Published in final edited form as:

J Nat Prod. 2015 August 28; 78(8): 2007–2022. doi:10.1021/acs.jnatprod.5b00342.

Metabolite Profiling and Classification of DNA-Authenticated Licorice Botanicals

Charlotte Simmler^{*,†}, Jeffrey R. Anderson[†], Laura Gauthier[†], David C. Lankin[†], James B. McAlpine[†], Shao-Nong Chen[†], and Guido F. Pauli^{*,†}

[†]UIC/NIH Center for Botanical Dietary Supplements Research, Department of Medicinal Chemistry and Pharmacognosy, College of Pharmacy, University of Illinois at Chicago, 833 S. Wood Street, Chicago, Illinois, United States

[‡]Institute for Tuberculosis Research, College of Pharmacy, University of Illinois at Chicago, 833 S. Wood Street, Chicago, Illinois, United States

Abstract

Raw licorice roots represent heterogeneous materials obtained from mainly three *Glycyrrhiza* species. *G. glabra*, *G. uralensis*, and *G. inflata* exhibit marked metabolite differences in terms of flavanones (Fs), chalcones (Cs), and other phenolic constituents. The principal objective of this work was to develop complementary chemometric models for the metabolite profiling, classification, and quality control of authenticated licorice. A total of 51 commercial and macroscopically verified samples were DNA authenticated. Principal component analysis and canonical discriminant analysis were performed on ¹H NMR spectra and area under the curve values obtained from UHPLC-UV chromatograms, respectively. The developed chemometric models enable the identification and classification of *Glycyrrhiza* species according to their composition in major Fs, Cs, and species specific phenolic compounds. Further key outcomes demonstrated that DNA authentication combined with chemometric analyses enabled the characterization of mixtures, hybrids, and species outliers. This study provides a new foundation for the botanical and chemical authentication, classification, and metabolomic characterization of crude licorice botanicals and derived materials. Collectively, the proposed methods offer a comprehensive approach for the quality control of licorice as one of the most widely used botanical dietary supplements.

*Corresponding Authors: (G. F. Pauli): gfp@uic.edu. Tel: +1 (312) 355-1949. Fax: +1 (312) 355-2693. (C. Simmler): simmler@uic.edu.

Notes

The authors declare no competing financial interest.

Supporting Information

Microscopic analyses of *Glycyrrhiza* powder; ¹H NMR spectra subjected to PCA; comparative PCA results including the statistical properties; UHPLC-UV chromatograms, calculation method for the determination of AUC ratios, and the resulting workbook table; SIMCA properties and calculated distances for the classification of mixtures, hybrids, and outliers; as well as ¹H NMR data of characteristic licorice compounds. The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jnatprod.5b00342.

In the United States herbal/dietary supplement market, consumers have presumably a 50% chance of choosing an authentic product containing both the correct species and characteristic bioactive compounds at their appropriate strength.^{1,2} Notably, the botanical authentication including the Latin binomial name is not always specified on the product, whether the botanical dietary supplement (BDS) is sold in the U.S., in the European Union, or in other market places. Traditionally, botanists rely upon morphological features to identify plants. However, the raw materials used to manufacture BDSs are typically ground to a powder, or otherwise processed, thereby hampering morphologic and macroscopic authentication. Investigators wishing to botanically identify these plant-derived materials, prior to any chemical or biological investigation, must turn to other intrinsic properties such as DNA barcoding, underlining genetic differences between species.^{3,4} DNA-based assays are not biased by the growth environment or the organ sampled and have become central to the identification of all types of biological samples including plant material.^{5,6} DNA barcoding encompasses the technique of identifying biological specimens using short DNA sequences from a standardized region of the nuclear and/or organelle genomes.^{4,7} On the other hand, metabolomic techniques for fingerprinting and profiling employing near-infrared (NIR) and Fourier transform infrared (FT-IR) spectroscopy, as well as chromatographic systems hyphenated with various detection modes (LC-UV, GC/LC-MS), and more recently NMR spectroscopy, have been successfully used for the chemical analysis and characterization of plant organs, as part of the quality control (QC) of medicinal plant products.^{3,8–10}

