Use of the Metabolomics Approach to Characterize Chinese Medicinal Material Huangqi

Li-Xin Duan[a,2], Tian-Lu Chen[b,2], Ming Li[a], Min Chen[c], Ye-Qing Zhou[a], Guang-Hong Cui[c], Ai-Hua Zhao[d], Wei Jia[b], Lu-Qi Huang[c,1] and Xiaoquan Qi[a,1]

a The Key Laboratory of Plant Molecular Physiology, Institute of Botany, Chinese Academy of Sciences, 20 Nanxincun Xiangshan Road, Beijing 100093, China
b Shanghai Center for Systems Biomedicine (Ministry of Education Key Laboratory of Systems Biomedicine), Shanghai Jiao Tong University, 800 Dongchuan Road, Shanghai 200240, China
c Institute of Chinese Materia Medica, Chinese Academy of Chinese Medical Sciences, 16 Nanxiao Road, Beijing 100700, China
d 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 became 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-requirement 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., 1998a), and mapping loci for quantitative traits (Qi et al., 1998b, 1999, 2000).

Apart from genetic differences at species, variety, and genotype levels, environments including growing locations and conditions (e.g. cultivated and wild) directly affect the levels of active components, and pharmacological and clinical effects (Obradovic 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 metabonomics, 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.

Table 1. The Collected Samples of Mojia Huangqi and Menggu Huangqi.

Index	Species	Growing locations	Genotypes
A*	Mojia Huangqi	Jilin Sipin (JL)	Cultivated (C)
B*	Mojia Huangqi	Jilin Tonghua (JL)	Cultivated (C)
C△	Mojia Huangqi	Gansu Weiyuan (GS)	Wild (W)
D△	Mojia Huangqi	Gansu Zhangxian (GS)	Wild (W)
E△	Menggu Huangqi	Gansu Longxi (GS)	Cultivated (C)
F◇	Menggu Huangqi	Shanxi Hunyuan (SX)	Cultivated (C)
G△	Menggu Huangqi	Gansu Zhuangxian (GS)	Cultivated (C)
H◇	Menggu Huangqi	Shanxi Yingxian (SX)	Wild (W)

* The north east of China; △ the north west of China; ◇ the middle of north China.

