more to rice flavor (Daygon et al., 2016). In addition, 3-octene-2-one is considered to be the most active ketone in rice, with a rose aroma and excellent long-lasting flavor, leading to the most intense flavor of rice.

Evaluation of biomarker efficiency

From previous studies, 2-AP was effective in identifying the geographical origins of fragrant rice and non-fragrant rice (Mathure et al., 2014). However, the potential of other biomarkers was not known. On this basis, the reliability and validity of the selected marker compounds for source discrimination were evaluated. The distinguishing ability was evaluated in this study using various multivariate analyses (Kyu et al., 2018). Data mining and visualization were performed using PCA and heat map analysis. At the same time, the reliability, significance and superiority of the discriminant models of different structures were tested. In addition, random forest (RF) classifier was also used to test the classification accuracy of the marker group (Lim et al., 2018). Through these analyses, we can use eight biomarkers to distinguish rice samples from different sources.

Tree diagram and heat map analyses

In the (coherent) hierarchical clustering analysis, potential markers are constructed to visualize clustering based on Euclidean distance measurements using the Ward clustering algorithm. Each sample begins as a separate cluster and the algorithm continues to combine them until all samples belong to one cluster (Zhao et al., 2014). As shown in Fig. 3a, according to the Euclidean distance, all samples were divided into two categories, one typical category was composed of rice from Wu Chang area, the other contained rice varieties from the rest areas. The distance between the two types was far, showing the Wu Chang rice was significantly different from other rice varieities. The Euclidean distance between all the samples in the Wu Chang area was less than 5, while the Euclidean distance between the Wu Chang area and the other areas was more than 20, indicating that the Wu Chang rice is significantly different from other rice varieities.

In addition to the tree diagram, heat map analysis was conducted. As shown in Fig. 3b, weak correlation between the variables was represented by blue and white, while the strong correlation was represented by white pink and brown, and the classification results were indicated by green and red. The results showed that the 20 rice samples were divided into two categories according to the contents of various VOCs, which were samples from Wu Chang area and other areas. The first category included 9 samples,these rice samples were highly correlated with the eight iconic compounds, and the content of 2-AP was higher, and the rice flavor was better. The second category consisted of 11 samples which contained less of the eight compounds and the rice flavor was relatively poor.

PCA analyses

PCA as an unsupervised method was used to comprehensively evaluate the volatile substances of different varieties of rice. Using the principle of dimensionality reduction, the original evaluation indicators were replaced by fewer comprehensive indicators, the comprehensive indicators retain the original indicators. The vast majority of information on indicators is irrelevant to each other (Xia et al., 2014). The analysis results showed that the cumulative variance contribution rates of the first five principal components were 33.9%, 26.8%, 11.1%, 8.4%, and 6.7%. The total cumulative variance contribution rate was 86.9%, and the number of principal components was determined to be 5. As shown in Fig. 4a, the use of the first three principal components could basically replace the main information features of rice samples. Nine rice samples in the Wu Chang area were concentrated in one space, and the distribution areas were small and overlapping, and the difference was not significant ($P > 0.05$). In contrast, the spatial distribution of 11 other rice samples were far apart, and the difference was significant ($P < 0.05$). It showed that the PCA analysis could clearly distinguish rice samples from Wu Chang and other regions.

PLS-DA Model

Since each classification requires rigorous validation, cross-validation is currently the method of choice for measuring predictive performance and predicts the reliability of a finite-sample classification model. The variable importance (VIP) in the projection is the weighted sum of the squares of the PLS loading to calculate the feature importance (Kim et al., 2015). As shown in Fig. 4b, the sample aggregation and dispersion can be observed in composition analysis score maps of the rice samples from the 20 different regions. The model generated by PLS-DA clusters the rice samples from the Wu Chang area into one cluster. Further analysis found that the distribution of rice samples in the Wu Chang area indicated that the composition and content of these nine samples were similar, while the other 11 samples were more discrete, indicating a large difference in composition and content. It also showed that the eight biomarkers had the best discrimination performance (fitness R^2 was 0.900, prediction goodness Q^2 was 0.853, the displacement test p value was 0.012).

RF classification model

Random Forest is a supervised learning algorithm suitable for high dimensional data analysis. It uses a set of classification trees, each of which grows by randomly selecting features from the data. With sampling at each branch, the class prediction is based on the majority vote of the whole. Random Forest (RF) also provides other useful information such as OOB (out-of-bag) error, variable importance measurement, and outlier measurement. In this study, we focused on the classification accuracy of complex interactions between the predictors (Cutler et al., 2007). As shown in Fig. 4c, the selected eight biomarkers had excellent classification ability, showing the smallest classification error (the error rate of rice in Wu Chang and other regions was 0).

