JOURNAL OF AGRICULTURAL AND FOOD CHEMISTRY

Article

pubs.acs.org/JAFC

Metabolic Profiling of Chinese Tobacco Leaf of Different Geographical Origins by GC-MS

Li Zhang,[†] Xiaoyu Wang,[†] Jizhao Guo,[†] Qiaoling Xia,[†] Ge Zhao,[†] Huina Zhou,[†] and Fuwei Xie*[,†,‡]

[†]Zhengzhou Tobacco Research Institute of CNTC, Zhengzhou, People's Republic of China

[‡]Chemistry Department, Zhengzhou University, Zhengzhou, People's Republic of China

Supporting Information

ABSTRACT: Tobacco leaf obtained from different geographical areas in China was profiled using gas chromatography−mass spectrometry (GC-MS) coupled with multivariate data analyses. Hierarchical cluster analysis (HCA) and principal component analysis (PCA) showed that the tobacco metabolome was clearly dependent on geographical origins; climatic conditions, such as temperature and precipitation, imposed a greater impact on metabolite levels than the cultivars. By orthogonal partial least-squares-discrimination analysis (OPLS-DA), 20 metabolites that contributed to the discrimination were screened, including primary metabolites (sucrose, D-fructose, D-mannose, D-glucose, inositol, maleic acid, citric acid, malic acid, L-threonic acid, L-proline, L-phenylalanine), secondary metabolites (chlorogenic acid, α- and β-4,8,13-duvatriene-1,3-diol, nicotine, quinic acid), and four unknown metabolites. The results suggest that metabolic profiling using GC-MS combined with multivariate analysis can be used to discriminate tobacco leaf of different geographical origins and to provide potential indicators of tobacco origins.

KEYWORDS: *GC-MS, metabolic profiling, geographical origins, Chinese tobacco leaf*

INTRODUCTION

It is well known that tobacco leaf is the important material of the tobacco industry; its metabolites are closely related to the flavor of cigarettes.[1,2] The metabolites of tobacco leaf are affected by several factors including cultivar, growth altitude, soil, and climate conditions such as temperature, sun exposure time, and rainfall.[3] There were some investigations about the effects of cultivar or geographical factors on tobacco composition and quality. Sakaki et al.[4] investigated the influence of tobacco cultivar on the smoking quality with respect to the volatiles. Yang et al.[5] reported that the content of carotenes tended to increase with altitude, while Weeks et al.[6] showed that tobacco leaf of high quality contained larger amounts of degradation products of carotenes. The effects of photoperiod and end-of-day light quality on alkaloids and phenolic compounds of tobacco were also reported.[7] Recently, Gang et al.[8] reported that polyphenols and organic acids in tobacco leaf varied in different planting regions in China. However, these studies focus only on specific groups of chemical components that are believed to be important to the flavor quality of flue-cured tobacco,[9] such as sugars, organic acids, alkaloids, and volatile aroma substances, so the results could not reveal global changes of tobacco leaf. Besides, these studies mainly took cured leaves as the study subject. The changes of chemical components in the cured tobacco leaf were the effects of not only cultivar or geographical factors but also a series of processes for curing.[10] A direct relationship between global metabolome of tobacco leaf and geographic factors (i.e., growth altitude, temperature, sun exposure time, light quality, precipitation, etc.) has not been well studied to date.

Metabolomics is a nontargeted method that is applied to characterize global changes of metabolites comprehensively.[11,12] By analyzing the metabolites in fresh plant tissues, which are the end products of cellular regulatory processes, we can know about responses of a biological system to perturbations imposed on plants, including genetic or environmental changes, using metabolomics.[13] At present, metabolomics has been applied in diverse analysis fields, such as medical diagnosis, [14] Traditional Chinese Medical (TCM) materials,[15] and food quality and safety.[16,17] Recently, studies on differentiating the geographical origins of food (green tea, grape berries, wines, *Angelica gigas*, etc.) using metabolic profiling approaches have also been reported.[18−21] For tobacco, Zhang et al.[22] studied the dynamic responses of tobacco to salt stress using metabolomics, and Choi et al.[23] analyzed the metabolic fingerprinting of wild-type and transgenic tobacco plants. However, metabolic profiling analysis of tobacco leaf of different geographical origins has not been reported.

The main analysis platforms in metabolomics studies are gas chromatography−mass spectrometry (GC-MS), liquid chromatography−mass spectrometry (LC-MS), and nuclear magnetic resonance (NMR).[11,24] For its repeatability, wide ranges of measurements, and straightforward peak identification, GC-MS has been regarded as one of the most applicable and versatile methods in metabolomics.[25−27] In this research, a GC-MS-based metabolic profiling method coupled with pattern recognition was used for discriminating tobacco leaf from different planting regions in China and identifying the significant metabolites responsible for the discrimination.

MATERIALS AND METHODS

Materials. Nine samples of middle and upper fresh flue-cured tobacco leaf were collected from three main planting regions in China:

Received: October 22, 2012
Revised: February 26, 2013
Accepted: February 27, 2013
Published: February 27, 2013

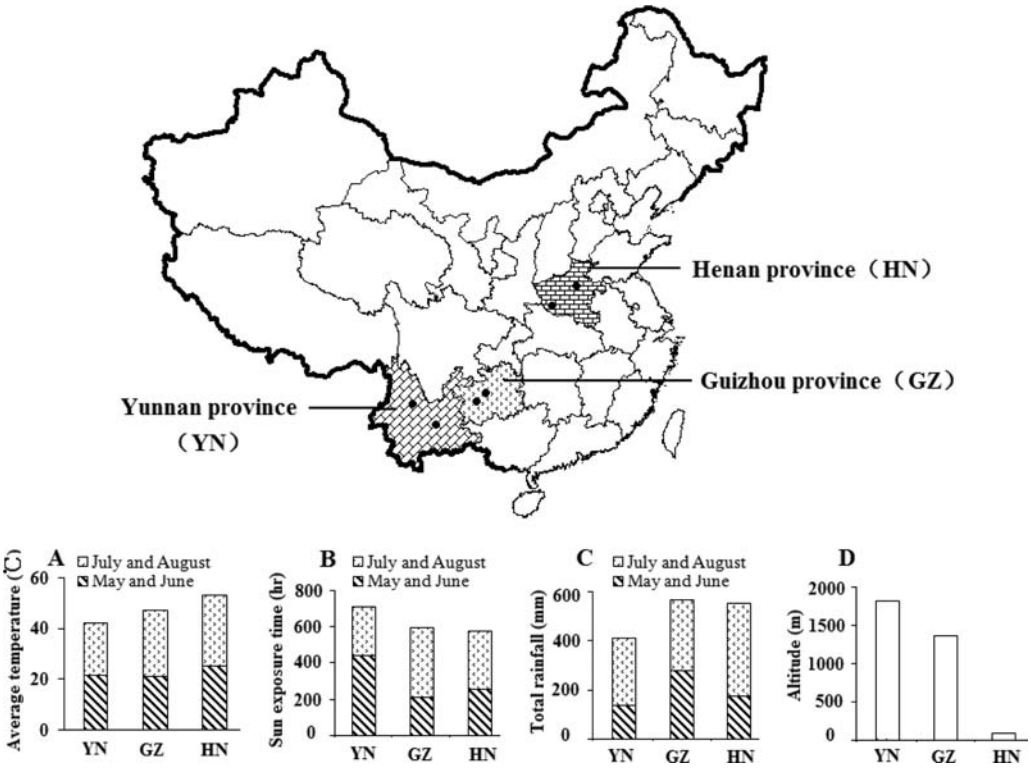

Figure 1. Geographical planting regions of tobacco leaf in China and climate and geographical conditions of (A) average temperature (°C); (B) total sun exposure time (h); (C) total rainfall (mm); and (D) altitude (m) of the three planting regions in China. The black dots represent the different tobacco-growing areas.

