Use of the Metabolomics Approach to Characterize Chinese Medicinal Material Huangqi

Li-Xin Duan, Tian-Lu Chen, Ming Li, Min Chen, Ye-Qing Zhou, Guang-Hong Cui, Ai-Hua Zhao, Wei Jia, Lu-Qi Huang, and Xiaoquan Qi

The Key Laboratory of Plant Molecular Physiology, Institute of Botany, Chinese Academy of Sciences, 20 Nanxincun Xiangshan Road, Beijing 100093, China

Shanghai Center for Systems Biomedicine (Ministry of Education Key Laboratory of Systems Biomedicine), Shanghai Jiao Tong University, 800 Dongchuan Road, Shanghai 200240, China

Institute of Chinese Materia Medica, Chinese Academy of Chinese Medical Sciences, 16 Nanxiao Road, Beijing 100700, China

School of Pharmacy, Shanghai Jiao Tong University, 800 Dongchuan Road, Shanghai 200240, China

ABSTRACT Integration of the genetic and metabolic fingerprinting can provide a new approach to differentiate similar Traditional Chinese Medical (TCM) materials. Two leguminous plants, Mojia Huangqi and Menggu Huangqi, are important medical herbs and share great similarities in morphology, chemical constituent, and genomic DNA sequence. The taxonomy of Mojia Huangqi and Menggu Huangqi has been debated for more than 50 years and discrimination of TCM materials directly affects the pharmacological and clinical effects. AFLP based genetic fingerprinting and GC–TOF/MS-based metabolic fingerprinting were used to successfully discriminate the two species. The results of AFLP supported the opinion that Menggu Huangqi was a variant of Mojia Huangqi. The metabolic fingerprinting showed growth locations have greater impacts on the metabolite composition and quantity than the genotypes (cultivated versus wild) in Menggu Huangqi. The difference of some soluble sugars, fatty acids, proline, and polyamine reflected plant adaptation to different growth environments. Using multivariate and univariate statistical analysis, three AFLP markers and eight metabolites were identified as candidate DNA and metabolic markers to distinguish the two herb materials. The correlation network between AFLP markers and metabolites revealed a complex correlation network, which indicated the special metabolic pathways and the regulation networks of Huangqi.

Key words: Astragalus membranaceus; metabolic fingerprinting; amplified fragment length polymorphisms; gas chromatography time-of-flight mass spectrometry.

INTRODUCTION

Astragalus L. is the largest genus in the Leguminosae family, comprising ~2000 species, grouped into more than 100 subdivisions, and there are 278 species in China (Fu, 1993). Two species, Menggu Huangqi and Mojia Huangqi, which belong to Subgenus Phaca (Linn.) Bunge, are widely used as Traditional Chinese Medicinal herbs for anti-perspirant and anti-diuretic (Cui et al., 2003; Ma et al., 2004; Cho and Leung, 2007; Kuo et al., 2009). The taxonomy of Menggu Huangqi and Mojia Huangqi has been debated for more than 50 years. The most clear morphological difference between the two species is the presence or absence of hair on ovary and pod. In 1964, Peigeng Xiao et al. (1964) argued that the distribution of Menggu Huangqi and Mojia Huangqi was not completely isolated. Some transitional herbaria suggested the possibility of hybridization between them. The flower, fruit, and other major genital organs did not show significant difference. So he believed Menggu Huangqi should be the variety of Mojia Huangqi and suggested naming Menggu Huangqi as A. membranaceus (Fisch.) Bge. var. mongholicus (Bge.) Hsiao and to keep A. membranaceus (Fisch.) Bge for Mojia Huangqi. Nowadays, this has become the official standard and has been adopted in ‘Flora of China’ and Chinese Pharmacopoeia (2005). However, Wang and Liu (1996) believed that the difference in flowering time caused the reproductive isolation, suggesting restoring Menggu Huangqi to independent species.