Conclusion

To authenticate rice from different geographic sources, it is important to select a series of reliable features. In this study, HS-SPME/GC-MS was successfully employed to analyze volatile compounds in rice cultivated in Wu Chang and other areas. Eight typical chemical biomarkers were identified and their relative levels were evaluated by non-target metabolomics. Among them, 2-acetyl-1-pyrroline is the most typical biomarker and has a close correlation with the aroma of rice. The experimental results showed that the eight typical biomarkers combined with multivariate analysis can be used to distinguish Wu Chang rice samples from other regions with high reliability. This work therefore represents an important contribution to the field of non-destructive geographical authentication of rice grain samples, and has broad impacts on agriculture, trade and economy.

Materials And Methods

Materials

Twenty different rice grain samples grown in 2018 were randomly collected from various locations at the harvest. Each sample has three biological replicates. All samples were stored at -80°C until analysis. Details of rice samples were list in Table 2. In addition, headspace vials (with 20 ml volume) were obtained from Agilent, each with a silver aluminum cap and a polytetrafluoroethylene (PTFE)/silicone rubber septum. Solid phase microextraction (SPME) manual injection handle and a 75 μm carboxen/polydimethylsiloxane (CAR/PDMS) extraction head were purchased from SUPELCO (St. Louis, USA).

Table 2
The origins of the 20 analyzed rice samples from Wuchang and other areas

Sample	Origin	Number	Cultivar
Wuchang rice	Changbao	CB	Wuyoudao 1
	Yingchengzi	YCZ	Wuyoudao 1
	Longfengshan	LFS	Wuyoudao 1
	Lalin	LL	Wuyoudao 1
	Minle	ML1	Wuyoudao 1
	Minle	ML2	Wuyoudao 1
	Anjia	AJ1	Wuyoudao 1
	Anjia	AJ2	Wuyoudao 1
	Xiangyang	XY	Wuyoudao 1
Non-wuchang rice	Jiansanjiang	JSJ	Kongyu 131
	Zhaoyuan	ZY	Songjing 22
	Xiangshui	XS	Shangyu 397
	Jilin	JL	Jijing 83
	Panjin	PJ	Liaojing 9
	Sheyang	SY	Huaidao 5
	Hubei	HB	Liangyoupeijiu
	Jiangxi	JX	Jinyou 463
	Henan	HN	Yuandao 109
	Korea	HG	Dongjin
	Thailand	TG	Jasmine 105

Extraction of volatiles from samples (HS-SPME)

Rice grains of 4 gram were accurately weighed and stored in a 20 mL headspace vial. The headspace bottle was heated at 80 ^{o}C for 30 minutes in water bath before a SPME fiber was inserted into the headspace portion of the vial. The SPME fiber was used to absorb VOCs emitted from the rice grains for 60 minutes.

GC-MS analysis

GC-MS analysis was performed using a TRACE GC ULTRA (Thermo) and SQ QUANTUM XLS(Thermo). A DB-5 capillary column (30 m × 0.25 mm × 0.25 μm) column was used. The SPME fiber was inserted into the GC inlet of the GC-MS instrument, pushing out the fiber head, and desorbing at 250 ^{o}C for 5 min. The inlet temperature was 250 ^{o}C, the carrier gas was He with the flow rate of

1 mL/min, and splitless injection was used. The temperature program was: initial temperature 40 ^{o}C for 3 min; then increased to 100 ^{o}C at 5 ^{o}C/min and held for 3 min; Finally, ramping to 250 ^{o}C at 10 ^{o}C/min and held for 4 min. Electron ionization source was used with an electron energy of 70 eV. Ion source temperature was 230 ^{o}C, and mass scan ranged from 40 to 300 m/z in a qualitative full scan mode.

Data processing

The MS spectra were matched to the National Institute of Standards and Technology reference spectra (NIST 08 and Wiley 7). The retention index (RI) values obtained in the study were also compared with the RI values reported in the literature to identify each volatile compound. The compound content was retrieved and analyzed using a peak area normalization method. All the data were obtained from three biological replicates.

Statistical analysis

Prior to statistical analysis, the processed data were further treated using log transformation and Pareto scaling. Fold change (FC), volcano plot, t-test and Empirical Bayesian Analysis of Microarray (EBAM) were applied for individual feature selection. A feature was considered to be differentially expressed when it has a FC < 0.8 or FC > 1.2, p-value < 0.05 and FDR < 0.1. Only the overlapping features between FC, volcano plot, t-test and EBAM were considered to be significant. In addition, hierarchical cluster analysis (HCA) was conducted to explore the variance tendency, examine the natural grouping structure, and visualize the sample grouping patterns. Principal component analysis (PCA), partial least square-differential analysis (PLS-DA), 10-fold cross validation and Heatmap were performed to seek the best discrimination model for the rice samples from Wu Chang and other areas. Finally, all the VOCs with variable importance in the projection (VIP) Score ≥ 0.8 and differentially expressed in the univariate analysis were considered to be potential candidates for the discrimination of rice samples from Wu Chang and other areas. The analyses were conducted using Metaboanalyst 4.0.