The global economic value of licorice (*Glycyrrhiza* sp., Leguminosae) is unquestionable, as evidenced by its multiple uses as a pharmaceutical agent, BDS, cosmetic, sweetener, food additive, flavor additive for tobacco, and confectionery food.^{11–13} BDSs containing licorice roots, rhizomes, and stolons are sourced globally from mainly three different species: *Glycyrrhiza glabra* L., *G. uralensis* Fisch. ex DC., and *G. inflata* Batalin. These species are listed in the pharmacopeias worldwide and have been employed in traditional remedies since the beginning of medicine.¹² In the Chinese Gan-Cao¹⁴ and the Japanese Kampo medicines,^{15,16} the three *Glycyrrhiza* species are even considered as being equivalent and, thus, are combined and utilized without distinction. Similar concepts apply to other pharmacopeias.^{17,18}

The traditional botanical authentication of commercial licorice is complex, as root powders from all three major *Glycyrrhiza* species share the same microscopic characteristics

(Supporting Information S1).¹⁷ Moreover, the heterogeneity of the plant material, which may be due to contamination or purposeful combination with other *Glycyrrhiza* species, can depreciate the QC outcomes and impede all efforts aimed at their chemical standardization. Major secondary metabolites occurring in the roots and stolons of all three species are the glycosides of liquiritigenin (flavanones: **1-4F**) and isoliquiritigenin (2'-hydroxychalcones: **1-4C**), as well as glycyrrhizin (**5**), an oleanane-type triterpenoid saponin responsible for the sweet taste of licorice (Scheme 1). Prior research has shown that the roots of each of the three *Glycyrrhiza* species display unique chemical profiles, determined by the presence of species-specific metabolites (glabridin (**6**) for *G. glabra*, licochalcone A (**7**) for *G. inflata*, and glycy coumarin (**9**) for *G. uralensis*) and by the relative composition of major flavanone (F), chalcone (C), and triterpenoid compounds.^{16,19–24} Metabolomic analyses of licorice roots were initiated by Yang and co-workers,²² who performed ¹H NMR multivariate analysis on botanically and macroscopically identified *Glycyrrhiza* species focusing on primary metabolites.²² Subsequently, Farag and co-workers²³ combined LC-MS and ¹H NMR metabolite profiling of botanically and macroscopically identified *Glycyrrhiza* species, using three sample acquisitions per species. Recently, we have demonstrated that both major Fs and their C isomers coexist in extracts from licorice roots and can be determined as pairs of isomers.²⁰ In licorice, those major Fs and Cs are known to be biosynthetically related and chemically interchangeable, exemplifying the notion of dynamic residual complexity.^{25,26} Accordingly, each C in licorice extract has a corresponding F isomer, which together form an F/C pair (Scheme 1). Despite the widespread use of licorice, the levels of most of its metabolites, including the F/C pairs and the species-specific compounds, are generally not disclosed in commercial BDSs. In fact, the QC of pharmacopeial licorice relies mainly on the detection and quantitation of glycyrrhizin (**5**), a major constituent that is present in all three *Glycyrrhiza* species and, thus, is the least (species) specific marker.^{17,18} To improve this situation, modern and reliable QC methodologies that are more comprehensive and permit the simultaneous authentication, differentiation/classification, and chemical characterization of the pharmacopeial *Glycyrrhiza* species are needed.

The principal objective of this study was to perform a classification and metabolomic characterization of all three pharmacopeial *Glycyrrhiza* species used as “licorice” in BDSs. A total of 51 samples were acquired as root sticks, bulk powders, or capsules. All samples were authenticated by DNA barcoding, adjusting the method originally developed by Kondo and coworkers.²⁷ Two orthogonal metabolomic techniques, untargeted 1D ¹H NMR and targeted UHPLC-UV,²⁰ were performed on crude extracts of each sample. The data obtained from these two analytical platforms were subjected to multivariate data analyses. All NMR spectra were analyzed through unsupervised principal component analysis (PCA) and supervised soft independent modeling of class analogy (SIMCA). Supervised canonical discriminant analysis (CDA) was employed to develop a classification model based on selected area under the curve (AUC) values from UHPLC-UV chromatograms, integrating the parameters linked to the F/C composition of each species. The second major objective was to demonstrate that DNA authentication is fundamental for the meaningful construction of chemometric models PCA/SIMCA and CDA, which collectively will enable the classification and characterization of various licorice samples sold as BDSs. Finally, our

study proposes a comprehensive approach to the QC of licorice botanicals, combining botanical authentication with DNA analysis together with metabolomic and chemometric classification.