RESULTS AND DISCUSSION

Species Discrimination by Use of DNA and Metabolic Fingerprintings

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
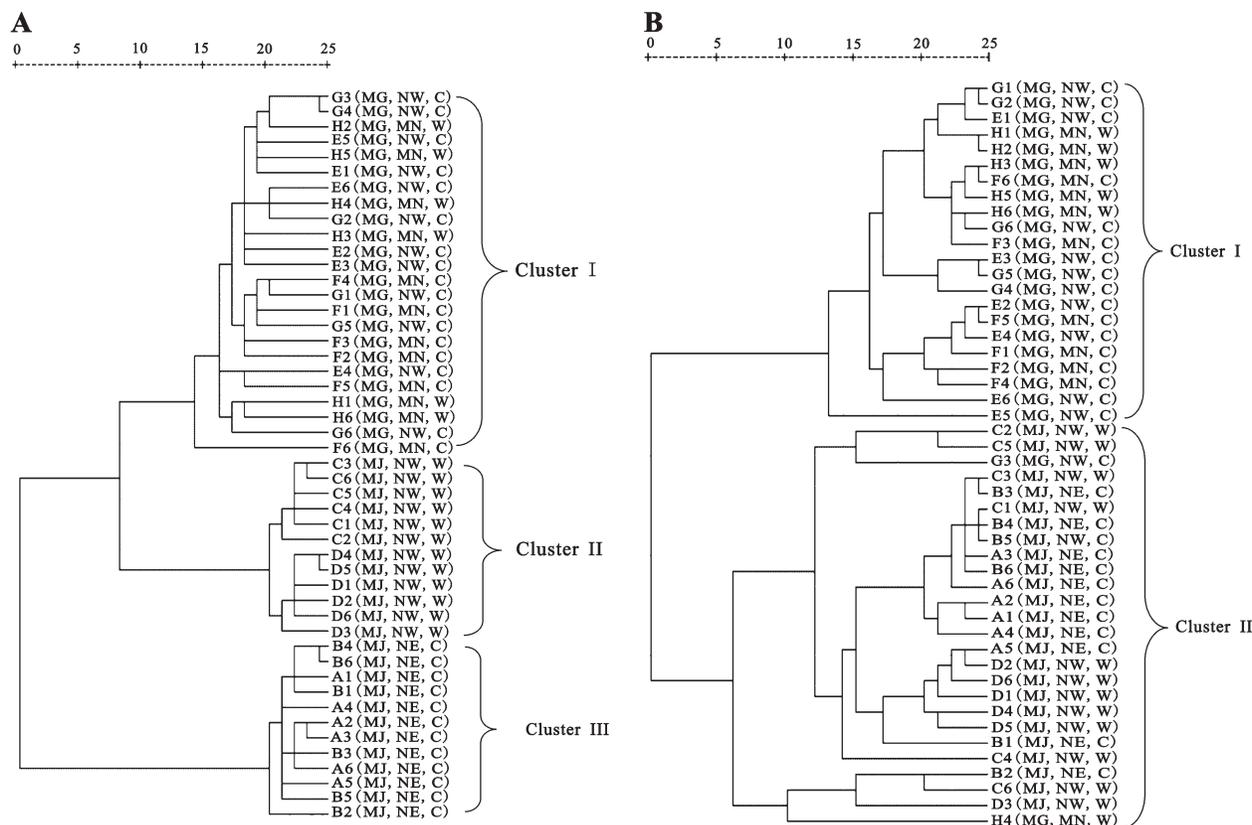
Figure 1. Dendrogram of Genetic and Metabolic Fingerprintings.
(A) Genetic similarity of Mojia Huangqi and Menggu Huangqi evaluated based on AFLP analysis using the between-groups linkage method with squared Euclidean distance as the interval program of SPSS.
(B) Dendrogram obtained after hierarchical cluster analysis (HCA) of the metabolic fingerpringters of Mojia Huangqi and Menggu Huangqi. The distances between these populations were calculated as described in Methods, using the normalized data of the single measurements. The letters of A, B, . . ., H were the indexes of different samples and the numbers 1, 2, . . ., 6 meant the repeats of the same sample. MJ and MG represent the species of *A. membranaceus* (**M**o**j**ia Huangqi) and *A. membranaceus* var. *mongholicus* (**M**eng**g**u Huangqi), respectively. NE, NW and MN were the abbreviations of growing locations and C and W were the abbreviations of cultivated and wild (see Table 1).

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 exist-

ing 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
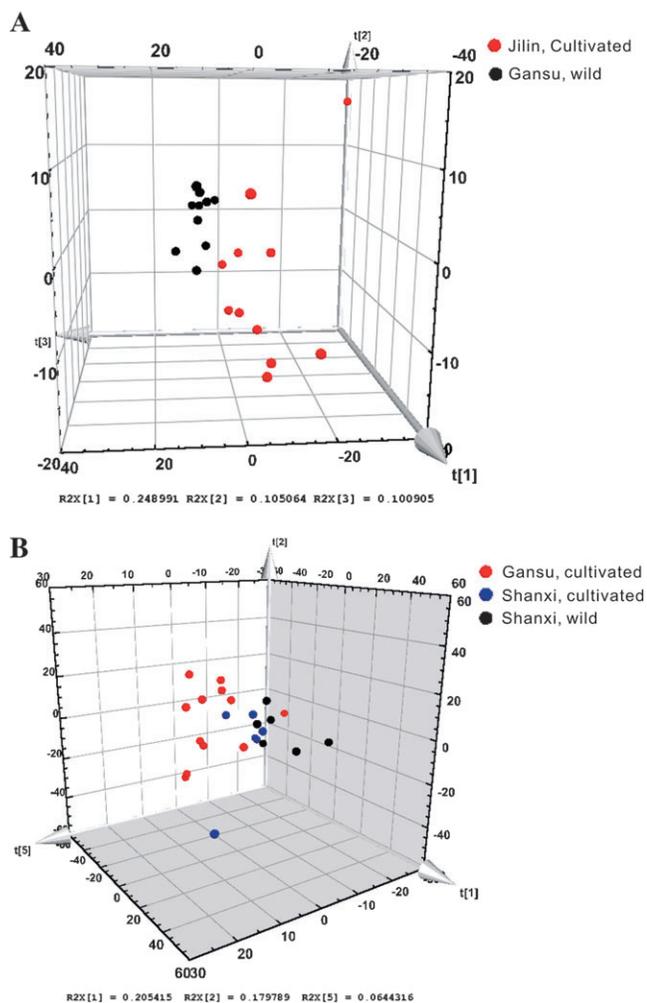
Figure 2. PCA Score Plot Marked by Different Growing Locations and Genotypes.