Abbreviations

VOCs, volatile organic compounds; HS-SPME, headspace solid phase microextraction; GC-MS, gas chromatography mass spectrometry; 2-AP, 2-acetyl-1-pyrroline; HCA, hierarchical cluster analysis; PCA, principal component analysis; PLS-DA, partial least squares discriminant analysis; RF, Random forest; NIR, near-infrared reflectance spectroscopy; NMR, nuclear magnetic resonance spectroscopy; RS, raman spectroscopy; LC, liquid chromatography; NIST, National Institute of Standards and Technology; PTFE, polytetrafluoroethylene; CAR/PDMS, carboxen/polydimethylsiloxane; RI, retention index; FC, fold change; EBAM, Empirical Bayesian Analysis of Microarray; VIP, variable importance; RSD, relative standard deviations; GC-O, gas chromatography olfactometry; UFA, unsaturated fatty acid; OOB, out-of-bag.

Appendix A. Supplementary Data

Supporting information of the 22 volatile compounds. RT indicates the retention time of the compound, RI and RI^{a} represent the experimental retention index and the standard retention index, respectively (Table S1);

Siloxane derivatives detected using SPME GC/MS. RT indicates the retention time of the compound, RI and RI^{a} represent the experimental retention index and the standard retention index, respectively (Table S2);

Supporting information of venn diagrams showing selection of typical biomarkers in combination with four statistical analysis methods (Figure S1) (PDF).

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for piblication

All authors agree to publish.

Availability of data and material

All data and materials from experiments are available.

Competing interests

The authors declare no competing financial interest.

Funding

This work was financially supported by the National Key Research and Development Program of China (no. 2017YFC1601905). Heilongjiang Province University Basic

Research Operating Costs Heilongjiang University Special Project (no.KJCX201817).

Authors' Contributions

Shengying Hu performed the experiments and wrote the manuscript draft; Hongbo Ren helped with the aroma experiments. Songyong and Siyuan Gao helped with data analysis. Li Meng designed the experiments, provided guidance and finalized the manuscript

Acknowledgments

The authors thank Engineering Research Center of Agricultural Microbiology Technology and Quality and Safety Institute of Agricultural Products, Heilongjiang Academy of Agricultural Sciences for instrumental support and Xiaoyun Han at Heilongjiang University for detection helping.

Authors' information

Corresponding Autor:

Li Meng, E-mail: mengli198026@126.com, Phone:+86 18845122960

ORCID: 0000-0003-0724-4094

References

Bryant R, McClung A (2011) Volatile profiles of aromatic and non-aromatic rice cultivars using SPME/GC-MS. Food Chem 124:501-13. doi:10.1016/j:foodchem.201

0.06.061.

Buttery RG, Ling LC, Juliano BO (1982) 2-Acetyl-1-pyrroline: An important aroma component of cooked rice. Chem Indus 12:958-959.

Concepcion JCT, Ouk S, Riedel A, Calingacion M, Zhao D, Ouk M, Garson M. J, Fitzgerald MA (2018) Quality evaluation, fatty acid analysis and untargeted profiling of volatiles in Cambodian rice. Food Chem 240:1014-1021. doi:10.1016/j.foodchem.

2017.08.019.

Cutler DR, Edwards Jr TC, Beard KH, Cutler A, Hess KT, Gibson J, Lawler JJ (2007) Random forests for classification in ecology. Ecol 88:2783-2792. doi:10.1890/07-053

9.1.

Daygon VD, Prakash S, Calingacion M, Riedel A, Ovenden B, Snell P, Fitzgerald MA (2016) Understanding the jasmine phenotype of rice through metabolite profiling and sensory evaluation. Metabolom 12(4):63-78. doi:10.1007/s11306-016-0989-6.

Emwas A-HM (2015) The strengths and weaknesses of NMR spectroscopy and mass spectrometry with particular focus on metabolomics research. Metabonomics Methods Protoc 1277:161-193. doi:10.1007/978-1-4939-2377-9_13

Engel W, Bahr W, Schieberle P (1999) Solvent assisted flavour evaporation – A new and versatile technique for the careful and direct isolation of aroma compounds from complex food matrices. Eur Food Res Technol 209:237-241. doi:10.1007/s002170050

486.