With regard to nomenclature, the contemporary literature uses the terms “*metabolite profiling*” and “*metabolite/metabolic fingerprinting*” widely and interchangeably.⁸ Both terms are used to describe more comprehensive studies of the (plant) metabolome.^{3,8} Considering that “*metabolite profiling*” is predominantly associated with the identification of key metabolites that lead to the differentiation of samples, its usage is preferred in this study.

RESULTS AND DISCUSSION

Authentication of Licorice Roots by DNA Barcoding

Commercially available licorice materials can be composed of a single *Glycyrrhiza* species, *Glycyrrhiza* hybrids, or a mixture of *Glycyrrhiza* species, but also mixtures of *Glycyrrhiza* species with other plants or contaminants. Over the past two decades, the international initiative, Consortium for the Barcode of Life (CBOL, <http://barcoding.si.edu>), has developed aims in global standards for DNA barcoding leading to species identification. With regard to plant barcoding, CBOL has sanctioned the use of two protein coding regions, *matK* and *rbcl*, as core DNA barcodes and two noncoding regions, *nrITS* and *psbA-trnH*, as supplemental markers.²⁸ The *matK*, *rbcL*, and *psbA-trnH* loci are found in the chloroplast (cp) genome (cpDNA), while the *nrITS* locus is part of the nuclear DNA (nrDNA). The generation of DNA barcodes can generally be achieved from fresh and dried plant parts. Under certain circumstances,^{29–32} plant DNA can also be isolated from plant extracts, processed herbal drugs, and finished products such as herbal teas, tablets, and capsules, provided that industrial processing did not lead to extensive DNA degradation.

Multiloci DNA characterization of the three principal botanically identified *Glycyrrhiza* species (205 analyzed leaf samples) was initially developed by Kondo and co-workers in 2007.²⁷ According to the molecular phylogeny set forth in this work, each *Glycyrrhiza* species or hybrid is characterized by different genotypes (Figure 1). The combination of *nrITS*, *rbcL*, *matK*, and *psbA-trnH* markers from each sample allows for the definition of different total genotypes (TGs), leading to the identification of each species or hybrid.²⁷ Following this framework, *G. uralensis* is defined by four total genotypes (TG6–9), *G. glabra* has a total of two genotypes (TG2, TG3), and *G. inflata* comprises three genotypes (TG2, TG4, TG5). As the present study was dealing with partially processed raw material of below-ground plant parts, such as ground roots, root powders, and mixtures of root powders, some modifications were made to Kondo's original protocol.²⁷ Specifically, the *nrITS* “universal” primers designed by White and co-workers³³ and employed by Kondo and co-workers²⁷ were initially used for amplification. Owing to their universality, these *nrITS* primers can also amplify sequences belonging to other taxa, thus leading to interference, notably with the amplification of fungal DNA.³⁴ To avoid such cross contaminations, *Glycyrrhiza*-specific internal *nrITS* primers were designed (see Experimental Section). Moreover, a second set of primers, which were 100% conserved between the three

Glycyrrhiza species, were developed for commercial samples with degraded DNA, allowing the production of smaller amplicons (~200 bp).

Glycyrrhiza hybrids were identified by comparing the nucleotide sequence chromatograms of the *nrITS* locus (Figure 1C). Localized in the nucleus, this DNA marker is inherited from both parents. The occurrence of both I-2 and I-3 sequences for the *nrITS* has been called ADD genotype by Kondo and corresponds to “nucleotide additivities” often observed in hybrids between two species.²⁷ In such a case, the *nrITS* DNA sequencing traces will show secondary peaks only at the variable species-specific nucleotides in both sequencing directions. For pure *Glycyrrhiza* hybrids, bidirectional DNA sequencing traces from cp loci will always show single peaks due to the chloroplast’s haploid genome (Figure 1C). In fact, it is generally accepted that angiosperm plastids are inherited maternally, although numerous reports have underlined that plastid inheritance pattern can be influenced by both maternal and paternal genotypes.³⁵ This biparental inheritance has not been demonstrated yet for *Glycyrrhiza* species; thus, we conclude presently that the cpDNA sequences of hybrids should identify the maternal species (Table 3 in bold font). Specifically, Kondo and co-workers²⁷ reported that the cpDNA markers R-2 and T-4 were located only in *G. uralensis*, whereas the markers M-2 and T-3 were found in *G. inflata*. This information was considered to specify the maternal species of hybrids.