the southwestern Yunnan Province and Guizhou Province and the central Henan Province (Figure 1). The average daily temperature, total sun exposure time, and total rainfall from May to August 2011 for Yunnan, Guizhou, and Henan are shown in Figure 1. The data were obtained from China Meterological Administration. These regions represented the main tobacco farming environments, and tobacco leaf of these regions represented three representative flavor types. Six collections were obtained for each sample,[15] so a total of 54 collections were used for tobacco metabolic profiling analysis (Table 1). The six collections of each sample were from individual plants that were grown in the same field. Quality control (QC) samples were obtained by blending the same amount of each milled collection thoroughly. [26]

Table 1. Flue-Cured Tobacco Samples Collected from Different Geographical Origins

index	cultivars	growing locations	sampling position	sampling date
A	Hongda (HD)	Yunnan Xiangyun (YN)	upper leaf (B)	September 1
B	Hongda (HD)	Yunnan Xiangyun (YN)	middle leaf (C)	August 10
C	K326	Yunnan Yuxi (YN)	middle leaf (C)	July 23
D	K326	Guizhou Zunyi (GZ)	upper leaf (B)	August 18
E	K326	Guizhou Zunyi (GZ)	middle leaf (C)	July 28
F	Yunyan97 (Y97)	Guizhou Qianxi (GZ)	Middle leaf (C)	August 19
G	Zhongyan100 (Z100)	Henan Xiangxian (HN)	upper leaf (B)	August 29
H	Zhongyan100 (Z100)	Henan Xiangxian (HN)	middle leaf (C)	August 9
I	Yunyan87 (Y87)	Henan Nanyang (HN)	middle leaf (C)	August 24

Reagents. Methanol and chloroform used as extraction solvents were purchased from J T Baker (Phillipsburg, NJ, USA). Forty-two standard substances used for the identification, derivatization reagents including *N*-methyl-*N*-(trimethylsilyl)trifluoroacetamide (MSTFA) and methoxyamine hydrochloride, and pyridine used as a solvent were all purchased from J&K (Beijing, China). Hexadecanoic-7,7,8,8-d_4 acid used as an internal standard was purchased from C/D/N Isotopes, Inc. (Pointe-Claire, Quebec, Canada). Distilled water was purified in-house using the Millipore Milli-Q system (Billerica, MA, USA).

Sample Preparation for GC-MS Analysis. The sample preparation protocols were mainly based on the methods published in the literature.[11,27,28,30] Fresh tobacco leaf was immediately frozen in liquid nitrogen and stored at −80 °C until further analysis. The fresh tobacco leaf was freeze-dried in a 24 h period and then ground to a powder and filtered through a 40-mesh sieve. The fresh tobacco metabolites were extracted using a single-phase solvent mixture of MeOH, H_2O, and $CHCl_3$ in a ratio of 5:2:2 (v/v/v).[15,16] Twenty milligrams of fresh tobacco powder was transferred to a 5 mL Eppendorf tube, and then 1 mL of solvent mixture and 200 μL of hexadecanoic-7,7,8,8-d_4 acid (diluted with a solvent mixture to a concentration of 40 μg/mL) were added, respectively. Samples were extracted by an ultrasonic method with the solvent mixture for 40 min at room temperature. The extraction solution was then centrifuged at 10000*g* for 10 min. Subsequently, a 400 μL portion of the supernatant was transferred to a 2.5 mL Eppendorf tube and was dried by an N-EVAP concentrator (Organomation Associates, Inc., Berlin, MA, USA).

For derivatization, 50 μL of methoxyamine hydrochloride in pyridine (20 mg/mL) was added to reduce the number of derivatives of reducing sugars. The mixture was vortexed for 1 min and then incubated at 37 °C for 90 min. The sample was then silylated for 30 min at 37 °C by adding 70 μL of MSTFA and vortex-mixing for 30 s.[11,25,30] After at least 2 h at room temperature, the samples were analyzed by GC-MS.

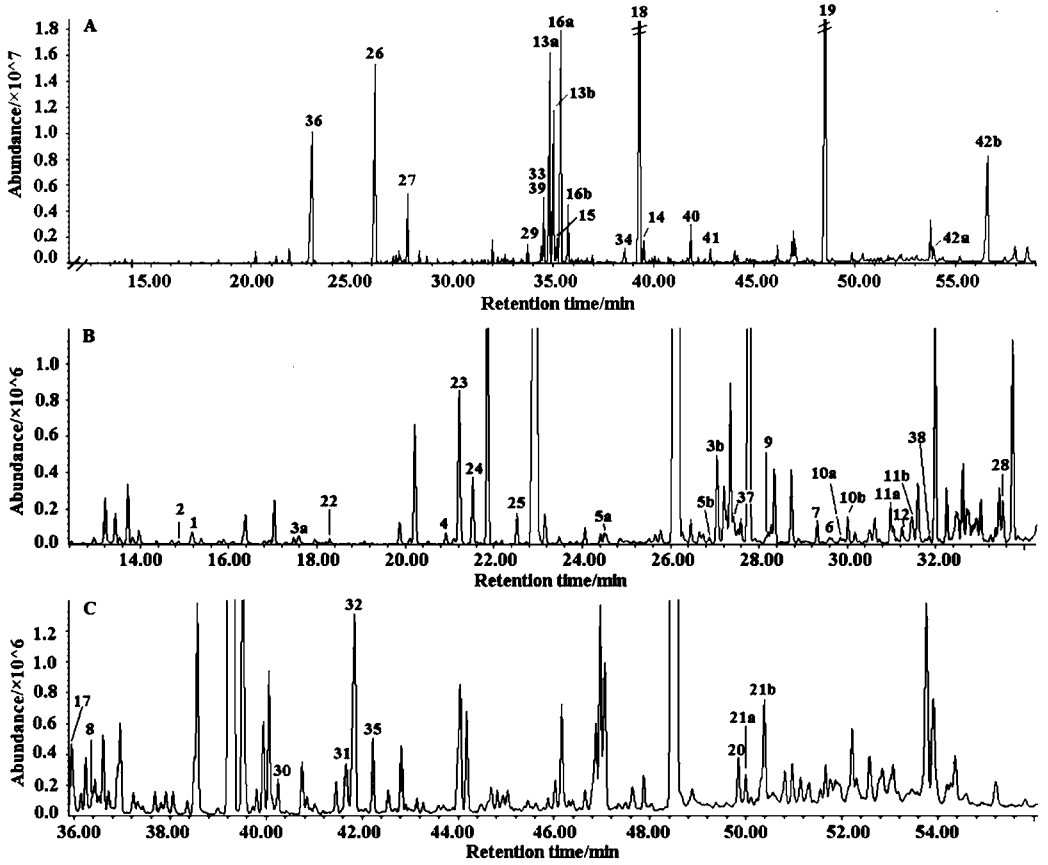

Figure 2. Representative GC-MS total ion chromatograms of tobacco leaf. Plots B and C are the large version of chromatogram A in 14−32 min and 36−54 min. The labeled peaks are listed in Table 2.