Frankel EN (2012) Hydroperoxide decomposition. In: Frankel EN (ed) Vol 4E. Woodhead Publishing, p 67-98.

Grimm CC, Bergman C, Delgado JT, Bryant R (2001) Screening for 2-acetyl-1-pyrroline in the headspace of rice using SPME/GC–MS. J Agric Food Chem 49:245-249. doi:10.1021/jf0008902.

Huang HS (2005) Study on the Influence of Aroma Analysis Method and Processing Process on the Aroma of Aromatic Rice. Hunan Agricultural University, Changsha.

Kim S, Shin B-K, Lim DK, Yang T-J, Lim J, Park JH, Kwon SW (2015) Expeditious discrimination of four species of the Panax genus using direct infusion-MS/MS combined with multivariate statistical analysis. J Chromatogr B Anal Technol Biomed Life Sci 1002:329-336. doi:10.1016/j.jchromb.2015.08.034.

Kang DF, He JM, Wang XC (2007) Analysis of Odor Components of Rice and Its Products by Headspace Solid Phase Microextraction and GC-MS. J chin cereals 22(5):147-149.

Kyu D, Phuoc N, Won S (2018) Non-destructive profiling of volatile organic compounds using HS-SPME/GC-MS and its application for the geographical discrimination of white rice. J Food Drug Anal 26(1):261-269. doi:10.1016/j.jfda.201

7.04.005.

Li Y, Zhang YJ, Nie Y (2017) Japanese rice traceability system establishes enlightenment to China. World agric 2017;11:40-46. doi: 1013856/j.cn11-1097/s.201

7.11.07.

Li Q, Liu WL, Lu YN, Lv YK (2018) Progress in application of headspace gas chromatography-mass spectrometry. Chromatogr 36(10):962-971. doi:10.3724/SP.J.11

23.2018.05013.

Lee D-K, Yi T, Park K-E, Lee H-J, Cho Y-K, Lee SJ, Lee J, Park JH, Lee M-Y, Song SU, Kwon SW (2014) Non-invasive characterization of the adipogenic differentiation of human bone marrow-derived mesenchymal stromal cells by HS-SPME/GC-MS. Sci Rep 4:6550-6558. doi:10.1038/srep06550.

Lashgari M, Singh V, Pawliszyn J (2019) A critical review on regulatory sample preparation methods: Validating solid-phase microextraction techniques. Trend Anal Chem 119:1-11. doi:10.1016/j.trac.2019.07.029.

Laguerre M, Mestres C, Davrieux F, Ringuet J, Boulanger R (2007) Rapid discrimination of scented rice by solid-phase microextraction, mass spectrometry and multivariate analysis used as a mass sensor. J Agric Food Chem 55:1077-1083. doi:10.1021/jf0623992.

Limpawattana M, Shewfelt RL (2010) Flavor lecicon for sensoy descriptive profiling of different rice types. J Food Sci 75:199-205. doi:10.1111/j.1750-384.2010.01577.x.

Lim DK, Mo C, Lee JH, Long NP, Dong Z, Li J, Lim J, Kwon SW (2018) The integration of multiplatform ms-based metabolomics and multivariate analysis for the geographical origin discrimination of *Oryza sativa* L. J Food Drug Anal 26(2):769-777. doi:10.1016/j.jfda.2017.09.004.

Mahattanatawee K, Rouseff RL (2014) Comparison of aroma active and sulfur volatiles in three fragrant rice cultivars using GC-Olfactometry and GC-PFPD. Food Chem 154:1-6. doi:10.1016/j.foodchem.2013.12.105.

Mathure SV, Jawali N, Thengane RJ, Nadaf AB (2014) Comparative quantitative analysis of headspace volatiles and their association with BADH2 marker in non-basmati scented, basmati and non-scented rice (*Oryza sativa* L.) cultivars of India. Food Chem 142:383-391. doi:10.1016/j.foodchem.2013.07.066.

Nazife N, Yazgan K, Tugba B, Ali T, Boyaci IH (2019) Raman spectroscopy coupled with chemometric methods for the discrimination of foreign fats and oils in cream and yogurt. J Food Drug Anal 27:101-110. doi:10.1016/j.jfda.2018.06.008.

Pramai P, Hamid NAA, Mediani A, Maulidiani M, Abas F, Jiamyangyuen S (2018) Metabolite profiling, antioxidant, and aglucosidase inhibitory activities of germinated rice: nuclearmagnetic-resonance-based metabolomics study. J Food Drug Anal 26:47-57. doi:10.1016/j.jfda.2016.11.023.

Reig-Valiente JL, Viruel J, Sales E, Marques L, Terol J, Gut M, Derdar S, Talón M, Domingo C (2016) Genetic diversity and population structure of rice varieties cultivated in temperate regions. Rice 9:58-70. doi:10.1186/s12284-016-0130-5.