In light of the undifferentiated pharmacopeial definitions of licorice, commercial BDSs are likely to represent mixtures of the three *Glycyrrhiza* species or even contain other plant species as adulterants. In order to evaluate the quality of the authentication and classification methods in the present study, in-house mixtures were prepared from powders of genetically uniform DNA-authenticated species. All mixtures were characterized by the proportional quantity (weight %) of each authenticated species (Table 4). The mixtures were DNA identified through the observations of secondary peaks on the nucleotide sequence chromatogram at variable species-specific nucleotides for all four DNA markers (Figure 1C). When mixing occurs between samples that differ at the species-specific nucleotides of chloroplast loci, the DNA sequencing traces for those nucleotides will show secondary peaks to some degree. Secondary peak height ratio, provided by Sequencher v5.2, was used to determine primary (1°) and secondary (2°) genotypes in both hybrid and mixture cases (Tables 3 and 4).

Considering that the sample requirements for DNA extraction are rather small, the variability of DNA barcoding results is inevitably linked to the homogeneity of the powder mixtures. Thus, sampling representativeness has to be considered as a key factor when analyzing mixtures. Results obtained from the typical 50–100 mg of powdered sample might not represent an accurate composition of the one pound or larger batch, from which the sample was taken. In light of this variability, all DNA extractions from the prepared mixtures, amplifications, and sequencing were performed several times in order to avoid misinterpretation. With the genotype combination, and knowing that one species is characterized by at least two genotypes, it was possible to detect and identify *Glycyrrhiza* species, even when present at only ~1% w/w in the mixture.

Among the 51 acquired licorice samples, 11 (~21%) turned out to be misidentified *Glycyrrhiza* species; eight (~16%) were found to be mixtures of *Glycyrrhiza* species (Table 1); and the majority (37; 72%) were authenticated as being composed of one single *Glycyrrhiza* species (Table 2). The 11 misidentified samples comprised six hybrids (Tables 1 and 3) and three misidentified as *G. inflata* samples, which were authenticated as *G. glabra* (BC748), *G. uralensis* (BC629 and BC778), and one *Glycyrrhiza* mixture (BC736). Most of the mixtures were sold as BDS capsules or powders and were found in both U.S. and E.U. markets. Interestingly, their BDS label indicated either *G. glabra* or simply “licorice”, while lacking any species identification. Among the 37 samples containing single *Glycyrrhiza* species (Table 2), three were identified as being *G. inflata* (TG5), 13 belonged to the species *G. uralensis* (TG7–9), and 21 were *G. glabra* (TG2, 3). As Kondo and coworkers reported,²⁷ the TG4 and TG6 genotypes are rare variants and, accordingly, were not detected in the present study. Interestingly, among the seven claimed *G. inflata* samples, only three were identified as pure *G. inflata* (TG5 genotype), while the others were found to be *G. uralensis*, a plant cultivated in the same region in China, or *G. glabra*, known to be also cultivated in Afghanistan, but also *Glycyrrhiza* hybrids (Tables 1 and 3). Moreover, when sold as BDSs, *G. inflata* was mostly found in mixtures with other *Glycyrrhiza* species. These observations collectively suggested that *G. inflata* is not readily available as pure single species and its sourcing requires working closely with local growers. *G. inflata* grows naturally in the northwest of China in the Xinjiang Uygur Autonomous Region^{36–38} and in Afghanistan.²⁷ The relatively harsh natural geographical habitat, additionally subject to political tensions, combined with an increased likelihood of hybridization with *G. uralensis* might contribute to the difficulties in sourcing authentic *G. inflata*.

In summary, DNA authentication was employed as an essential qualitative step of botanical verification and yielded three groups of samples: single species, hybrids, and mixtures of *Glycyrrhiza* species. Notably, and by a significant margin, DNA-based species identification does not suffice for the assessment of botanical quality, especially because bioactive compounds are mostly small metabolites rather than DNA molecules. Moreover, a given genotype does not necessarily determine the phenotype and overall plant metabolome. In addition, *G. glabra* and *G. inflata* samples share the same TG2 genotype (Figure 1, Table 2). The authentication of *Glycyrrhiza* species belonging to the TG2 genotype alone will, therefore, require the implementation of orthogonal methods based on metabolite profiling. The present results demonstrated that an initial DNA authentication step was still crucial for the selection of 37 samples containing single *Glycyrrhiza* species in order to perform the following chemometric analyses.