GC-MS Analysis. Samples were analyzed in random order with QC samples inserted with every eight samples in the running sequence. For each run, 1 μL of derivatized sample was injected by an Agilent G4513A autosampler (Agilent, Atlanta, GA, USA) into an Agilent 7890-5975 GC-MS system with a split ratio of 10:1. The column used was a 30 m × 0.25 mm i.d., 0.25 μm, DB-5 MS fused-silica capillary column (Agilent, Palo Alto, CA, USA). The inlet temperature was 290 °C, and the helium gas flow rate through the column was 1 mL/min. The initial oven temperature was held at 70 °C for 4 min, ramped to 310 °C at 5 °C/min, and then held there for 10 min. The transfer line and the ion source temperatures were 280 and 230 °C, respectively. The ionization mode was the electron impact at 70 eV. The mass spectra plot was acquired using full scan monitoring mode with a mass scan range of m/z 40−510. The acceleration voltage was turned on after a solvent delay of 8 min.

Data Preprocessing. The components eluting within the total ion chromatogram were extracted in an Automatic Mass Spectral Deconvolution and Identification System (AMDIS) (NIST, Gaithersburg, MD, USA). Significant metabolites were identified by comparing their mass spectra and retention index with those of commercial standards.[31,32] An in-house automatic integration method was established in an Agilent MSD ChemStation and applied to the quantitation of selective ion traces. To guarantee the accuracy of integration, manual corrections were performed. A three-dimensional matrix including sample information, peak retention time, and peak relative intensities was generated. The relative intensities of the various metabolites were obtained by normalizing the intensity of individual ion traces, which are indicative of the respective compounds, to the response of an internal reference compound. Internal standard and any known artificial peaks, such as peaks caused by noise, column bleed, and derivatization procedure, were removed from the matrix. According to a modified "80% rule", only peaks detected in more than 80% of the samples of any given group were kept for further multivariate analysis.

Multivariate Analysis of GC-MS. Hierarchical cluster analysis (HCA) and principal component analysis (PCA) were selected to obtain the first understanding of the relationships among the data matrix. The data matrix was mean centered and Pareto (Par) scaled with the SIMCA-P 13.0 Demo (Umetrics, Umeå, Sweden) before the multivariate analysis.[28] For HCA, the squared Euclidean distance and the incremental linkage method were used in the assignment of clusters for HCA (SPSS 17.0, SPSS Inc., USA).

In order to examine the differences in the metabolite levels of fresh tobacco leaf of different geographical origins, an orthogonal partial least-squares discriminate analysis (OPLS-DA) (SIMCA-P 13.0 Demo, Umetrics, Umeå, Sweden) was used. The data matrix was also mean centered and Pareto (Par) scaled prior to modeling.[33] The efficiency and reliability of the OPLS-DA model were verified by percent variation of the x and y variables explained by the model (R2X, R2Y) and the predictive performance of the model (Q2).

RESULTS AND DISCUSSION

Reproducibility. In the metabolic profiling analysis, the absolute amounts of metabolites are not required. However, the reproducibility of results is an essential factor in assessing the quality of the analytical technique.[12,25] To test the reproducibility of our system, three components contributing to the observed variability were determined, including variability caused by GC-MS analysis and stability of chemically modified samples, the sample preparation, and the biological material. To gain insight into the reproducibility of the analysis, a single chemically modified sample was divided into 16 aliquots and measured in sequence. The result showed that the deviations of these detections were controlled within a 2 SD range except for

Table 2. Primary and Secondary Metabolites Identified by GC-MS from Chinese Tobacco Leaf[a]

primary metabolites			
amino acids	sugars and polyols	organic and fatty acids	secondary metabolites
L-alanine, **1**	2-deoxy-D-ribose, **9**	propanedioic acid, **22**	nicotine, **36**
L-valine, **2**	D-xylose, **10**	maleic acid, **23**	anatabine, **37**
L-proline, **3**	xylitol, **11**	succinic acid, **24**	cotinine, **38**
L-threonine, **4**	L-rhamnose, **12**	fumaric acid, **25**	quinic acid, **39**
L-aspartic acid, **5**	D-fructose, **13**	malic acid, **26**	α-4,8,13-duvatriene-1,3-diol, **40**
L-phenylalanine, **6**	D-mannose, **14**	L-threonic acid, **27**	β-4,8,13-duvatriene-1,3-diol, **41**
L-glutamine, **7**	D-galactose, **15**	shikimic acid, **28**	chlorogenic acid, **42**
L-tyrosine, **8**	D-glucose, **16**	citric acid, **29**	
	sorbitol, **17**	caffeic acid, **30**	
	inositol, **18**	linoleic acid, **31**	
	sucrose, **19**	linolenic acid, **32**	
	α-lactose, **20**	tetradecanoic acid, **33**	
	maltose, **21**	hexadecanoic acid, **34**	
		octadecanoic acid, **35**	

[a]Identification was performed by comparing the mass spectra and retention index with those of commercial standards.