Raro M, Moco S, Lommen A, Keurentjes JJ, Bino RJ, Hall RD (2007) Untargeted large-scale plant metabolomics using liquid chromatography coupled to mass spectrometry. Nat Protoc 2:778-791. doi:10.1038/nprot.2007.95.

Suzuki Y, Chikaraishi Y, Ogawa NO, Ohkouchi N, Korenaga T (2008) Geographical origin of polished rice based on multiple element and stable isotope analyses. Food Chem 109:470-475. doi:10.1016/j.foodchem.2007.12.063.

Sakthivel K, Sundaram RM, Rani NS, Balachandran SM, Neeraja CN (2009) Genetic and molecular basis of fragrance in rice. Biotechnology Adv 27(4):468-473. doi:10.10

16/j.biotechadv.2009.04.001.

Vemireddy LR, Satyavathi V, Siddiq E, Nagaraju J (2015) Review of methods for the detection and quantification of adulteration of rice: basmati as a case study. J Food Sci Technol 52:3187-3202. doi:10.1007/s13197-014-1579-0.

Wang B. (2017). Five golden flowers of Heilongjiang rice. Heilongjiang Grain. **11**, 40-44.

Widjaja R, Craske JD, Wootton M (1996) Comparative studies on volatile components of non-fragrant and fragrant rices. J Sci Food Agric 70(2): 151-161. doi:1

0.1002/(sici)1097-0010(199602)70:2<151::aid-jsfa478>3.0.co;2-u.

Wilkens WF, Lin FM (1970) Gas chromatographic and mass spectral analyses of soybean milk volatiles. J Agric Food Chem 18(3):333-336. doi:10.1021/jf60169a003.

Xia J, Psychogios N, Young N, Wishart DS (2009) MetaboAnalyst: a web server for metabolomic data analysis and interpretation. Nucleic Acids Res 37:652-660. doi: 10.1093/nar/gkp356.

Yaylayan VA, Keyhani A (1999) Origin of 2,3-pentanedione and 2,3-butanedione in d-glucose/l-alanine maillard model systems. J Agric Food Chem 47(8):3280-3284. doi:10.1021/jf9902292.

Zhao Y, Zhao C, Li Y, Chang Y, Zhang J, Zeng Z, Guo WX (2014) Study of metabolite differences of flue-cured tobacco from different regions using a pseudotargeted gas chromatography with mass spectrometry selected-ion monitoring method. J Sep Sci 37:2177-2184. doi:10.1002/jssc.201400097.