Chemometric Classification of DNA-Authenticated *Glycyrrhiza* Species

The DNA-authenticated samples were extracted with 95% EtOH using ASE 350 instrumentation. This reproducible and standardized extraction protocol allowed for a rapid processing of several samples simultaneously, which is a fundamental requirement of any comparative metabolite profiling study. As suggested previously,²⁰ the use of EtOH as solvent yielded extracts enriched in phenolic metabolites and depleted in highly polar metabolites. Each licorice material was extracted twice, when possible, and each extract was

analyzed in duplicate, in order to evaluate both sample representativeness and reproducibility of the chemical profiles.

Two complementary analytical techniques were used for metabolite profiling: one spectroscopic, 1D ^1H NMR, and one chromatographic, UHPLC-UV. The NMR technique offers a holistic view of the extract composition, allowing the simultaneous detection of UV-visible and -transparent metabolites. To enhance reproducibility, all spectra were acquired under quantitative conditions (qHNMR).³⁹ The UHPLC-UV technique enabled a rather more targeted analysis of principal UV-visible metabolites such as the major F/C pairs (pairs **1–4**), glycyrrhizin (**5**), glabridin (**6**), licochalcone A (**7**), the recently identified *p*-hydroxybenzylmalonic acid HBMA (**8**), and glycy coumarin (**9**).⁴⁰ The major investigated F/C pairs were as follows: liquiritin (**1F**) and isoliquiritin (**1C**) form the first F/C pair (F/C pair **1**); F/C pair **2** is defined by liquiritigenin 7-*O*-apiosylglucoside (**2F**) and licuraside (**2C**); liquiritin apioside (**3F**) and isoliquiritin apioside (**3C**) form F/C pair **3**, and F/C pair **4** consists of liquiritigenin (**4F**) and isoliquiritigenin (**4C**). With regard to their analytical sensitivity,⁴¹ both ^1H NMR and UHPLC-UV techniques will principally detect the more concentrated compounds (low μM range). Both analyses characterize and classify *Glycyrrhiza* extracts according to their major to moderately abundant metabolites (Figure 2).

NMR-Based Metabolomic Analysis

Multivariate statistical analysis with PCA has been performed previously with either ^1H NMR spectra or MS data for the differentiation of botanically and macroscopically authenticated *Glycyrrhiza* species.^{22–24} However, as demonstrated above, macroscopic identification is insufficient to detect *Glycyrrhiza* hybrids. All licorice roots/rhizomes are similar macroscopically and microscopically. A potential botanical misidentification, in particular with *G. inflata* samples, could lead to misinterpretation or erroneous classification. Hence, our PCA analyses included only DNA-authenticated samples containing single *Glycyrrhiza* species. This statistical analysis was chosen for its capacity to group spectra according to their similarities without using the knowledge of sample class. Accordingly, PCA performed with ^1H NMR spectra generated an unbiased clustering among chemically similar samples and produced an untargeted classification of the plant extracts. The PCA was performed with the AMIX software, applying rectangular bucketing on processed qHNMR spectra, and a final Pareto scaling, giving weight to minor signals.⁴² The first PCA result was obtained considering the full NMR spectrum (Supporting Information S2) and excluding only the solvent signals (DMSO-*d*₅, HOD, EtOH residue), the sugar region, and the proline signals.⁴⁰ The second PCA used only the downfield region from 4.50 to 8.50 ppm, without the sucrose anomeric proton signal at 5.169 ppm (d, *J* = 3.75 Hz).⁴⁰ Both PCA models excluded the chemical fluctuation linked to the presence of sucrose and proline, as their concentrations are known to be related to the conditions of cultivation, drought, and time of harvest.^{43,44} Therefore, those metabolites were regarded as nonspecific for the distinction of *Glycyrrhiza* species. Interestingly, the score and loading plots obtained with both models (full spectrum vs downfield region) led to similar clustering results (Supporting Information S3–4), suggesting that the high-field NMR signals were less discriminant. Using the downfield region only, PC1 and PC2 together accounted for 58.2% of the metabolomic variance, compared to 47.2% when considering the full NMR spectrum.