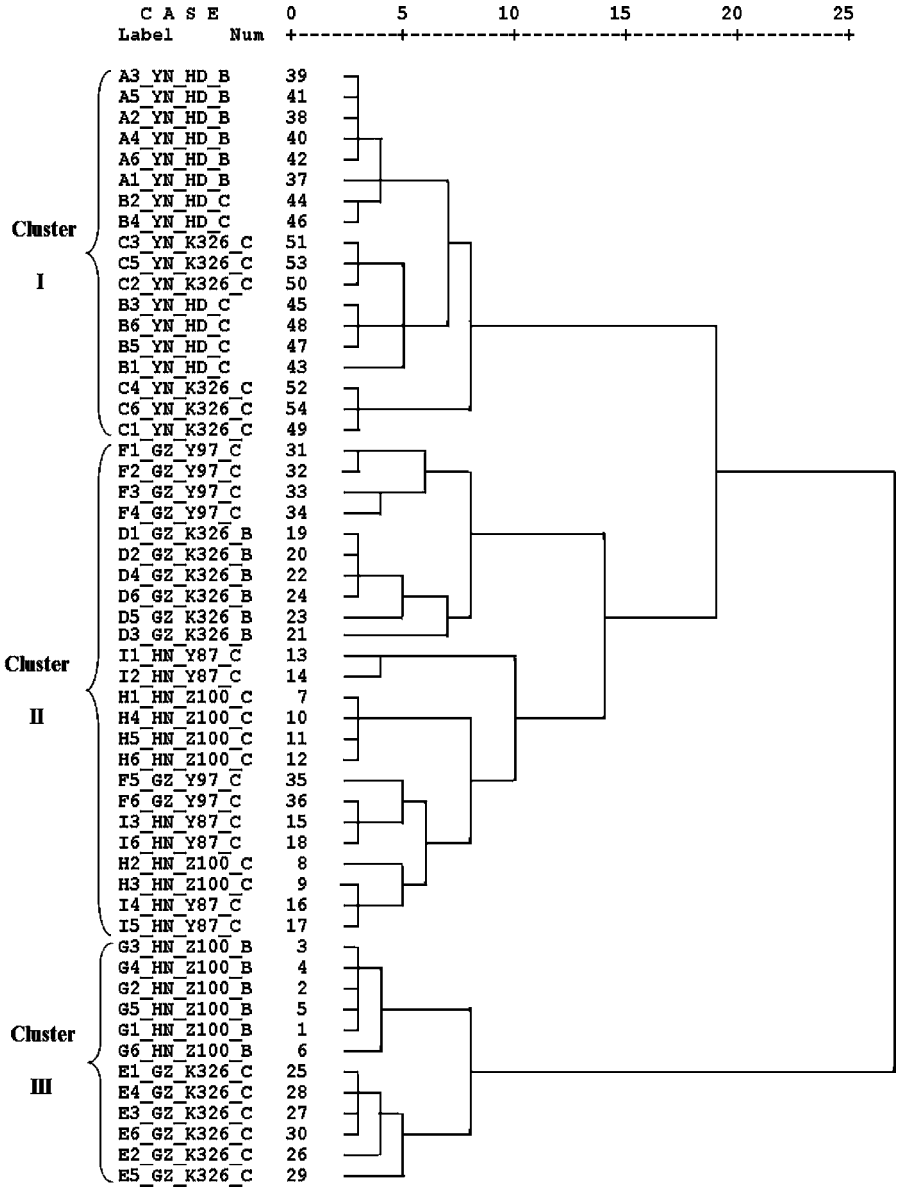

Figure 3. Dendrogram obtained from HCA analysis of the metabolic profiling of tobacco leaf of different geographical origins.

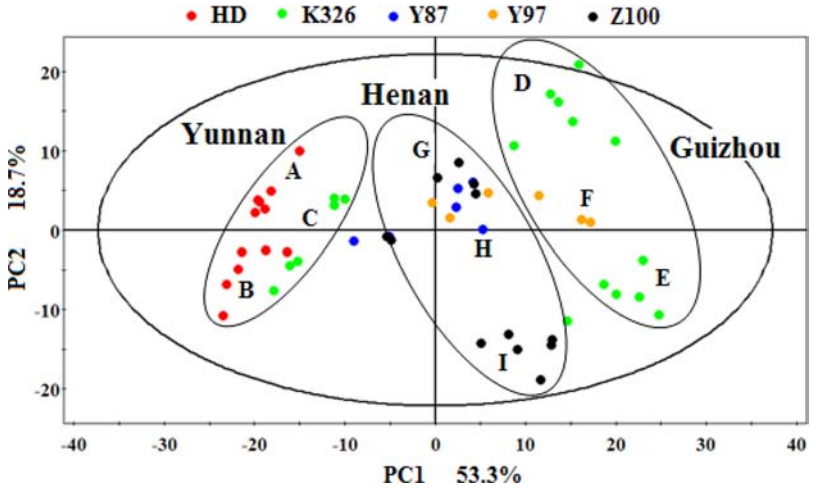

Figure 4. PCA score plot marked by different geographical origins and cultivars. Samples A, B, and C are all from Yunnan; samples D, E, and F are all from Guizhou; and samples G, H, and I are all from Henan. Samples C, D, and E all belong to the K326 cultivar. The PCA score plot of PC1 and PC2 includes 72% of the total information derived from data variances and distinguishes collections of Yunnan, Guizhou, and Henan according to their different geographical origins.

the first run. Therefore, the samples were analyzed by GC-MS after at least 2 h at room temperature. In order to avoid the influence of different derivatization times, all samples were analyzed in random order. QC samples, which were obtained by thoroughly blending the same amount of each milled sample, were inserted every eight samples in the running sequence to monitor and evaluate the reproducibility of the whole process. The mean relative standard deviation (RSD) of all the resolved peaks' retention time and relative peak areas was 0.03% ± 0.03% and 22% ± 16%, respectively. The stability of retention time makes it possible to automatically integrate. In order to obtain insight into the biological variability, the middle maturity leaves of Yunnan Province from six individual tobacco plants grown side by side under identical conditions were analyzed. The results showed that the variability caused by GC-MS analysis, sample preparation, and biological samples was small and could be tolerated. Also, the biological variability was in clear excess of the variability caused by sample preparation and GC-MS analysis. Therefore, variability in results is mainly from the variability within the biological samples themselves. All the results showed that GC-MS was a reliable method for metabolite analysis.

Characterization of Metabolites by GC-MS. Nine samples (54 collections) including different cultivars, different geographical origins, and different sampling positions, representing large varieties of tobacco leaf in China, were investigated. Figure 2 shows one of the typical total ion chromatograms (TIC) of the QC sample, which was obtained by thoroughly blending the same amount of each milled sample. In case of partially overlapped peaks, the AMDIS software was used to extract the pure compound. By comparing their mass spectra and retention index with those of commercial standards, a total of 42 metabolites were identified in the chromatogram (Table 2).[32] Primary metabolites, such as sugars, organic acids, and amino acids, were detected with a high concentration in the tobacco leaf extracts. Several secondary metabolites closely related to tobacco quality, such as nicotine, α- and β-4,8,13-duvatriene-1,3-diol, and chlorogenic acid, were also observed in the chromatogram.

Metabolic Profiling with HCA and PCA. HCA and PCA analyses were applied to obtain a preliminary overview of similarities and differences among the collections. The HCA dendrogram (Figure 3) and PCA score plot (Figure 4) showed that samples of Yunnan Province were clearly different from those of the other two regions. In the PCA score plot, 54 collections were obviously separated into three groups according to the different geographical origins of the tobacco samples; the PC1/PC2 plot described 72% of the total variance (Figure 4). Samples from the same region even if they belonged to different cultivars were clustered into the same group, such as samples A, B, and C, samples D, E, and F, and samples G, H, and I, while samples that belonged to the same cultivar, but from different regions, were assigned to different groups, for example C and E. These results showed that geographical origins might have a greater impact on the metabolite levels than the cultivars.

The differences in geographical conditions and climatic conditions might be the main factors in the different metabolite levels of different geographical origins.[34] Yunnan Province is located in southwest China and is influenced by a low latitude plateau, mountainous country monsoon climate. Yunnan has strong ultraviolet light, long total sun exposure time, relatively stable average temperature, and distinct, alternating dry and rainy seasons.[26] Guizhou Province is located northeast of Yunnan Province and is influenced by a subtropical plateau monsoon climate. The average altitude and total sun exposure time of Guizhou are lower than that of Yunnan, and the average temperature is not as stable as that of Yunnan, either. Henan Province lies in the central plain of China and is influenced by a temperate continental climate. The temperature differences in different seasons of Henan are large, with high temperatures in summer, the field period of tobacco growth. The average daily temperature, total sun exposure times, and total rainfall from May to August 2011 for Yunnan, Guizhou, and Henan are shown in Figure 1. The data were obtained from China Meterological Administration.