DNA fingerprinting as a tool enhances our understanding of the genetic control of morphological traits and has
applications in determining genetic diversity and distinctness (Schoen and Brown, 1991; Turpeinen et al., 2003) and also in determining phylogeny of species and populations (Wang et al., 1992; Muminovic et al., 2004). Thus, genetic fingerprinting provides a faster, more discriminating and more cost-effective method of plant identification than the analysis of morphological characters. However, Yip and Kwan (2006) could not distinguish the two Huangqi types using the internal transcribed spacer 1 (ITS1) sequence of the nuclear ribosomal RNA gene. Using correct DNA sequences is critical in plant identification when a large number DNA sequences is not available for a species. Amplified Fragment Length Polymorphism (AFLP) technology (Zabeau and Vos, 1993) has been proved to be a robust DNA fingerprinting and marker system in distinguishing plant genotypes with no pre-requrement of the genome sequence (Vos et al., 1995) and is widely used for polymorphism detection (Qi and Lindhout, 1997), construction of genetic linkage maps (Qi et al., 1998), and mapping loci for quantitative traits (Qi et al., 1998b, 1999, 2000).

Apart from genetic differences at species, variety, and geno-type levels, environments including growing locations and conditions (e.g. cultivated and wild) directly affect the levels of active components, and pharmacological and clinical effects (Obadovic et al., 2007). Consistency in contents and compositions of active compounds in the crude medical materials are the prerequisite for chemical and pharmacological investigations. The traditional morphological identification needs sophisticated experience and is insufficient. Metabolomics or metabolonomics, the analysis of total population of metabolites in a given biological sample, and the integration of the data in the context of functional genomics are an important field of ‘omics’ studies and metabolite levels reflect the genetic morphology and its interactions with the environments (Fiehn, 2002). Metabolomics analysis has been successfully applied to discriminate and differentiate plant phenotypes (Fiehn et al., 2000; Roessner et al., 2001; Choi et al., 2004) or the latent phenotypes associated with silent mutations (Blaise et al., 2007). It has been long recognized that the efficacy of Chinese medicinal herbs is usually attributed to multiple compounds, but not a single compound. Plant metabolomics is a promising tool to understand the intervention mechanisms of Traditional Chinese Medicine. Gas chromatography time-of-flight mass spectrometry (GC–TOF/MS) technology, which has a fast scanning speed and improves the deconvolution of complex mixtures such as plant extracts (Lisec et al., 2006), was preferentially used in plant metabolomics (Catchpole et al., 2005; Cook et al., 2004; Fiehn et al., 2008; Lee and Fiehn, 2008).

In this study, we conducted DNA and metabolite fingerprinting analyses on a set of 48 collections of the two types of Huangqi from different locations and genotypes (wild and cultivated) (Table 1) by using AFLP technology and GC–TOF/MS, respectively, aiming to solve the taxonomy uncertainty of the two Huangqi types and to explore the metabolomics approach on discrimination and study of Traditional Chinese Medical materials.

### RESULTS AND DISCUSSION

#### Species Discrimination by Use of DNA and Metabolic Fingerprints

The AFLP fingerprinting analysis by using 11 primer pairs had shown that the majority of AFLP bands are common to the two groups of Huangqi. This indicates that the two Huangqi are closely related and could belong to the same species. A total of 85 polymorphic AFLP bands ranging from 100 to 500 bp were scored. The average number of polymorphic bands was 7.7, with a minimum number of 3 to a maximum number of 13 polymorphic bands per primer pair. The 48 Huangqi collections were unambiguously grouped into three main clusters (Figure 1A). All collections from Menggu Huangqi were grouped into Cluster I, of which collections from different growth locations or the plants from cultivated and wild could not be further clustered into subgroups. While collections of the Mojia Huangqi from two different growth locations and conditions (north-west, plants from wild versus north-east, cultivated plants) were grouped into two clusters: Cluster II and Cluster III, respectively. Cluster analysis shown that Mojia Huangqi from wild in the north-west (Cluster II) is closer to Menggu Huangqi from cultivated and wild in both the north-west and the middle of the north (Cluster I) than Mojia Huangqi from cultivated in the north-east (Cluster III). This relationship may reflect the geographic locations at which the plant is cultivated, from Shanxi (MN) and Gansu (NW), which are neighboring provinces. Gansu province is the largest Huangqi cultivation area, where the two Huangqi have been cultivated together for a long time. Exchange of genes could have happened very often among the two Huangqi groups. Indeed, the collections from Jilin province (NE) were clearly distinguished from collections from the north-west and the middle north. Other DNA fingerprinting analysis by using the arbitrarily primed polymerase chain reaction (APPCR) method also found that the samples from Heilongjiang province (near the Jilin province) were obviously distinguished from samples of the north-west and the middle north locations (Yip and Kwan, 2006). These results support the opinion that Menggu Huangqi is the variety of Mojia Huangqi (Xiao