(A) PCA score plot of Mojia Huangqi. The red dot represents the collections cultivated in Jilin province and the black dot represents the collections that were wild and grow in Gansu province.

(B) PCA plot of Menggu Huangqi. The red dot represents the collections cultivated in Gansu province and the blue dot represents the collections cultivated in Shanxi province. The black dot represents the collections that were wild and grow in Shanxi province.

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-sensitive to environmental factors and growth stages in comparison to AFLP markers. In this study, six biological repeats were collected 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 metabolomics 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
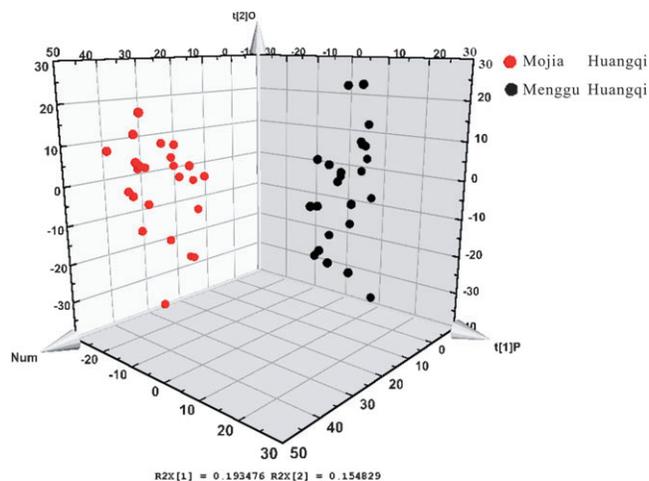
Figure 3. Score Plot of OPLS–DA Analysis for Mojia Huangqi and Menggu Huangqi Collections.

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 province (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 component (PC 5), except two individual collections (Figure 2B). However 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, microbial invasion, etc. affect the levels of metabolites. Previous research 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

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

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

* *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
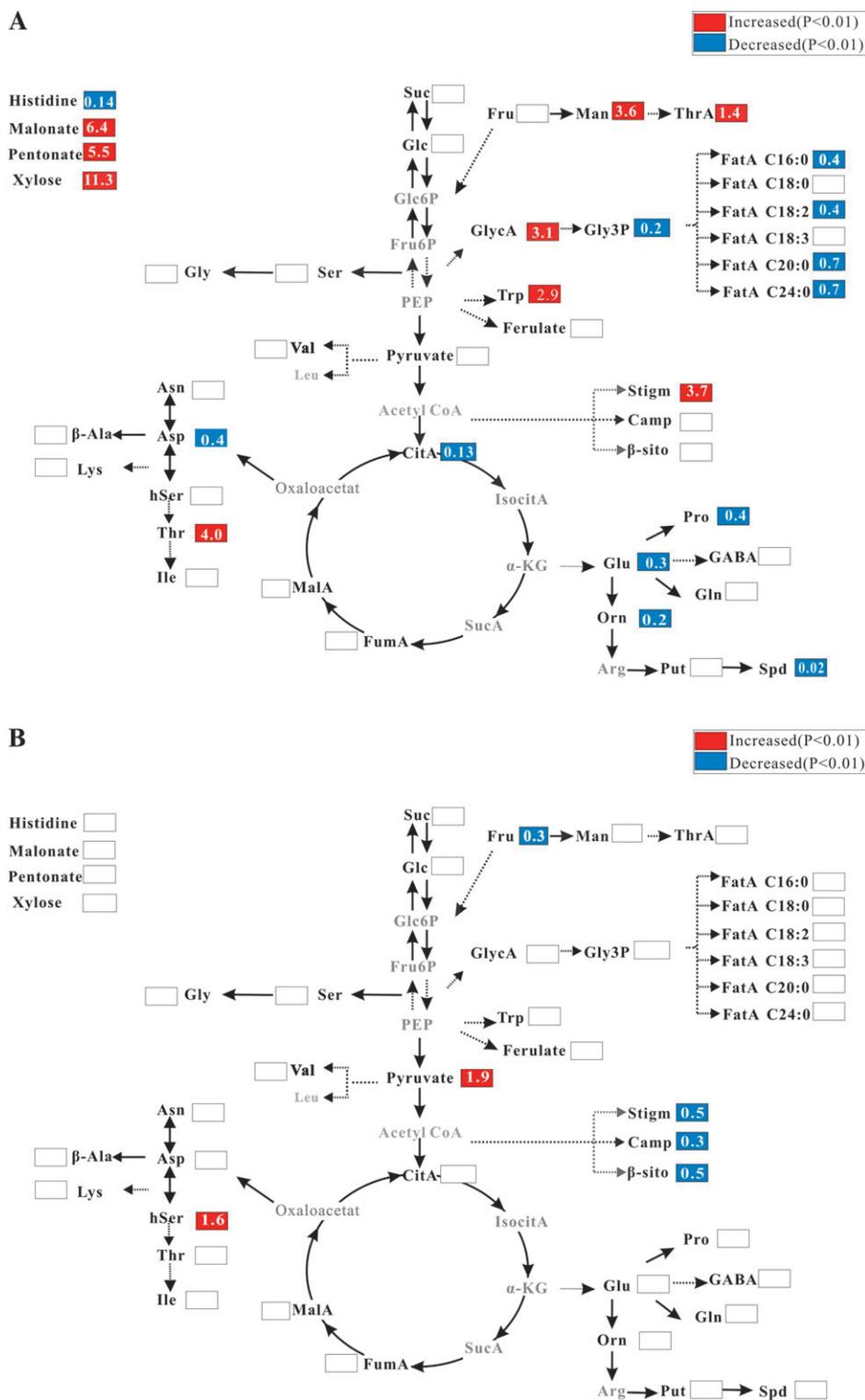
Figure 4. Levels of Metabolites in Main Metabolic Maps.