In light of these observations, the PCA model associated with the NMR downfield region was selected. The score plot in Figure 3A shows three distinct clusters, corresponding to the three DNA-authenticated species, and distributed over three opposite regions. By comparing the NMR signals (variables as buckets) in the loading plots in Figure 3B with the results in the score plot, discriminating variables were identified as being responsible for the sample differentiation. As expected, the signal corresponding to the omnipresent licorice compound **5** (H-12, s, at 5.38 ppm) was not discriminant and did not lead to species differentiation. Without considering the anomeric protons of the glycosidic moiety, and due to extensive signal overlap, the differentiation of **1F** and **3F** in the crude extracts required attention to detail (Table 5, Figure 2). However, the H-3'/5' proton signals were slightly shifted, with the protons of resonating at 7.052 ppm vs those of **3F** at 7.028 ppm. The anomeric protons from the glycosidic moiety of the major F/C pairs, as well as the aromatic signals of **8** (buckets at 6.56 and 6.92 ppm), clearly contributed to the differentiation between *G. glabra* and *G. inflata* on one hand vs *G. uralensis* on the other. The anomeric glucose proton of F/C pair **1** was included in the bucket at 4.88 ppm (d, $J = 7.46$ Hz), whereas the anomeric proton from the apiose moiety of the F/C pair **3** was found at 5.32 ppm (d, $J = 1.20$ Hz). Regarding the identification of discriminative variables, this led to the conclusion that NMR signals corresponding to the F/C pair **1** and to compound **4F** (buckets at 5.48, 6.48, and 7.40 ppm) determined the group formed by *G. uralensis* extracts. The major variables enabling the distinction of *G. glabra* samples were signals from the apiose anomeric proton, the F/C pair **2** (bucket centered at 8.24 ppm of **2C**), the signal from **3F** (bucket centered at 7.00 ppm), and signals from glabridin (**6**) derivatives (Table 5, Figures 2 and 3). Finally, the distinct aromatic signals corresponding to **7** led to the clustering of *G. inflata* samples. As shown in the score plot, this species displayed a unique ^1H NMR chemical profile, forming a group independent from the close clustering observed between *G. glabra* and *G. uralensis*. Interestingly, both **6** and **7** were clearly detectable by NMR and UHPLC-UV, thereby suggesting their relative abundance in the corresponding *Glycyrrhiza* extracts (Figure 2). In contrast, the ^1H NMR conditions did not lead to the detection of the *G. uralensis*-specific compound, glycycomarin (**9**). This observation can be explained by two factors: the ^1H NMR spectrum of **9** does not offer characteristic isolated signals within the complex spectrum of *G. uralensis* extract, and **9** is present at a lower level compared to **6** and **7**.

Using PCA, the authenticated samples with TG2 genotype were chemically identified as *G. glabra*. Within the cluster formed by the *G. glabra* samples, three groups were observed. Samples containing proportionally more **3F**, while depleted in **6**, were closest to the *G. uralensis* group on the score plot. When indicated, these samples (BC695, BC726) originated from Egypt. *G. glabra* samples from Iran and Afghanistan clustered at the other extremity of this species group (Figure 3). Such observations were in agreement with the results obtained by Farag and co-workers, who also observed that *G. glabra* samples from Egypt clustered apart from those originating from Syria and Afghanistan.²³ Further analyses of the qHNMR spectra suggested that samples BC745, BC749, and BC757 were characterized by a higher proportion of the F/C pair **2**, with proportionally more Cs than Fs compared to other *G. glabra* samples (Supporting Information S5).

In conclusion, the PCA analysis performed with ^1H NMR data enabled the metabolomic differentiation/classification of DNA-authenticated *Glycyrrhiza* species according to their composition in major F/C pairs and the presence of the detectable species-specific metabolites **6** and **7**.