<table>
<thead>
<tr>
<th>Index</th>
<th>Species</th>
<th>Growing locations</th>
<th>Genotypes</th>
</tr>
</thead>
<tbody>
<tr>
<td>A*</td>
<td>Mojia Huangqi</td>
<td>Jilin Sipin (JL)</td>
<td>Cultivated (C)</td>
</tr>
<tr>
<td>B*</td>
<td>Mojia Huangqi</td>
<td>Jilin Tonghua (JL)</td>
<td>Cultivated (C)</td>
</tr>
<tr>
<td>C*</td>
<td>Mojia Huangqi</td>
<td>Gansu Weiyan (GS)</td>
<td>Wild (W)</td>
</tr>
<tr>
<td>D*</td>
<td>Mojia Huangqi</td>
<td>Gansu Zhangxian (GS)</td>
<td>Wild (W)</td>
</tr>
<tr>
<td>E*</td>
<td>Menggu Huangqi</td>
<td>Gansu Longxi (GS)</td>
<td>Cultivated (C)</td>
</tr>
<tr>
<td>F*</td>
<td>Menggu Huangqi</td>
<td>Shanxi Hunyuan (SX)</td>
<td>Cultivated (C)</td>
</tr>
<tr>
<td>G*</td>
<td>Menggu Huangqi</td>
<td>Gansu Zhuangxian (GS)</td>
<td>Cultivated (C)</td>
</tr>
<tr>
<td>H*</td>
<td>Menggu Huangqi</td>
<td>Shanxi Yinxing (SX)</td>
<td>Wild (W)</td>
</tr>
</tbody>
</table>

* The north east of China; △ the north west of China; ◇ the middle of north China.
et al., 1964) and it is appropriate to name Menggu Huangqi as *A. membranaceus* (Fisch.) Bge. var. *mongholicus* (Bge.) Hsiao and to keep *A. membranaceus* (Fisch.) Bge for Mojia Huangqi. Our data demonstrated that AFLP is efficient DNA fingerprinting technology to reveal the genetic relationship of two Huangqi and has been used to successfully solve the taxonomy debate about Menggu Huangqi and Mojia Huangqi. Among the 85 polymorphic AFLP bands, M40E41-5 and M33E41-2 exist only in Mojia Huangqi collections, and M38E35-3 was identified only in Menggu Huangqi collections. So, M40E41-5, M33E41-2, and M38E35-3 could be the candidate AFLP markers for differentiation of the two Huangqi species.

A total of 1 193 peaks were quantified from the non-polar (759 peaks) and polar profiles (434 peaks) by use of GC–TOF/MS. Compound identification was performed by comparing the mass fragments with NIST 05 standard mass spectral databases and the Golm Metabolome Datebase (Kopka et al., 2005) in NIST MS search 2.0 (NIST, Gaithersburg, MD) software with a similarity of more than 80%. In total, 118 compounds were tentatively identified, 86 compounds and 54 compounds existing in polar phase and lipophilic phase, respectively, with 22 compounds existing in both of the two phases, such as Glycine (3TMS), L-Asparagine (3TMS), and L-Homoserine (2TMS) (Supplemental Table 1). Cluster analysis using these metabolite data assigned the 48 Huangqi collections into two distinguish clusters (Figure 1B). Cluster I contains 22 collections of Menggu Huangqi and Cluster II contains all 24 collections of Mojia Huangqi and two collections of Menggu Huangqi. This result indicates that substantial differences in metabolites exist in Mojia Huangqi and Menggu Huangqi. The cluster analysis shown a certain influence on metabolite composition by growth locations; for example, collections of Mojia Huangqi collected from the north east (NE) and the north west (NW) have clustered together, respectively, in Cluster II.

The cluster analyses have demonstrated that genetic markers seemed more reliable in distinguishing Mojia Huangqi species than metabolic profiles. However, it is difficult to use genetic markers to inspect the quality of Chinese medicinal herbs, since functional compounds in the same herb significantly depend on growth conditions. Thus, metabolic profiling...
is the right tool to evaluate the quality of Chinese medicinal
herbs, although repeats from same sample were not tightly
clustered together. The levels of metabolites are hyper-sensi-
tive to environmental factors and growth stages in comparison
to AFLP markers. In this study, six biological repeats were col-
lected from slightly different locations and, certainly, there
were certain variations in growth age among the repeats.
So, four to six biological repeats are required in the metabo-
lomics analysis to provide more reliable data.