(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 $P < 0.01$. α-KG, alpha-ketoglutarate; Arg, arginine; Asn, asparagine; Asp, aspartate; β-Ala, beta-alanine; β-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 chro-

mosome 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. 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; Gln, glutamine; Glu, glutamate; Gly, glycine; Gly3P, glycerol-3-phosphate; GlycA, glycerate; hSer, homoserine; Ile, isoleucine; IsocitA, isocitric acid; Leu, leucine; Lys, lysine; MalA, malate; Man, mannose; Orn, Ornithine; PEP, Phosyphoenol-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), methoxy amination 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^{-1}) and 100 μl ribitol (0.2 mg ml^{-1}) as internal quantitative standards were added to an Eppendorf tube containing 100 mg of the powder. The mixture was extracted using a supersonic instrument for 30 min and centrifuged at 11 000 rpm for 10 min. One milliliter of the supernatant was transferred to another tube, mixed with

300 μl chloroform, and 600 μl dH$_2$O, 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^{-1}) in anhydrous pyridine at 37°C for 2 h, and then were silylated at 37°C for 30 min with 70 μl of N-methyl-N-(trimethylsilyl)-trifluoro acetamide (MSTFA). The derivatized samples were transferred into 250 μl glass vials (Agilent) for GC–TOF/MS analysis.

GC–TOF/MS

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^{-1}, the column temperature was held at 80°C for 4 min, then increased 5°C min^{-1} to 330°C, and held there for 5 min. The column effluent was introduced into the ion source of a Pegasus IV time-of-fight mass spectrometer. The transfer line and the ion source temperatures were 250 and 200°C, respectively. Ions were generated by a 70-eV electron beam at an ionization current of 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 Hompson, 1980; Pattanayak et al., 2000) with a few modifications. The root powder (100 mg) was dispersed in 1 ml preheated 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 saturant hydroxybenzene/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/isoamylalcohol (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^{-1} of RNAs for 1 h at 37°C.