Figures

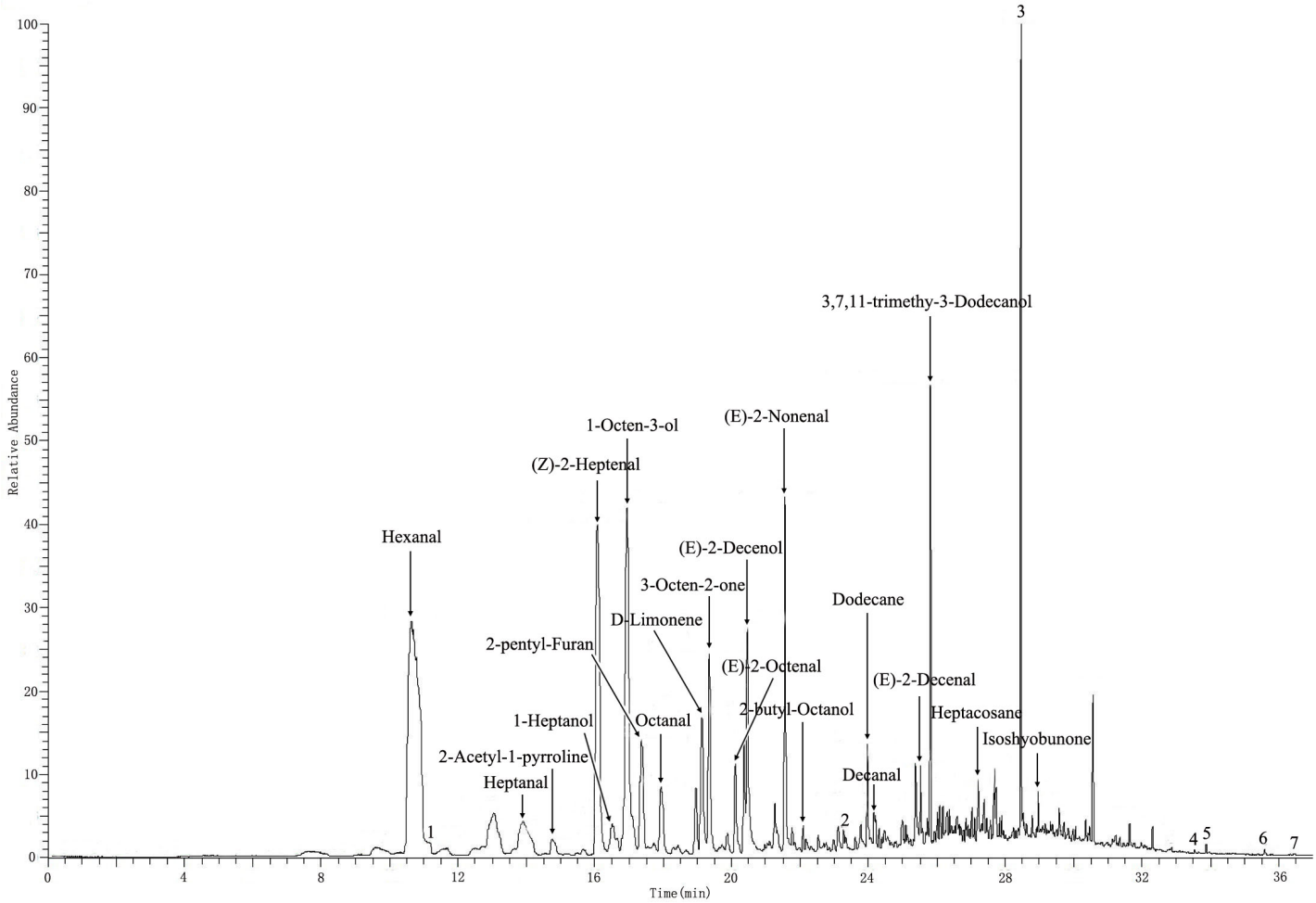


Figure 1

VOCs profiling of Wu Chang rice obtained by GC-MS using CAR/PDMS SPME at extraction temperature of 80°C for 1h.