UHPLC-UV-Based Metabolomic Analysis

As a second approach, an orthogonal UHPLC-UV analysis was performed to determine whether all *Glycyrrhiza* extracts can be classified according to their composition in major F/C pairs, solely (Figure 4, Supporting Information S6), or whether their classification requires the additional consideration of species-specific metabolites. For this purpose, a methodology using only the AUC ratios of selected metabolites was developed. The AUC values for each F and for compound **6** were taken at 275 nm, whereas the AUCs for each C and for compounds **7** and **9** were measured at 360 nm. Each AUC value was divided by the total calculated AUC at each wavelength (Supporting Information S7). The total F/C ratio was then determined by dividing the total $\text{AUC}_{275\text{ nm}}$ by the total $\text{AUC}_{360\text{ nm}}$. The results obtained for the 37 authenticated species demonstrated that they all displayed a unique F/C composition (Figure 4A). The F/C pairs **2** and **3** clearly characterized *G. glabra* extracts, whereas the F/C pairs **1** and **3** predominated in all the *G. uralensis* and *G. inflata* extracts. With regard to their major F/C pairs, *G. uralensis* extracts were similar to *G. inflata*. However, the total F/C ratio clearly indicated that *G. glabra* and *G. inflata* extracts have proportionally more C than F compared to *G. uralensis* samples (Figure 2, Figure 4A). Interestingly, the calculated ratios for samples originating from Iran and Afghanistan (BC745, BC749) revealed that the F/C pair **2** represented 50% of the total calculated F/C pairs, compared to the other *G. glabra* samples ($28 \pm 11\%$). Moreover the total F/C ratio determined for the species from Iran and Afghanistan was found to be the lowest (0.53 ± 0.03) of all *G. glabra* samples (0.74 ± 0.16). In conclusion, both the composition in major F/C pairs and the calculated F/C ratios can contribute to the authentication of *Glycyrrhiza* species.

The ratios representing the composition in F/C pairs **1–3**, combined with the total F/C ratio, were utilized to perform a statistical classification by CDA (Figure 4B, Supporting Information S8). This multivariate analysis is used to distinguish sets of observations (e.g., extracts from authenticated *Glycyrrhiza* species) and allocate new observations to previously defined groups (e.g., species). Notably, such a method based on AUC ratios is independent from the retention time and from the status of the UV lamp. Therefore, UHPLC-UV-based CDA can integrate AUC ratios acquired at different times, on different columns and instruments, and does not rely on the generation of calibration curves, collectively promoting the universality of the proposed classification method.

The first canonical score plot (Figure 4B) confirmed the box chart results (Figure 4A). The integration of the AUC ratios from **5** taken at 254 nm ($[\text{AUC glycyrrhizin}]/[\text{total AUC}]$) did not alter nor optimize the classification results, thereby confirming again that the ubiquitous *Glycyrrhiza* metabolite **5** does not allow the distinction of *Glycyrrhiza* species. The canonical score plot 1 (Figure 4B) demonstrated the heterogeneity within all *G. uralensis* extracts regarding their F/C composition. These observations complement the results of

Kondo and co-workers,¹⁹ who also noted the content variation of the F/C pair **1**, as well as compounds **4F** and **5**, in their DNA-authenticated samples of *G. uralensis*. All *G. glabra* samples clustered independently of their geographical origins.

Recently Yu and co-workers⁴⁵ demonstrated that the presence of **1F** and **1C** is genetically determined in *G. uralensis*, whereas their respective concentrations together with the abundance of **4F**, **4C**, and **5** are determined by environmental pressure. Despite the fact that the absolute concentrations of all principal F and C isomers have been shown to greatly fluctuate between samples,^{21,45–47} we demonstrate here that the proportionality between major F and C, as well as the composition in F/C pairs, remained stable between various samples of the same species. Hence, the F/C composition is unique for each of the three principal *Glycyrrhiza* species and can be considered as being determined genetically. This is the first study to show that pharmacopeial *Glycyrrhiza* species can be classified according to the proportionality of their major F/C pairs, regardless of the absolute concentration of each individual F and C.

In a subsequent step, the AUC ratios obtained for the species-specific compounds **6**, **7**, and **9** were added to the classification table, and a second CDA was performed (Figure 4C). The canonical score plot 2 offered an improved clustering of *Glycyrrhiza* species, thereby fostering their differentiation. Both canonical score plots were utilized for further identification and classification of *Glycyrrhiza* mixtures, hybrids, and outliers.