DNA concentration as etermined 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 *Eco*RI and *Mse*I at 37°C for 3 h, and the enzymes were inactivated at 65°C for 15 min. The restricted DNA fragments were ligated to *Eco*RI (2 μl, 5 pmol μl^{-1}) and *Mse*I (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}) *Eco*RI + GACTGCGTACCAATTCA and 0.6 μl (50 ng μl^{-1}) *Mse*I + GATGAGTCCTGAGTAAA 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, 5 μl of the diluted product was added in a reaction tube containing 0.8 μl of *Eco*RI primer, 0.8 μl of *Mse*I primer, 9.4 μl of ddH$_2$O, 1.6 μl of 2.5 mM dNTP, 2.0 μl of 10 reaction buffer, and 0.2 μl of Taq polymerase. Eleven primer pairs (*Eco*RI-ACG/*Mse*I-AAC, E37M32; *Eco*RI-AGG/*Mse*I-AAC, E41M32; *Eco*RI-ACG/*Mse*I-AAG, E37M33; *Eco*RI-ACT/*Mse*I-AAG, E38M33; *Eco*RI-AGA/*Mse*I-AAG, E39M33; *Eco*RI-AGG/*Mse*EI-AAG, E41M33; *Eco*RI-AAG/*Mse*I-ACT, E33M38; *Eco*RI-ACA/*Mse*I-ACT, E35M38; *Eco*RI-ACG/*Mse*I-AGC, E37M40; *Eco*RI-ACT/*Mse*I-AGC, E38M40; *Eco*RI-AGG/*Mse*I-AGC, E41M40) were used for the selective amplification. Reactions were heated at 94°C for 2 min, and amplified by 33 cycles at: 94°C for 30 s, 56°C for 30 s, and 72°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 display the amplified DNA fragments as described in the reference (Ude et al., 2002).

Data Procession

Polymorphic AFLP bands were scored visually from the gels with the aid of a light box. A band was considered polymorphic if it was present in at least one genotype and absent in the others. A matrix (referred to as the gene matrix hereafter) was generated in which each band was scored as '1' if present and '0' if absent.

The acquired MS files from GC–TOFMS analysis were exported in NetCDF format by Chromatof software (v3.30, Leco Co., CA, USA). NetCDF files were extracted using custom scripts (revised Matlab toolbox hierarchical multivariate curve resolution (H-MCR), developed by Par Jonsson (Jonsson et al., 2004, 2005) in the Matlab7.0 (The MathWorks, Inc., USA) for data pretreatment procedures such as baseline correction, de-noising, smoothing, alignment, time-window splitting, and multivariate curve resolution (based on multivariate curve resolution algorithm) (Ni et al., 2008). The resulting three dimension matrix (referred to as the metabolic matrix hereafter) includes sample information, peak retention time, and peak intensities. Internal standards and any known artificial peaks, such as peaks caused by noise, column bleed, and derivatization procedure, were removed from the matrix. The resultant data were normalized to the maximum chromatogram intensity for each sample in order to minimize the discrepancy resulting from different volumes of samples.

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 sophisticated 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. R^2X and R^2Y represent the fraction of the variance of the *x* and *y* variable explained by a model, while Q^2Y suggests the predictive performance of the model. The cumulative values of R^2X, R^2Y, and Q^2Y 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.*

REFERENCES

Blaise, B.J., Giacomotto, J., Elena, B., Dumas, M.E., Toulhoat, P., Segalat, L., and Emsley, L. (2007). Metabotyping of *Caenorhabditis elegans* reveals latent phenotypes. Proc. Natl Acad. Sci. U S A. **104,** 19808–19812.

Bylesjö, M., Rantalainen, M., Cloarec, O., Nicholson, J.K., Holmes, E., and Trygg, J. (2006). OPLS discriminant analysis: combining the strengths of PLS-DA and SIMCA classification. J. Chemom. **20,** 341–351.

Catchpole, G.S., et al. (2005). Hierarchical metabolomics demonstrates substantial compositional similarity between genetically modified and conventional potato crops. Proc. Natl Acad. Sci. U S A. **102,** 14458–14462.

Cho, W.C., and Leung, K.N. (2007). *In vitro* and *in vivo* immunomodulating and immunorestorative effects of *Astragalus membranaceus*. J. Ethnopharmacol. **113,** 132–141.

Choi, H.K., Choi, Y.H., Verberne, M., Lefeber, A.W.M., Erkelensc, C., and Verpoorte, R. (2004). Metabolic fingerprinting of wild type and transgenic tobacco plants by ^{1}H NMR and multivariate analysis technique. Phytochemistry. **65,** 857–866.

Cook, D., Fowler, S., Fiehn, O., and Thomashow, M.F. (2004). A prominent role for the CBF cold response pathway in configuring the low-temperature metabolomem of *Arabidopsis*. Proc. Natl Acad. Sci. U S A. **101,** 15243–15248.

Cui, R., He, J.C., Wang, B., Zhang, F., Chen, G.Y., Yin, S., and Shen, H. (2003). Suppressive effect of *Astragalus membranaceus* Bunge on chemical hepatocarcinogenesis in rats. Cancer Chemoth. Pharm. **51,** 75–80.