The Effect of Growth Locations and Conditions on the
Huangqi Metabolism

To reveal the effect of environments on the metabolites, Mojia
Huangqi and Menggu Huangqi were analyzed separately. An
unsupervised principal component analysis (PCA) model was
used, since it can represent the intrinsic characteristics of
the data. In Mojia Huangqi, the wild samples from Gansu prov-
ince (NE) and the cultivated samples from Jilin province (NE)
were separated by the third principle component (PC 3) (Figure
2A). This result is consistent with the result obtained by cluster
analysis of AFLP makers (Figure 1A). Due to a lack of cultivated
collections from Gansu province (NE) and wild collections from
Jilin province (NE), it is hard to justify whether these differences
are caused by the environmental factors or the geographical
isolations.

In Menggu Huangqi, all samples were collected from two
neighboring provinces: Gansu (NW) and Shanxi (MN). The PCA
score plot revealed that the collections collected from Shanxi
province (MN, marked by black dot and blue dot) and Gansu
province (NW, marked by red dot) are separated by the fifth com-
ponent (PC 5), except two individual collections (Figure 2B). How-
ever the collections grouped by growing conditions were mixed
in the PCA score plot at any two components (red dot and blue
dot). These results show that growth locations may have greater
impacts on the metabolite composition and quantity than the
genotypes (cultivated versus wild) in Menggu Huangqi.

Chinese medicinal herbs in different growing localities are
different in therapeutic potency. Environmental conditions such
as light/dark, drought, temperature, nutritional supply, micro-
bial invasion, etc. affect the levels of metabolites. Previous re-
search demonstrated that the main constituents of Menggu
Huangqi changed according to seasonal variation and age of
the plant (Ma et al., 2002). Owing to the intrinsic bias of GC–
TOF/MS to secondary metabolites, some active metabolites,
such as isoflavonoids and saponins, cannot be detected and
identified in this analysis. It is conductive to use other methods
such as the LC–MS system to complement and perfect this study.
Also, a designed experiment with a set of identical samples
grown in different geographical locations would give direct
comparisons among the different environmental factors in
Figure 2), in the projection) value of OPLS–DA model consequently, relying on the three criteria—VIP (variable importance validated the stability and reliability of this OPLS–DA model. Sub-are 0.349, 0.962, and 0.940, respectively. The permutation result

acid, Xylose, and Pentonic acid) could be presumably considered as candidate biomarkers (Table 2). The starting letters of the in-

olites (including the tentatively identified metabolites, Malonic

tation irrelevant to species was removed prior to model building deriving from a more sophisticated OPLS–DA (orthogonal partial

total 64.5% of the variance of this dataset. The two Huangqi

by genes in the genome. The partial correlation analysis

Biosynthesis and accumulation of metabolites are mainly controlled by enzymes and regulation factors, which are encoded

Duan et al. • Metabolomics to Characterize Medicinal Materials | 5

Table 2. A List of Significant Difference Compounds for Species Discrimination.

<table>
<thead>
<tr>
<th>Index</th>
<th>Rt(min)</th>
<th>VIP</th>
<th>Name</th>
<th>Similarity</th>
<th>MW-U*</th>
<th>ROC^*</th>
<th>Mean (MJ)</th>
<th>Mean(MG)</th>
<th>Ratio(MJ/MG)</th>
</tr>
</thead>
<tbody>
<tr>
<td>p54^h</td>
<td>10.181</td>
<td>2.177</td>
<td>Malonic acid</td>
<td>954</td>
<td>&lt;0.001</td>
<td>1.000</td>
<td>152.819</td>
<td>23.971</td>
<td>6.375</td>
</tr>
<tr>
<td>o540^h</td>
<td>33.034</td>
<td>2.135</td>
<td>–</td>
<td>–</td>
<td>&lt;0.001</td>
<td>1.000</td>
<td>6.765</td>
<td>49.950</td>
<td>0.135</td>
</tr>
<tr>
<td>p15</td>
<td>6.214</td>
<td>2.120</td>
<td>–</td>
<td>–</td>
<td>&lt;0.001</td>
<td>1.000</td>
<td>9.103</td>
<td>0.955</td>
<td>9.529</td>
</tr>
<tr>
<td>p48</td>
<td>9.343</td>
<td>2.107</td>
<td>–</td>
<td>–</td>
<td>&lt;0.001</td>
<td>1.000</td>
<td>21.718</td>
<td>7.886</td>
<td>2.754</td>
</tr>
<tr>
<td>o546</td>
<td>33.418</td>
<td>2.072</td>
<td>–</td>
<td>–</td>
<td>&lt;0.001</td>
<td>1.000</td>
<td>1.259</td>
<td>15.401</td>
<td>0.082</td>
</tr>
<tr>
<td>p188</td>
<td>22.019</td>
<td>1.963</td>
<td>D-Xylose</td>
<td>950</td>
<td>&lt;0.001</td>
<td>0.993</td>
<td>4.568</td>
<td>0.407</td>
<td>11.223</td>
</tr>
<tr>
<td>p122</td>
<td>16.284</td>
<td>1.854</td>
<td>–</td>
<td>–</td>
<td>&lt;0.001</td>
<td>1.000</td>
<td>12.391</td>
<td>0.818</td>
<td>15.155</td>
</tr>
<tr>
<td>p205</td>
<td>23.924</td>
<td>1.851</td>
<td>Pentonic acid</td>
<td>903</td>
<td>&lt;0.001</td>
<td>0.958</td>
<td>12.970</td>
<td>2.355</td>
<td>5.507</td>
</tr>
</tbody>
</table>