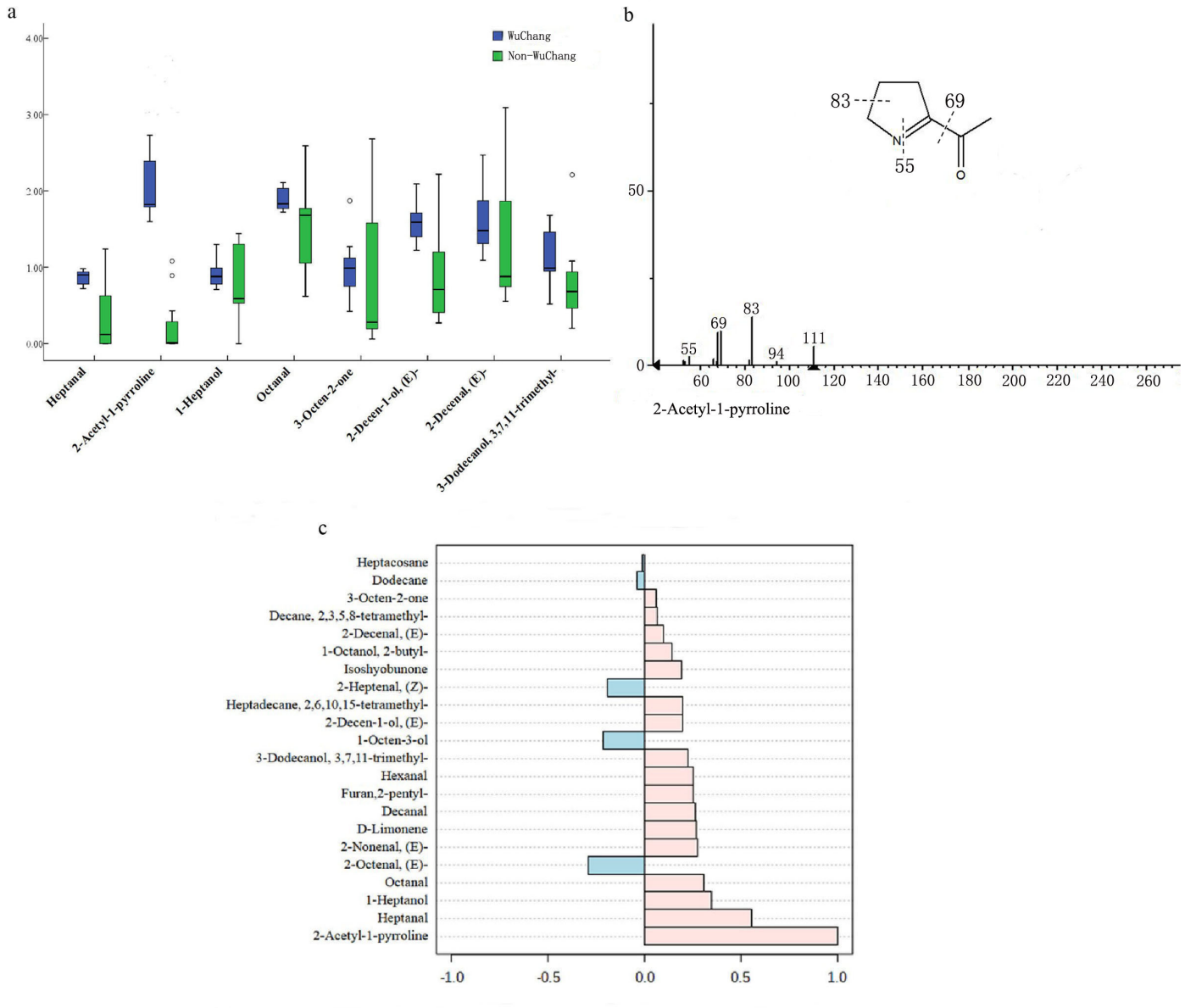


Figure 2

Box plots and 2-AP analysis of eight chemical biomarkers in rice samples. (a) Box diagram of eight typical biomarkers for rice samples from Wu Chang and other regions; (b) 2-AP ion map; (c) Correlation analysis of 2-AP with other volatile compounds