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

Fiehn, O., Kopka, J., Dörmann, P., Altmann, T., Trethewey, R.N., and Willmitzer, L. (2000). Metabolite profiling for plant functional genomics. Nat. Biotechnol. **18,** 1157–1164.

Fiehn, O., Wohlgemuth, G., Scholz, M., Kind, T., Lee, D.Y., Lu, Y., Moon, S., and Nikolau, B. (2008). Quality control for plant metabolomics: reporting MSI-compliant studies. Plant J. **53,** 691–704.

Fu, K.-tsun (1993). Flora of *Reipublicae Popularis Sinicae*, Vol. 42 (1) (Beijing: Science Press).

Jonsson, P., Gullberg, J., Nordstrom, A., Kusano, M., Kowalczyk, M., Sjostrom, M., and Moritz, T. (2004). A strategy for identifying differences in large series of metabolomic samples analyzed by GC/MS. Anal. Chem. **76,** 1738–1745.

Jonsson, P., Johansson, A.I., Gullberg, J., Trygg, J.A.J., Grung, B., Marklund, S., Sjöström, M., Antti, H., and Moritz, T. (2005). High-throughput data analysis for detecting and identifying differences between samples in GC/MS-based metabolomic analyses. Anal. Chem. **77,** 5635–5642.

Kopka, J., et al. (2005). GMD@CSB.DB: the Golm Metabolome Database. Bioinformatics. **21,** 1635–1638.

Kuo, Y.H., Tsai, W.J., Loke, S.H., Wu, T.S., and Chiou, W.F. (2009). *Astragalus membranaceus* flavonoids (AMF) ameliorate chronic fatigue syndrome induced by food intake restriction plus forced swimming. J. Ethnopharmacol. **122,** 28–34.

Kusano, M., Fukushima, A., Arita, M., Jonsson, P., Moritz, T., Kobayashi, M., Hayashi, N., Tohge, T., and Saito, K. (2007). Unbiased characterization of genotype-dependent metabolic regulations by metabolomic approach in *Arabidopsis thaliana*. BMC Syst. Biol. **1,** 53.

Kusano, T., Berberich, T., Tateda, C., and Takahashi, Y. (2008). Polyamines: essential factors for growth and survival. Planta. **228,** 367–381.

Lee, D.Y., and Fiehn, O. (2008). High quality metabolomic data for *Chlamydomonas reinhardtii*. Plant Methods. **4,** 1–13.

Lisec, J., Schauer, N., Kopka, J., Willmitzer, L., and Fernie, A.R. (2006). Gas chromatography mass spectrometry-based metabolite profiling in plants. Nat. Protoc. **1,** 387–398.

Ma, X.F., Tu, P.F., Chen, Y.J., Zhang, T.Y., Wei, Y., and Ito, Y. (2004). Preparative isolation and purification of isoflavan and pterocarpan glycosides from *Astragalus membranaceus* Bge. var. *mongholicus* (Bge.) Hsiao by high-speed counter-current chromatography. J. Chromatogr. A. **1023,** 311–317.

Ma, X.Q., Shi, Q., Duan, J.A., Dong, T.T., and Tsim, K.W. (2002). Chemical analysis of Radix Astragali (Huangqi) in China: a comparison with its adulterants and seasonal variations. J. Agric. Food Chem. **50,** 4861–4866.

Muminovic, J., Melchinger, A.E., and Luubberstedt, T. (2004). Genetic diversity in cornsalad (*Valerianella locusta*) and related species as determined by AFLP markers. Plant Breed. **123,** 460–468.

Murray, M.G., and Hompson, W.F. (1980). Rapid isolation of high molecular weight plant DNA. Nucleic Acids Res. **8,** 4321–4325.

Ni, Y., Su, M., Lin, J., Wang, X., Qiu, Y., Zhao, A., Chen, T., and Jia, W. (2008). Metabolic profiling reveals disorder of amino acid metabolism in four brain regions from a rat model of chronic unpredictable mild stress. FEBS Lett. **582,** 2627–2636.

Obradovic, M., Kraljsek, S.S., Dermastia, M., and Kreft, S. (2007). A new method for the authentication of plant samples by analyzing fingerprint chromatograms. Phytochem. Anal. **18,** 123–132.