* P-value of Mann-Whitney test.
∧ Area under the Receive Operation curve.
Δ The starting letters of the index ‘o’ and ‘p’ meant the metabolites from the organic (non-polar) and polar phase, respectively.

different locations and provide more reliable results on how the metabolism is influenced by environmental factors.

Identification of Metabolites Differentially Accumulated in Two Huangqi

The cluster analysis based on the metabolite data can generally group the 48 Huangqi collections into two clusters. The PCA model was also applied to obtain a preliminary overview of general similarities and differences between collections. The 48 collections of Huangqi were clearly separated into Mojia Huangqi (marked by a red dot) and Menggu Huangqi (marked by a black dot) on the PCA score plot with its first two components (Supplemental Figure 1). The first eight principal components derived from the above data matrix encompassed in total 64.5% of the variance of this dataset. The two Huangqi can be separated clearly by only one predictive component deriving from a more sophisticated OPLS–DA (orthogonal partial least squares discriminate analysis) model, since noisy information irrelevant to species was removed prior to model building (Bylesjö et al., 2006) (Figure 3). The R^2X, R^2Y, and Q^2 of this model are 0.349, 0.962, and 0.940, respectively. The permutation result validated the stability and reliability of this OPLS–DA model. Subsequently, relying on the three criteria—VIP (variable importance in the projection) value of OPLS–DA model > 1.8 (Supplemental Figure 2), P-value of Mann-Whitney test < 0.001, and AUC > 0.8 (area under the ROC curve, see data procession)—eight metabolites (including the tentatively identified metabolites, Malonic acid, Xylose, and Pentonic acid) could be presumably considered as candidate biomarkers (Table 2). The starting letters of the index ‘o’ and ‘p’ meant the metabolites from the organic (non-polar) and polar phase. The column ‘match’ presents the identification similarity of the detected compound with the libraries. The last two columns list the average spectrum intensities for different species of every metabolite.

Obviously, in the analysis of metabolite data, PCA and OPLS–DA are more powerful than cluster analysis. Moreover, the candidate biomarkers provide a basis for the development of simple and fast assay methods for the discrimination of Huangqi materials. Further analysis of the differentially accumulated metabolites in Mojia Huangqi and Menggu Huangqi would reveal the different metabolism in the roots of these two groups of Huangqi.

Differences in Metabolism of the Two Huangqi Groups

The tentatively identified compounds were assigned in the common metabolic pathways according to the literature works (Schauer et al., 2006; Kusano et al., 2007). The amounts of metabolites in Mojia Huangqi and Menggu Huangqi were compared. Metabolites with peak intensity ratios that were significantly different (P < 0.01) between the two Huangqi were labeled red or blue (Figure 4A). The metabolites with gray characters were extended ones. Fatty acids, proline, and metabolites in polyamine metabolic pathways were obviously decreased, but some soluble sugars, such as mannose, xylose, malonate, and pentanate were increased in Mojia Huangqi in comparison with the Menggu Huangqi. However, a few changes were observed in different geographic locations (Jilin province versus Gansu province) in Mojia Huangqi (Figure 4B). The major phytosterols were significantly decreased in Gansu province; in contrast, homoserine and pyruvate were increased in Jilin province.