Metabolites Differently Accumulated in Tobacco Leaf of Different Geographical Origins. The HCA and PCA analyses give us a preliminary overview of similarities and differences among collections and suggest that geographical origins impose a significant effect on metabolite levels. In order to further understand the differences among tobacco leaves of

different geographical origins, an orthogonal partial least-squares-discriminant analysis model was established. Noisy information irrelevant to geographical origins was removed prior to model building, so the OPLS-DA model was considered a more sophisticated algorithm. Six OSC components with 44.1% X variables independent of geographical origins were removed, leaving 50.8% of X variables related to geographical origins in the model. In the established OPLS-DA model, two significant components described 95.5% of the variation in Y (R2Y = 0.955) and predicted 91.3% (Q2Y = 0.913) according to cross-validation. Therefore, using the first two components to examine the collections was sufficient.

The score plot showed a clear separation of three groups: Yunnan, Guizhou, and Henan (Figure 5A). From the figure,

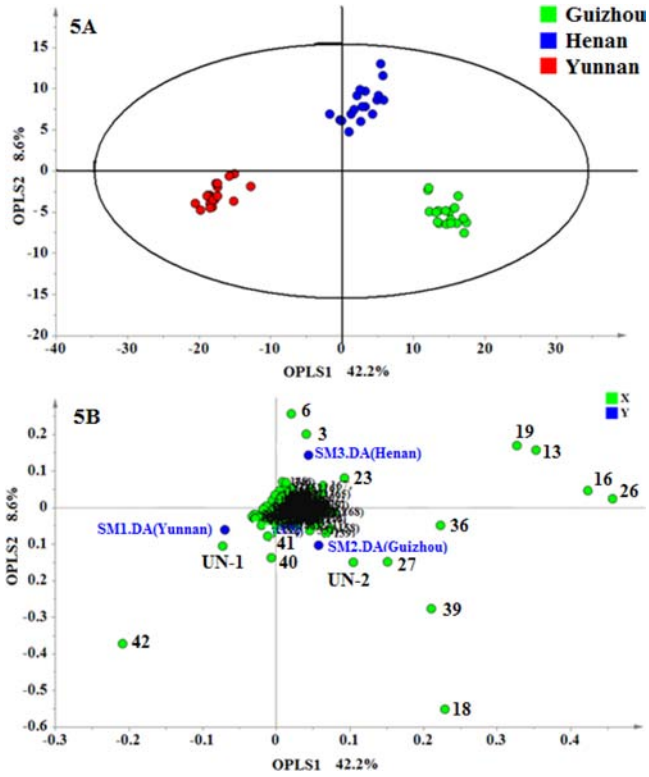

Figure 5. (A) Score plot and (B) loading plot of OPLS 1 and 2 generated from the OPLS model of Chinese tobacco leaf of different geographical origins. In the loading plot, the closer to the origin, the smaller contribution a metabolite makes to the discrimination. The most significant metabolites influencing the separation of tobacco leaf of different geographical origins were labeled. **3**, L-proline; **6**, L-phenylalanine; **13**, D-fructose; **16**, D-glucose; **18**, inositol; **19**, sucrose; **23**, maleic acid; **26**, malic acid; **27**, L-threonic acid; **36**, nicotine; **39**, quinic acid; **40**, α-4,8,13-duvatriene-1,3-diol; **41**, β-4,8,13-duvatriene-1,3-diol; **42**, chlorogenic acid, and two unknown metabolites, UN-1 and UN-2.

OPLS1 played a significant role in discriminating samples of Yunnan from the others, while OPLS2 had a comparatively strong impact on separating samples of Henan from those of Guizhou. More interesting, the locations of samples from Yunnan and Henan in the score plot were entirely opposite, while samples from Guizhou were located between them. The observations are in accordance with the flavor type of flue-cured tobacco leaf in China; the flue-cured tobacco leaves from different regions in China have their own characteristic flavor type; for example, tobacco leaves of Yunnan, Henan, and Guizhou have the typical characteristics of clear flavor, full flavor, and middle flavor, respectively.[35] It is reported that long total sun exposure time in the earlier stage of tobacco growth and the relatively low temperature, long rainy season, and sunless weather in the later stage of tobacco growth in Yunnan are the main causes of clear flavor type, while the relatively high temperature and plenty of sun exposure time in the later period in Henan are the primary factors in the formation of full flavor type.[36]

Furthermore, the contribution of each metabolite to OPLS1 and OPLS2 was computed, and each metabolite was given a loading or weight value for OPLS1 and OPLS2. In the corresponding loading plot (Figure 5B), chlorogenic acid, **42**, and an unknown metabolite, **UN-1**, had a negative contribution to OPLS1, which separated samples of Yunnan from the others, with Yunnan having a negative loading value. This showed chlorogenic acid and UN-1 were the discriminating components for tobacco leaf from Yunnan. In tobacco leaf from Henan, however, the levels of L-proline, **3**, L-phenylalamine, **6**, maleic acid, **23**, D-fructose, **13**, and sucrose, **19**, were relatively higher. For tobacco leaf from Guizhou, the levels of metabolites were similar to those of both Yunnan and Henan; the levels of α- and β-4,8,13-duvatriene-1,3-diol in Guizhou tobacco leaf were similar to those of Yunnan, while the levels of D-glucose and malic acid were similar to those of Henan. Compared with Yunnan and Henan, tobacco leaf from Guizhou contained higher levels of L-threonic acid, **27**, nicotine, **36**, quinic acid, **39**, inositol, **18**, and an unknown metabolite, UN-2.

Combining the VIP (variable importance in the projection) values with the loading plot, 20 metabolites with VIP being more than 1 were selected as biomarker metabolites. Among them, four metabolites still remained unknown; the relative contents of the identified metabolites in different planting regions are shown in Figure 6. The *t* tests showed that these biomarkers were all significantly different at least between two of the regions ($p \leq 0.05$). These biomarkers mainly included some primary metabolites (sucrose, D-fructose, D-mannose, D-glucose, inositol, maleic acid, citric acid, malic acid, L-threonic acid, L-proline, L-phenylalanine) as well as some secondary metabolites (chlorogenic acid, α- and β-4,8,13-duvatriene-1,3-diol, nicotine, quinic acid).