Pattanayak, D., Srinivasan, K., Mandaokar, A.D., Shukla, A., Bhalla, R., and Kumar, P.A. (2000). AFLP fingerprinting and genotypic characterization of some serovars of *Bacillus thuringiensis*. World J. Microb. Biot. **16,** 667–672.

Qi, X., and Lindhout, P. (1997). Development of AFLP markers in barley. Mol. Gen. Genet. **254,** 330–336.

Qi, X., Fufa, F., Sijtsma, D., Niks, R.E., Lindhout, P., and Stam, P. (2000). The evidence for abundance of QTLs for partial resistance to *Puccinia hordei* on the barley genome. Mol. Breeding. **6,** 1–9.

Qi, X., Jiang, G., Chen, W., Niks, R.E., Stam, P., and Lindhout, P. (1999). Isolate-specific QTLs for partial resistance to *Puccinia hordei* in barley. Theor. Appl. Genet. **99,** 877–884.

Qi, X., Niks, R.E., Stam, P., and Lindhout, P. (1998a). Identification of QTLs for partial resistance to leaf rust (*Puccinia hordei*) in barley. Theor. Appl. Genet. **96,** 1205–1215.

Qi, X., Stam, P., and Lindhout, P. (1998b). Use of locus-specific AFLP markers to construct a high-density molecular map in barley. Theor. Appl. Genet. **96,** 376–384.

Roessner, U., Luedemann, A., Brust, D., Fiehn, O., Linke, T., Willmitzer, L., and Fernie, A.R. (2001). Metabolic profiling allows comprehensive phenotyping of genetically or environmentally modified plant systems. Plant Cell. **13**, 11–29.

Schauer, N., et al. (2006). Comprehensive metabolic profiling and phenotyping of interspecific introgression lines for tomato improvement. Nat. Biotechnol. **24**, 447–454.

Schoen, D.J., and Brown, A.H.D. (1991). Intraspecific variation in population gene diversity and effective population size correlates with the mating system in plants. Proc. Natl Acad. Sci. U S A. **88**, 4494–4499.

Smoot, M., Ono, K., Ruscheinski, J., Wang, P.L., and Ideker, T. (2011). Cytoscape 2.8: new features for data integration and network visualization. Bioinformatics. **27**, 431–432.

Szabados, L., and Savouré, A. (2010). Proline: a multifunctional amino acid. Trends Plant Sci. **15**, 89–97.

Turpeinen, T., Vanhala, T., Nevo, E., and Nissila, E. (2003). AFLP genetic polymorphism in wild barley (*Hordeum spontaneum*) populations in Israel. Theor. Appl. Genet. **106**, 1333–1341.

Ude, G., Pillay, M., Nwakanma, D., and Tenkouano, A. (2002). Analysis of genetic diversity and sectional relationships in *Musa* using AFLP markers. Theor. Appl. Genet. **104**, 1239–1245.

Vos, P., et al. (1995). AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res. **23**, 4407–3314.

Wang, E.T., and Liu, M.Y. (1996). A comparative biology study on the two medicinal Astragelus. Bullet. Botan. Res. **16**, 85–91.

Wang, Z.Y., Second, G., and Tanksley, S.D. (1992). Polymorphism and phylogenetic- relationships among species in the genus Oryza as determined by analysis of nuclear RFLPs. Theor. Appl. Genet. **83**, 565–573.

Wanner, L.A., and Junttila, O. (1998). Cold-induced freezing tolerance in *Arabidopsis*. Plant Physiol. **120**, 391–399.

Xiao, P.G., Feng, M.X., Cheng, J.R., and Lou, Z.Q. (1964). Botanical and pharmacognostical studies of the Chinese Drug Huang-chi I. Identification of the botanical origins and review of the ancient herbs. Acta. Pharm. Sin. **11**, 114–119.

Xie, X.L., Wang, X.S., Zhao, L., Wang, L., and Li, Y. (2005). Review of studies on germplasm resources of Radix Astragali. J. Anhui Agri. Sci. **33**, 121–123.

Yip, P.Y., and Kwan, H.S. (2006). Molecular identification of *Astragalus membranaceus* at the species and locality levels. J. Ethnopharmacol. **106**, 222–229.

Zabeau, M., and Vos, P. (1993). Selective restriction fragment amplification: a general method for DNA fingerprinting. European Patent, EP0534858.