Huangqi mainly grow from the north-west to the north-east of China. Mojia Huangqi usually grows at higher elevations, with high humus soil with a cold climate and abundant rainfall, and Menggu Huangqi normally grows on droughty and sandy soil in loess plateau (Xie et al., 2005). Plants often accumulate polyamines and proline in response to abiotic and biotic stresses. Research has indicated that the accumulation of polyamines improves drought tolerance (Kusano et al., 2008) and proline can influence osmotic stress tolerance in multiple ways (Szabados and Savouré, 2010). Changes in sugar content were related to variations in freezing tolerance (Wanner and Junttila, 1998). Sugars protect the plasma membranes and proteins from freezing and dehydration. The accumulation of different metabolites in the roots of different Huangqi collections may reflect their adaptation to different environments.

Association of AFLP Markers with Metabolic Markers

Biosynthesis and accumulation of metabolites are mainly controlled by enzymes and regulation factors, which are encoded by genes in the genome. The partial correlation analysis...
Figure 4. Levels of Metabolites in Main Metabolic Maps.

(A) The comparisons of metabolite contents were calculated by dividing the metabolite level in Menggu Huangqi with that in Mojia Huangqi.

(B) The comparisons of metabolite levels of Jilin province to Gansu province in Menggu Huangqi collections. The level of significance was set at $P < 0.01$. \(\alpha\)-KG, alpha-ketoglutarate; Arg, arginine; Asn, asparagine; Asp, aspartate; \(\beta\)-Ala, beta-alanine; \(\beta\)-Sito, beta-sitosterol; Camp, campesterol; CitA, citrate; FatA C16:0, n-hexadecanoate; FatA C18:0, stearate; FatA C18:2, linoleate; FatA C18:3, linolenate; C20:0,
between metabolic peaks and AFLP markers has revealed this relationship. The Pearson partial correlation analysis between 1193 metabolic peaks and 85 AFLP markers has identified 122 metabolites and 21 AFLP markers that show significant correlations ($r > 0.95$ or $r < -0.95$, $P < 0.001$, $n = 24$) (Figure 5). One big correlation network and eight simple correlation groups were obtained. One to several metabolites were controlled and regulated by each locus represented by the AFLP marker. In the biggest correlation network, AFLP markers M33E38-2 and M38E35-2 were the most important notes that correlated with 15 metabolites and five other AFLP markers, namely M40E37-11, M33E39-1, M40E38-1, M33E41-5, and M33E41-9. Highly correlated AFLP markers indicate that these markers may be closely linked and located on the same chromosome region. This may represent an important region controlling and regulating many metabolites that were detected in this study. The identified metabolic markers o540 and o546 for discrimination of the two Huangqi were significantly correlated with M33E41-9. The correlation network among metabolites and DNA markers provided a basis for deeper insight into metabolic pathways and the regulation networks.

**CONCLUSION**

The integrated genetic and metabolic fingerprinting methods have demonstrated their potential for discriminating different Traditional Chinese Medical materials. Basically, the two methods elicit similar results on the levels of species varieties,

![Figure 5](https://example.com/final-graph.png)

*Figure 5. The Significant Correlation Network of Metabolites and AFLP Marks ($r > 0.95$ or $r < -0.95$, $P < 0.001$, $n = 24$). The Cytoscape software was used to visualize the network. The pink line indicates a positive correlation and the green line shows a negative correlation.*

Eicosanoic acid; FatA C24:0, n-tetracosanoate; Fru, fructose; Fru6P, fructose 6-phosphate; FumA, fumarate; GABA, gamma-amino-n-butyrate; Glc, glucose; Glc6P, glucose-6-phosphate; Gin, glutamine; Glu, glutamate; Gly, glycine; Gly3P, glycerol-3-phosphate; GlycA, glyc erate; hSer, homoserine; Ile, isoleucine; IsocitA, isocitric acid; Leu, leucine; Lys, lysine; MalA, malate; Man, mannose; Orn, Ornithine; PEP, Phosphoenol-pyruvate; Pro, proline; Put, putrescine; Ser, serine; Spd, spermidine; Stigm, stigmasterol; Suc, sucrose; SucA, succinate; Thr, threonine; ThrA, threonate; Trp, tryptophan; Val, valine.
growing locations, and growing conditions (wild and cultivated). The AFLP-based fingerprinting can identify the genotypes unequivocally, while the GC–TOF/MS-based metabolic fingerprinting is reliable for rapid analysis of a large number of herbal materials, revealing not only the varieties of genome and the environment, but also potential differential metabolites. The metabolic profiling could be easily used to discriminate the two species and the contents of polyamines and sugars in main metabolic maps revealed obvious differences related to their growing environment. Based on the multivariate and univariate statistical analysis, three AFLP markers and eight metabolic markers were considered as DNA markers and biomarkers to distinguish the two herb materials.