The accumulation of secondary metabolites in higher plants is a consequence of the interaction between plants and environments (biotic and abiotic) in the long-term process of plant evolution,[37] so changes in the levels of secondary metabolites are closely related to environmental factors, such as sun exposure time, light quality, temperature, and rainfall. According to the chemical structures and properties, the secondary metabolites included three major categories: phenolics, terpenes, and nitrogen compounds (mainly alkaloids). These metabolites play important roles in not only resistance, color, and signaling of tobacco leaf, but also the quality of tobacco leaf.[38]

Chlorogenic acid is the principle phenolic compound in tobacco leaf, which amounts to about 80% of the total phenolics. It was reported that the content of chlorogenic acid was positively correlated with tobacco quality.[39] It contributes to the color, antioxidant properties, and sensory properties of tobacco leaf.[40] For example, it could offer tobacco products a mild, sweet flavor and a baking spice scent.[41] A significantly higher level of chlorogenic acid was observed in tobacco leaf from Yunnan, which had the highest altitude, the longest sun exposure time, and the lowest average temperature. Moreover,

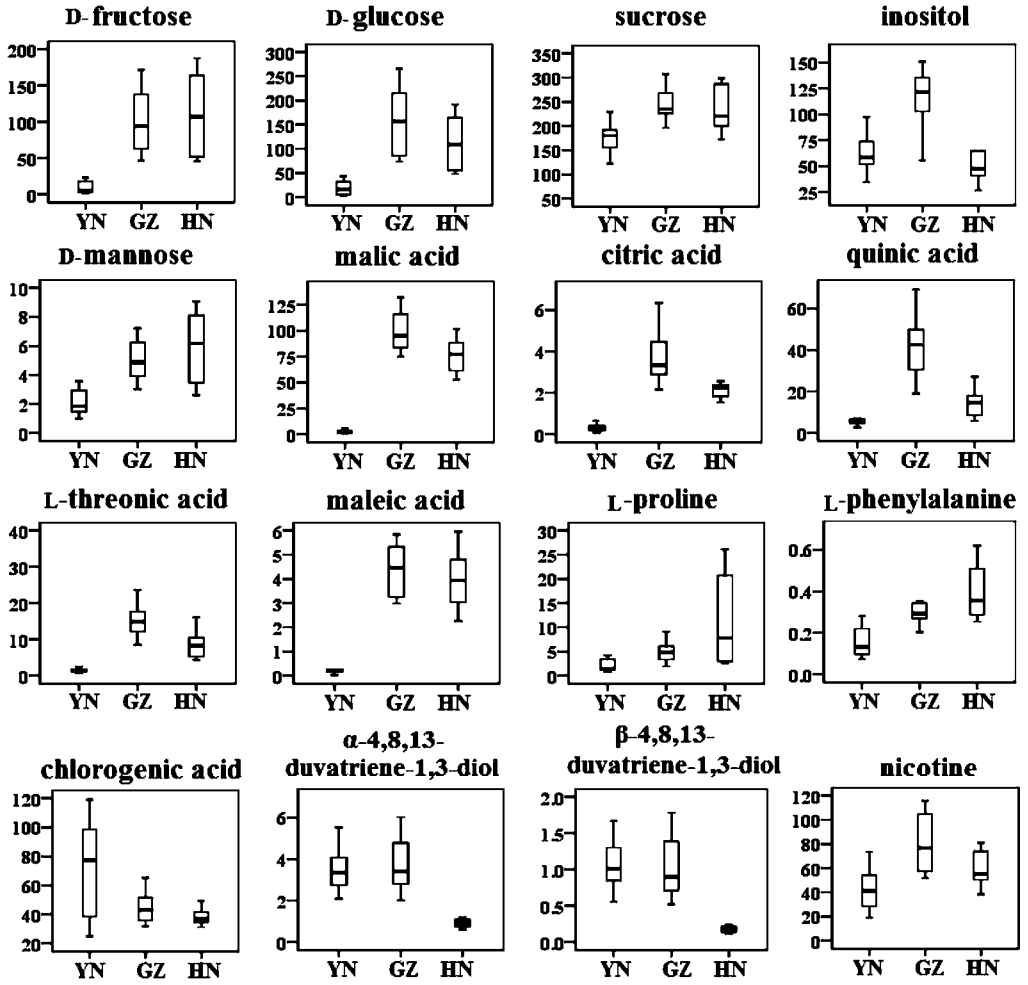

Figure 6. Box and whisker plots of significantly different metabolites. These primary and secondary metabolites were identified through commercial standards and were responsible for the discrimination of tobacco leaf of different geographical origins in the OPLS-DA model. YN, GZ, and HN represent Yunnan Province, Guizhou Province, and Henan Province, respectively.

the level of chlorogenic acid was lowest in tobacco leaf from Henan, which had the lowest altitude, the shortest sun exposure time, and the highest average temperature (Figures 1 and 6). The results suggest that chlorogenic acid increases with increasing altitude and sun exposure time and decreasing temperature. The effect of length of sun exposure time and altitude on phenolic synthesis has been reported by several authors. Tso et al.[7] suggested that chlorogenic acid increased with longer sun exposure time, and FR-irradiated plant tissue was higher in chlorogenic acid content than R-irradiated plant tissue. It was also reported that the level of chlorogenic acid was higher for tobacco leaf planted at higher altitude, which might be related to near-ultraviolet light and the intensity of visible light.[42−45] The present results from the three regions appear to be in agreement with these findings.

Cembranoids are the major components in the cuticular wax of *Nicotiana tabacum*, among which the content of α- and β-4,8,13-duvatriene-1,3-diol is highest in fresh tobacco leaf.[46] Although cembranoids are normally considered as lacking flavor properties, they have been proposed as precursors to solanone and its derivatives formed during curing and aging. Solanone and its derivatives are important neutral flavor constituents, which play important roles in improving the flavor quality of cigarettes.[47] Johnson et al.[48] showed that leaves from field plants contained much higher amounts of cembranoids than greenhouse plants of the same variety, which was thought to be related to the light transmission characteristics of glass in the wavelength region 290 to 330 nm. These findings may partly explain why the content of cembranoids in tobacco leaf from Henan is significantly lower than that from Yunnan and Guizhou. Yunnan and Guizhou both lie in the plateau, and the intensity of the UV radiation in these two regions is notably higher than that of Henan. However, the content of cembranoids in Yunnan did not significantly differ from that in Guizhou, although the altitude of Yunnan is higher than that of Guizhou. It might be because of the relatively lower average temperature in Yunnan.[49] These results showed that the accumulation of cembranoids increased with temperature and the intensity of the UV radiation. In addition, the heavy rainfall of Henan in the later stage of tobacco growth, which could wash away some of the cuticular wax on the tobacco leaf, is another factor in the lower content of cembranoids in Henan.

It is well known that tobacco usage is primarily due to the stimulant effect of nicotine, which plays an important role in the physiological strength of smoke. However, too much nicotine would increase the irritancy of smoke. It is reported that nicotine content of tobacco leaf increases with increased hours of sunshine.[7] In this study, the highest level of nicotine in fresh tobacco leaf was observed in Guizhou, which had the longest total sun exposure time in the later stage of tobacco

growth (July and August), and the lowest level was observed in Yunnan, which had the shortest sun exposure time in July and August. Moreover, the correlation coefficient between the average level of nicotine in the three regions and the total sun exposure time in July and August of these regions is as much as 0.998. This result showed that long sun exposure time in the later stage of tobacco growth could promote the accumulation of nicotine in tobacco leaf.