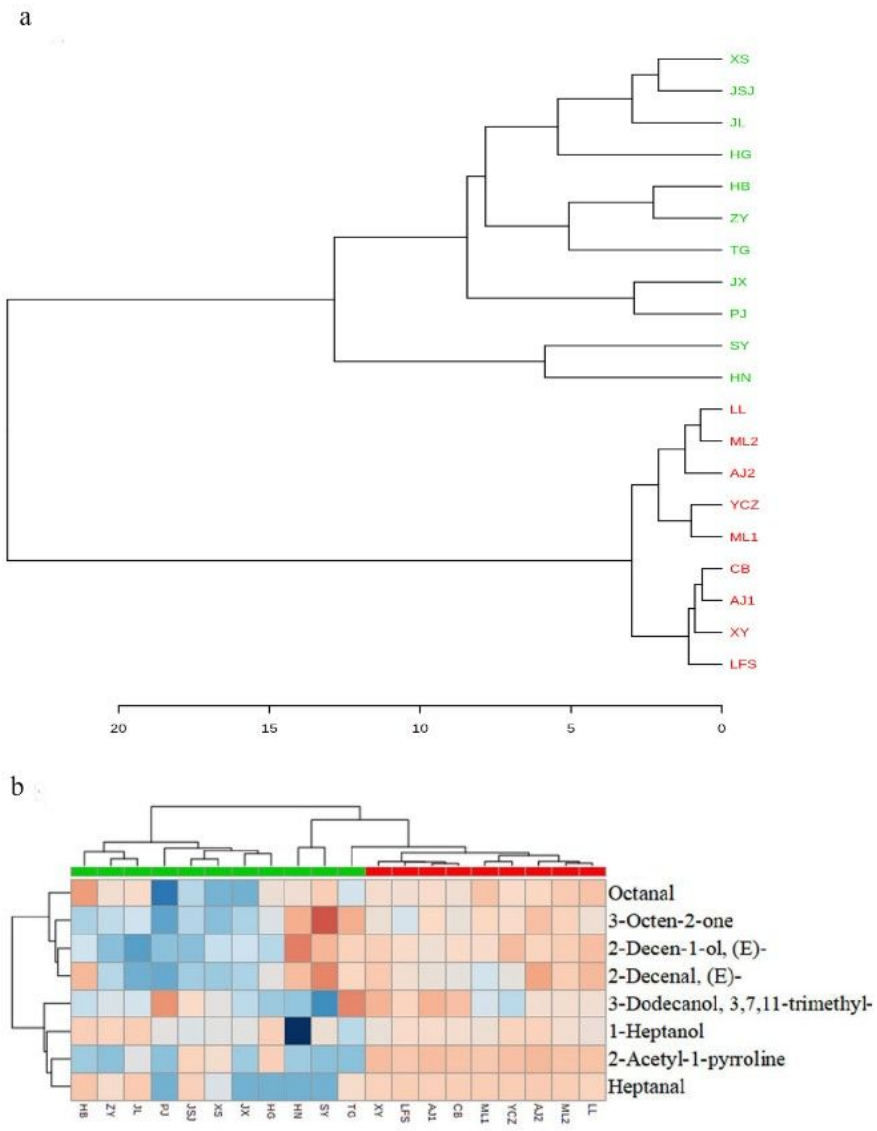
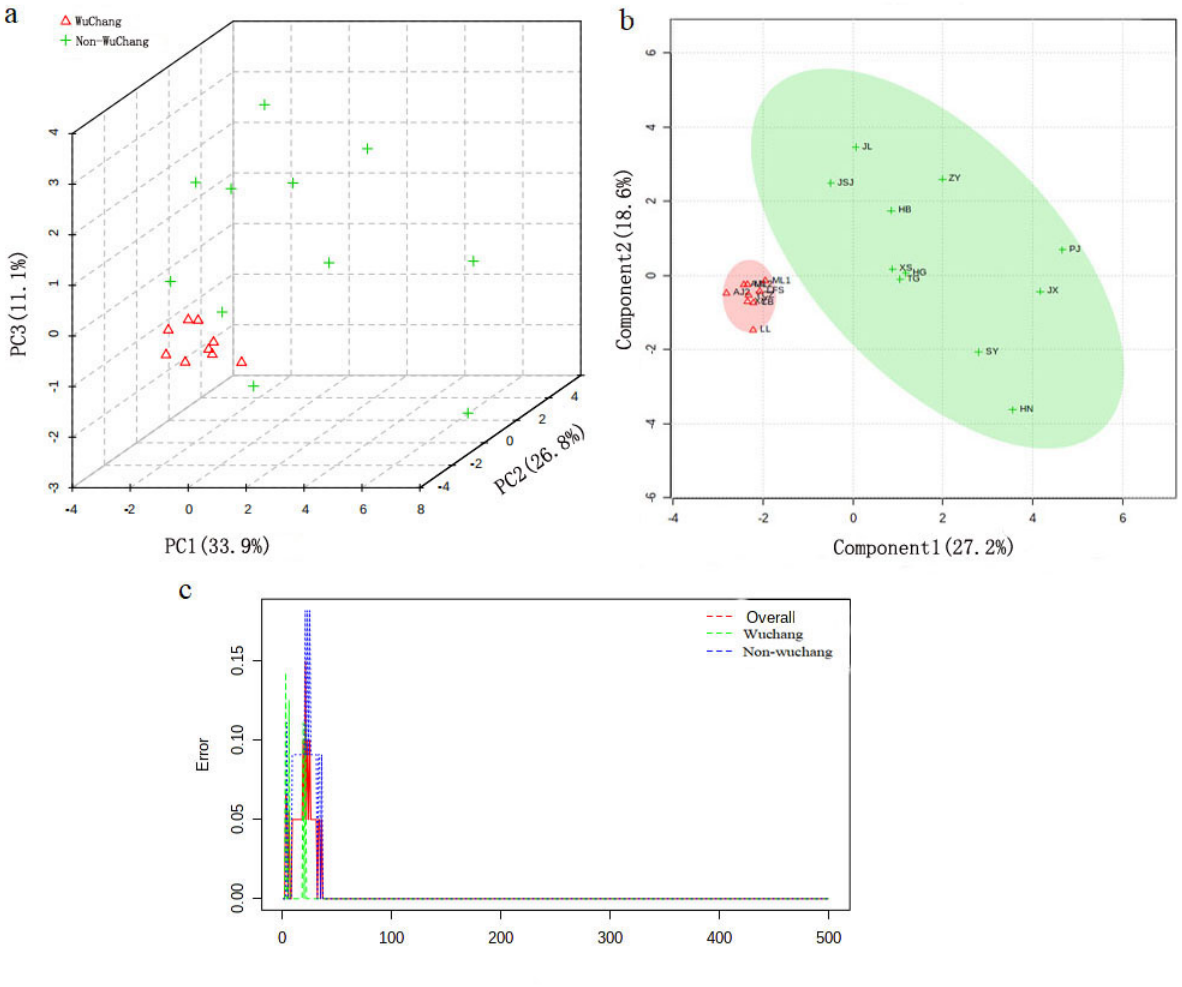


Figure 3

Dendrogram and the heatmap of the eight VOC typical biomarkers (distance measurement using euclidean, and clustering algorithm using ward.D).



	Wuchang	Non-wuchang	Class.error
Wuchang	9.00	0.00	0.00
Non-wuchang	0.00	11.00	0.00

Figure 4

Different statistical analyses of the chemical biomarkers. (a) 3D score plot between the selected PCs in PCA. The explained variances are shown in brackets; (b) PLS-DA models revealed good potential for discriminating rice samples between Wu Chang and other cultivars; (c) Cumulative error rates by Random Forest classification. The red lines indicate the overall error rate, and the green and blue lines represent the error rates for each class.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplementarydata.pdf