**METHODS**

**Plant Materials**

Eight samples of the two types of Huangqi were collected from different locations in the north-west Gansu province, the northeast Jilin province, and the middle of north China Shanxi province, representing large varieties of environments. Both wild and cultivated medical herbs were included in the collections (Table 1). Six collections (repeats) were obtained for each sample. These collections were individual plants that grew in the same province, but different counties. Identification of Menggu Huangqi and Mojia Huangqi was according to their morphology, as described by Xuefeng Feng (PhD of the Institute of Chinese Materia Medica, China Academy of Chinese Medical Sciences). In total, 48 collections of Menggu Huangqi and Mojia Huangqi were used for DNA and metabolic fingerprinting analysis.

**Chemicals**

Methanol and chloroform (HPLC grade) were purchased from Fisher Scientific (Hampton, NH). Pyridine for GC grade, N-Methyl-N-(trimethylsilyl) trifluoroacetamide (MSTFA reagent), methoxyamine reagent, and internal standards references ribitol and nonadecanoic acid were purchased from Sigma-Aldrich (MO). Distilled water was purified ‘in-house’ using a Milli-Q system Millipore (MA).

**Metabolites Extraction and Derivatization**

Metabolite extraction was performed according to the protocol described as in the references (Lisec et al., 2006; Fiehn et al., 2008; Lee and Fiehn, 2008). Radixes of Menggu Huangqi and Mojia Huangqi were frozen with liquid nitrogen and ground to powder with a mortar and pestle. One hundred milligrams of powder were used for extraction of metabolites. 1.5 ml extraction buffer (methanol/chloroform/water, 5:2:2) and 10 μl nonadecanoic acid (2.10 mg ml

300 μl chloroform, and 600 μl dH₂O, and subsequently centrifuged at 4 000 rpm for 5 min. The upper phase (polar phase) 100 μl and the lower phase (non-polar phase) 100 μl were reduced to dry under nitrogen gas stream and in a vacuum rotary evaporator without heating, respectively. The dried residue were oximated with 40 μl methoxylamine hydrochloride (20 mg ml

The polar and non-polar phases were analyzed using a LECO Pegasus IV gas chromatography time-of-flight mass spectrometry (GC–TOF/MS) system. The GC–TOF/MS system was composed of an Agilent autosampler, a gas chromatograph 6890 (Agilent Corporation, USA), and a LECO Pegasus IV time-of-flight mass spectrometer (LECO Corporation, USA). One microliter of each derivatized sample was injected by autosampler into a gas chromatograph equipped with a 30 m × 0.25 mm i.d. fused silica capillary column with a chemically bonded 0.25-μm DB-5 MS stationary phase (J&W ScientiWc, USA). The injector temperature was 280°C. The Helium gas flow rate through the column was 1.5 ml min

2.0 mA, and 20 spectras per second were recorded in the mass range 80–500 m/z. The acceleration voltage was turned on after a solvent delay of 300 s. The detector voltage was 1700 V.

**DNA Extraction**

Isolation of genomic DNA followed the protocol (Murray and Thompson, 1980; Pattanayak et al., 2000) with a few modifications. The root powder (100 mg) was dispersed in 1 ml pre-heated CTAB buffer (10 g CTAB, 40 g sodium chloride in a total volume of 500 ml of 10 mM EDTA, 50 mM Tris, pH 8.0) and was incubated in a 65°C water bath for 1 h, cooled to room temperature, and centrifuged at 12 000 rpm for 15 min. The supernatant was transferred to another centrifuge tube and was extracted with the same volume of saturat hydroxysilic acid/water/chloroform/isoamyl alcohol (25:24:1). Phases were separated by centrifugation at 12 000 rpm for 10 min at 4°C. The upper phase was extracted twice with the same volume chloroform/isoamyl alcohol (24:1). The DNA was precipitated by adding a two-thirds volume of ice-cold isopropanol and recovered by centrifugation at 12 000 rpm for 1 min. The DNA pellet was washed twice with 70% ethanol, air-dried briefly, and re-suspended in 200 μl of TE buffer (10 mM Tris, 1.0 mM EDTA, pH 8.0) and treated with 10 μg ml

2 of RNAs for 1 h at 37°C.
DNA concentration as determined by electrophoresis in a 0.8% agarose gel with lambda DNA standard.