In summary, the results demonstrated that the combination of GC-MS-based metabolic profiling with multivariate analysis could be a useful tool for discriminating tobacco leaf of different geographical origins and determining the potential indicators of tobacco provenance. The PCA and HCA analyses showed that the metabolite levels of tobacco leaf were clearly dependent on geographical conditions, such as temperature, precipitation, and light. By further investigation of metabolic profiling using OPLS-DA, the metabolites discriminating the geographical origins were screened out. These metabolites, especially the secondary metabolites, were closely related to the flavor type of tobacco leaf. For example, higher levels of chlorogenic acid and α- and β-4,8,13-duvatriene-1,3-diol and the lower level of nicotine in Yunnan contributed to the clear flavor type of Yunnan. Therefore, this global metabolic profiling using GC-MS combined with multivariate statistics could be useful not only for understanding tobacco plant metabolism but also for assessing tobacco flavor quality. A further study using GC-MS combined with other analytical techniques would extend the coverage of the metabolome of tobacco leaf and provide much more useful biomarkers responsible for the discrimination of tobacco leaf from different geographical areas.

■ ASSOCIATED CONTENT

Supporting Information

Table S1: Detailed information of the identified metabolites in Table 2, including the retention index on GC stationary phases of different polarities (DB-5MS and DB-FFAP), mass fragments, classification, and derivatives; Table S2: Analytical precision of the GC-MS method and biological variation in tobacco leaf from Xiangyun, Yunnan Province; Figure S1: Plot of 16 runs of GC-MS generated by the first component of PCA. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*Tel: +86-0371-67672502. Fax: 86-0371-67672390. E-mail: xiefuwei@sina.com.

Funding

This work was supported by grants from National Natural Science Foundation of China (Grant Nos.: 21205140 and 21005087).

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We would like to thank Dr. Lixin Duan and Dr. John Hugh Snyder at Institue of Botany, the Chinese Acadamy of Sciences, for help in using the SIMCA-P13.0 software.

■ REFERENCES

(1) Zhu, X. L.; Zheng, F. G.; Cao, Z. J. World tobacco production report in 2005. *Acta Tabacaria Sin.* **2006**, *12*, 58−64.

(2) Smeeton, B. W. Genetical control of tobacco quality. *Recent Adv. Tobacco Sci.* **1987**, *13*, 3−26.

(3) Kasperbauer, M. J.; Tso, T. C.; Sorokin, T. P. Effects of end-of-day red and far-red radiation on free sugars, organic acids and amino acids of tobacco. *Phytochemistry* **1970**, *9*, 2091−2095.

(4) Sakaki, T.; Niino, K.; Sakuma, H.; Sugawara, S. Relationship between tobacco headspace volatiles and their smoking quality. *Agric. Biol. Chem.* **1986**, *50*, 317−323.

(5) Yang, H. Q.; Zhou, J. H.; Luo, Z. M.; Yang, S. Y. Study on chromoplast pigments and its degrading products of the flue-cured tobacco from different producting regions in China. *J. Southwest Agric. Univ. (Nat. Sci.)* **2004**, *26*, 640−644.

(6) Week, W. W. Chemistry of tobacco constituents influencing flavor and aroma. *Recent Adv. Tobacco Sci.* **1985**, *11*, 175−200.

(7) Tso, T. C.; Kasperbauer, M. J.; Sorokin, T. P. Effect of photoperiod and end-of-day light quality on alkaloids and phenolic compounds of tobacco. *Plant Physiol.* **1970**, *45*, 330−333.

(8) Xiang, G.; Yang, H. Y.; Yang, L.; Zhang, X.; Cao, Q.; Miao, M. M. Multivariate statistical analysis of tobacco of different origin, grade and variety according to polyphenols and organic acids. *Microchem. J.* **2010**, *95*, 198−206.

(9) Severson, R. F.; Arrendale, R. F.; Chortyk, O. T.; Johnson, A. W.; Jackson, D. M.; Gwynn, G. R.; Chaplin, J. F.; Stephenson, M. G. Quantitation of the major cuticular components from green leaf of different tobacco types. *J. Agric. Food Chem.* **1984**, *32*, 566−570.

(10) Wahlberg, I.; Karlsson, K.; Austin, D. J.; Junker, N.; Roeraade, J.; Enzell, C. R.; Johnson, W. H. Effects of flue-curing and ageing on the volatile, neutral and acidic constituents of Virginia tobacco. *Phytochemistry* **1977**, *16*, 1217−1231.

(11) Moritz, T.; Johansson, A. I. Plant Metabolomics. In *Metabolomics, Metabonomics and Metabolite Profiling*; Griffiths, W. J., Ed.; The Royal Society of Chemistry: London, UK, 2008; pp 254−268.

(12) Fiehn, O.; Kopka, J.; Dormann, P.; Altmann, T.; Trethewey, R. N.; Willmitzer, L. Metabolite profiling for plant functional genomics. *Nat. Biotechnol.* **2000**, *18*, 1157−1161.

(13) Fiehn, O. Metabolomics−the link between genotypes and phenotypes. *Plant Mol. Biol.* **2002**, *48*, 155−171.

(14) Liu, C. X.; Li, C.; Lin, D. H.; Song, N. N. Significance of metabonomics in drug discovery and development. *Asian J. Drug Metab. Pharmacokinet.* **2004**, *4*, 87−96.

(15) Duan, L. X.; Chen, T. L.; Li, M.; Chen, M.; Zhou, Y. Q.; Cui, G. H.; Zhao, A. H.; Jia, W.; Huang, L. Q.; Qi, X. Q. Use of the metabolomics approach to characterize Chinese medicinal material huangqi. *Mol. Plant* **2012**, *5*, 376−386.

(16) Pongsuwan, W.; Fukusaki, E.; Bamba, T.; Yonetani, T.; Yamahara, T.; Kobayashi, A. Prediction of Japanese green tea ranking by gas chromatography/mass spectrometry-based hydrophilic metabolite fingerprinting. *J. Agric. Food Chem.* **2007**, *55*, 231−236.

(17) Pongsuwan, W.; Bamba, T.; Yonetani, T.; Kobayashi, A. K.; Fukusaki, E. Quality prediction of Japanese green tea using pyrolyzer coupled GC/MS based metabolic fingerprinting. *J. Agric. Food Chem.* **2008**, *56*, 744−750.

(18) Pereira, G. E.; Gaudillere, J.-P.; Van Leeuwen, C.; Hilbert, G.; Lavialle, O.; Maucourt, M.; Deborde, C.; Moing, A.; Rolin, D. ^{1}H NMR and chemometrics to characterize mature grape berries in four wine-growing areas in Bordeaux, France. *J. Agric. Food Chem.* **2005**, *53*, 6382−6389.

(19) Kim, E. J.; Kwon, J.; Park, S. H.; Park, C.; Seo, Y.-B.; Shin, H.-K.; Kim, H. K.; Lee, K.-S.; Choi, S.-Y.; Ryu, D. H.; Hwang, G.-S. Metabolite profiling of *Angelica gigas* from different geographical origins using ^{1}H NMR and UPLC-MS analyses. *J. Agric. Food Chem.* **2011**, *59*, 8806−8815.

(20) Son, H.-S.; Kim, K. M.; van den Berg, F.; Hwang, G.-S.; Park, W.-M.; Lee, C.-H.; Hong, Y.-S. ^{1}H nuclear magnetic resonance-based metabolomic characterization of wines by grape varieties and production areas. *J. Agric. Food Chem.* **2008**, *56*, 8007−8016.