AFLP Analysis

The AFLP technique has been described by Zabeau and Vos (1993; Vos et al., 1995). In brief, approximately 500 ng of DNA was digested with EcoRI and MseI at 37°C for 3 h, and the enzymes were inactivated at 65°C for 15 min. The restricted DNA fragments were ligated to EcoRI (2 μl, 5 pmol μl⁻¹) and MseI (2 μl, 50 pmol μl⁻¹) adapters by adding 1 μl of T4 Ligase and 5 μl of adapter ligation buffer to the reaction tubes and incubating at 4°C for 12 h, and then the enzymes were inactivated at 65°C for 15 min. Two microliters of the DNA solution were pre-amplified using 0.6 μl (50 ng μl⁻¹) EcoRI + GACTGCCTACCAAATCA and 0.6 μl (50 ng μl⁻¹) MseI + GATGAGTCTGAGTAAA primers. Each reaction was composed of 20 μl of pre-amp primer mix, 2 μl of 10 reaction buffer, 1.6 μl of 25 mM dNTP, and 0.2 μl of Taq polymerase. PCR was performed for 29 cycles at: 94°C for 30 s, 56°C for 30 s, and 72°C for 60 s. The pre-amplified DNA was diluted to 1:20 with ultra-pure water. For selective amplification, DNA was digested with EcoRI and MseI at 37°C for 60 s. Polyacrylamide gel-electrophoresis was carried out according to the method described by Debasis Pattanayak et al. (2000) and silver nitrate staining used to detect the enzymes were inactivated at 65°C.

Univariate and Multivariate Statistical Analysis

The AFLP marker and metabolic matrices were both analyzed by hierarchical cluster analysis with squared Euclidean distance as similarity (SPSS 16.0, SPSS Inc.).

The metabolic matrix was analyzed further by both univariate and multivariate statistical techniques, attempting not only to discriminate different species, growing locations, and genotypes, but also to evaluate the predictive power of each metabolite and to select the potential differential metabolites.

Non-parametric Mann-Whitney test and receiver operating characteristic (ROC) curve (SPSS16.0, SPSS Inc.) were selected as the univariate methods conducted for differential metabolites evaluation and selection. The smaller P-value of the Mann-Whitney test indicates the larger significance of the corresponding metabolite. The critical P-value was set at 0.001. The area under the ROC curve (denoted as AUC hereafter) was computed via numerical integration of the curves. The metabolite that has the larger area under the ROC curve was identified as having the stronger separation and predictive power. The cut point of AUC was set at 0.8.

In parallel, for multivariate statistical analysis, unsupervised principle component analysis (PCA) and supervised orthogonal partial least squares discriminate analysis (OPLS–DA) analysis were carried out by the SIMCA-P 12.0 Software package (Umetrics, Umeå, Sweden). The matrix was mean centered and unit variance scaled prior to modeling. The default seven-round cross-validation was applied with one-seventh of the samples being excluded from model building (as testing set) in each round, hoping to guard against over-fitting. R²X and R²Y represent the fraction of the variance of the x and y variable explained by a model, while Q²Y suggests the predictive performance of the model. The cumulative values of R²X, R²Y, and Q²Y vary from 0 to 1. Based on the variable importance (VIP) values (with the threshold of 1.8) from a OPLS–DA model, a number of metabolites responsible for the differentiation of the metabolic profiles of species could be obtained.

The correlations between the metabolic peaks and AFLP marks were analyzed by using partial correlation analysis (in program R 2.10.1). The correlation networks were presented by Cytoscape software (Cytoscape 2.6.3) (Smoot et al., 2011).

SUPPLEMENTARY DATA

Supplementary Data are available at Molecular Plant Online.
**FUNDING**

This work was financially supported by the National Key Technology R&D Program in the 11th Five Years Plan of China (No 2006BA108805-03) and the National Basic Research Program of China (2007CB108800).

No conflict of interest declared.

**REFERENCES**