(21) Lee, J.-E.; Lee, B.-J.; Chung, J.-O.; Hwang, J.-A.; Lee, S.-J.; Lee, C.-H.; Hong, Y.-S. Geographical and climatic dependencies of green

tea (*Camellia sinensis*) metabolites: A ^{1}H NMR-based metabolomics study. *J. Agric. Food Chem.* **2010**, *58*, 10582−10589.
(22) Zhang, J. T.; Zhang, Y.; Du, Y. Y.; Chen, S. Y.; Tang, H. R. Dynamic metabonomic responses of tobacco (*Nicotiana tabacum*) plants to salt stress. *J. Proteome Res.* **2011**, *10*, 1904−1914.
(23) Choi, H. K.; Choi, Y. H.; Verberne, M.; Lefeber, A. W. M.; Erkelens, C.; Verpoorte, R. Metabolic fingerprinting of wild type and transgenic tobacco plants by ^{1}H NMR and multivariate analysis technique. *Phytochemistry* **2004**, *65*, 857−864.
(24) Fukusaki, E.; Kobayashi, A. Plant metabolomics: Potential for practical operation. *J. Biosci. Bioeng.* **2005**, *100*, 347−354.
(25) Roessner, U.; Wagner, C.; Kopka, J.; Trethewey, R. N.; Willmitzer, L. Simultaneous analysis of metabolites in potato tuber by gas chromatography−mass spectrometry. *Plant J.* **2000**, *23*, 131−142.
(26) Li, Y.; Pang, T.; Li, Y. L.; Wang, X. L.; Li, Q. H.; Lu, X.; Xu, G. W. Gas chromatography-mass spectrometric method for metabolic profiling of tobacco leaves. *J. Sep. Sci.* **2011**, *34*, 1447−1454.
(27) Lisec, J.; Schauer, N.; Kopka, J.; Willmitzer, L.; Fernie, A. R. Gas chromatography mass spectrometry-based metabolite profiling in plants. *Nat. Protoc.* **2006**, *1*, 387−396.
(28) Kim, H. K.; Verpoorte, R. Sample preparation for plant metabolomics. *Phytochem. Anal.* **2010**, *21*, 4−13.
(29) Gullberg, J.; Jonsson, P.; Nordstrom, A.; Sjostrom, M.; Moritz, T. Design of experiments: an efficient strategy to identify factors influencing extraction and derivatization of *Arabidopsis thaliana* samples in metabolomic studies with gas chromatography/mass spectrometry. *Anal. Biochem.* **2004**, *331*, 283−295.
(30) Fiehn, O.; Kopka, J.; Trethewey, R. N.; Willmitzer, L. Identification of uncommon plant metabolites based on calculation of elemental compositions using gas chromatography and quadrupole mass spectrometry. *Anal. Chem.* **2000**, *72*, 3573−3580.
(31) Schauer, N.; Steinhauser, D.; Strelkov, S.; Schomburg, D.; Allison, G.; Moritz, T.; Lundgren, K.; Roessner-Tunali, U.; Forbes, M. G.; Willmitzer, L.; Fernie, A. R.; Kopka, J. GC-MS libraries for the rapid identification of metabolites in complex biological samples. *FEBS Lett.* **2005**, *579*, 1332−1337.
(32) Molyneux, R. J.; Schieberle, P. Compound identification: A journal of agricultural and food chemistry perspective. *J. Agric. Food Chem.* **2007**, *55*, 4625−4629.
(33) Noda, I. Scaling techniques to enhance two-dimensional correlation spectra. *J. Mol. Struct.* **2008**, *883*−*884*, 216−227.
(34) Stedman, R. L. Chemical composition of tobacco and tobacco smoke. *Chem. Rev.* **1968**, *68*, 153−207.
(35) Tang, Y. J. On aroma type of flue-cured tobacco. *Chin. Tobacco Sci.* **2011**, *32*, 1−7.
(36) Huang, Z. Y.; Zhu, Y.; Wang, S. H.; Li, T. F.; Deng, Y. L.; Sao, Y. Relationship between tobacco quality and climates in Yunnan. *Resources Sci.* **2007**, *29*, 83−90.
(37) Cutler, H. G.; Severson, R. F.; Cole, P. D.; Jackson, D. M.; Johnson, A. W. Secondary Metabolites from Higher Plants. In *Natural Resistance of Plants to Pests*; American Chemical Society: Washington, DC, 1986; Vol. *296*, pp 178−196.
(38) Nugroho, L.; Verpoorte, R. Secondary metabolism in tobacco. *Plant Cell Tiss. Org.* **2002**, *68*, 105−125.
(39) Bazinet, L.; DeGrandpré, Y.; Porter, A. Electromigration of tobacco polyphenols. *Sep. Purif. Technol.* **2005**, *41*, 101−107.
(40) Wang, H.; Zhao, M.; Yang, B.; Jiang, Y.; Rao, G. Identification of polyphenols in tobacco leaf and their antioxidant and antimicrobial activities. *Food Chem.* **2008**, *107*, 1399−1406.
(41) Kallianos, A. G. Phenolics and acid in leaf and their relationship to smoking and aroma. *Recent Adv. Tobacco Sci.* **1976**, *20*, 61−79.
(42) Penn, P. T.; Stephens, R. L.; Weybrew, J. A. Some factors affecting the content of principal polyphenols in tobacco leaves. *Tobacco Sci.* **1958**, *2*, 68−72.
(43) Zucker, M.; Nitsch, C.; Nitsch, J. P. The induction of flowering in Nicotiana II: Photoperiodic alternation of the chlorogenic acid concentration. *Am. J. Bot.* **1985**, *52*, 271−277.
(44) Andersen, R. A.; Kasperbauer, M. J. Effects of near-ultraviolet radiation and temperature on soluble phenols in *Nicotiana tabacum*. *Phytochemistry* **1971**, *10*, 1229−1232.
(45) Andersen, R.; Kasperbauer, M. J. Chemical composition of tobacco leaves altered by near-ultraviolet and intensity of visible light. *Plant Physiol.* **1973**, *51*, 723−726.
(46) Roberts, D. L.; Rowland, R. L. Macrocyclic diterpenes. α- and β-4,8,13-duvatriene-1,3-diols from tobacco. *J. Org. Chem.* **1962**, *27*, 3989−3995.
(47) Gamou, K.; Kawashima, N. Studies on leaf surface lipid of tobacco. I. Changes in leaf surface lipid and duvatrienediol during growth, senescence and curing of tobacco leaves. *Agric. Biol. Chem.* **1979**, *43*, 2163−2168.
(48) Johnson, A. W.; Severson, R. F.; Jackson, D. M.; Gwynn, G. R.; Chaplin, J. F. In *Tobacco budworm infestation on early, late-planted, and sucker regrowth resistant and susceptible tobacco*, 29th Tobacco Workers Conference, Lexington, KY, January 19−22, 1981.
(49) Severson, R. F.; Johnson, A. W.; Jackson, D. M. Cuticular constituents of tobacco: Factors affecting their production and their role in insect and disease resistance and smoke quality. *Recent Adv. Tobacco Sci.* **1985**, *11*, 105−174.