SIMCA and CDA for the Characterization of Outliers, Hybrids, and Mixtures

The second major objective of this study was to utilize the newly developed chemometric models for the classification and characterization of commercial licorice mixtures and *Glycyrrhiza* hybrids, in addition to the identification of potential outliers. The PCA result obtained with the downfield ¹H NMR region (84 buckets) was validated to build a classification model by SIMCA. By analyzing the influence plot, the SIMCA classification indicated whether a sample/spectrum belonged to the *Glycyrrhiza* model (model space). In the influence plot, the y-axis represents the calculated residual variances (Q residual) between the analyzed spectrum and the model space, whereas the x-axis indicates Hotelling's T^2 scores, representing the distance of the analyzed spectrum to the model center. The score plots revealed the position of the sample relative to each species cluster (Figures 5–7A and B). The AUC ratios calculated from the UHPLC-UV data were incorporated into each CDA table in order to classify these samples according to their F/C composition (score plot 1, Figures 5–7C), with and without considering species-specific compounds (score plot 2, Figures 5–7D). The CDA classification results automatically attributed each observation/sample within a predetermined group, whereas the canonical score plots reflected the overall chemical composition and trends.

Three Species Outliers Were Identified by ¹H NMR/SIMCA

The samples BC747, BC755, and BC686 were DNA authenticated as being *G. glabra*, *G. uralensis*, and a mixture of *G. glabra* and *G. uralensis*, respectively, as shown in Figure 5 as well as Tables 2 and 4. According to the SIMCA classification, all three samples fell outside the model space, above both confidence lines, and, thus, were characterized as outliers

(Figure 5B). Both canonical score plots, either when using solely the information from their F/C composition or the species-specific compounds, confirmed the DNA identification of each sample (Figures 5C and D). This outcome illustrates that both chemometric models are complementary: CDA can confirm the species identity, whereas SIMCA can facilitate the identification of species outliers with unusual chemical profiles.

Further analyses of qHNMR spectra and UHPLC-UV chromatograms highlighted the occurrence of major chemical differences compared to the other confirmed *Glycyrrhiza* species (Supporting Information S9). For example, *G. uralensis* sample BC755 did not contain any detectable polar primary metabolites such as sucrose, while **4F** was found to be particularly abundant. *G. glabra* sample BC747 was characterized by a high concentration of glycyroside [formononetin-7-*O*-apiosyl-(1→2)-glucoside], with a proportionally lower concentration of the F/C pairs **2** and **3**. The *Glycyrrhiza* mixture BC686 was characterized by a remarkably high concentration of phloretic acid and **5**, while being proportionally depleted in both Fs and Cs.

SIMCA and CDA Classification of *Glycyrrhiza* Hybrids

The SIMCA and CDA plots in Figure 6 demonstrate that all DNA-authenticated hybrids had distinct metabolite profiles. The *G. uralensis* × *G. inflata* hybrid, BC782c, had an F/C composition comparable to *G. glabra* and, together with the hybrids BC782d, contained compound **7**, which is *G. inflata* specific. Inversely, BC752, BC758, displayed an F/C composition comparable to *G. uralensis*, but compound **7** was not detected, and compound **9** was only found in BC752. The sample BC750 determined to be a *G. glabra* × *G. uralensis* hybrid was classified within the *G. glabra* group due to its characteristic F/C composition and despite the absence of **6**. The SIMCA results indicated that samples BC782c and d, BC772a, and BC750 displayed significant off-model space components due to minor differences in their NMR signals, which were not present in the NMR spectra of other *Glycyrrhiza* samples within the model space. Nevertheless, prior DNA authentication was found to be a prerequisite to make both chemometric models suitable for the characterization of *Glycyrrhiza* hybrids.

Analysis of Licorice Mixtures via in-House Mixture Validation

In order to better understand the results achievable with commercial mixtures (Figure 7), licorice mixtures of known composition (Table 4) were analyzed by SIMCA and CDA. Both chemometric models revealed that mixtures were classified according to their most abundant species. When *G. uralensis* (Mix 1, 2, and 4) and *G. glabra* (Mix 1, 8, and 10) were present at >50% w/w in the mixture, and *G. inflata* was present at >30% w/w (Mix 5 and 6), the samples fell within or close to their corresponding species groups. *G. inflata* displayed a sufficiently unique chemical profile to be detected chemically at a lower percentage in a mixture (30% vs 50%). On both PCA and CDA score plots, the comparison of the positions obtained for in-house mixtures with those of commercial mixtures enabled the identification of the second most preponderant species. Samples BC625, Mix 5, and Mix 6 could be classified similarly, revealing that the first most abundant species in BC625 is *G. inflata*, the second is *G. glabra*, closely followed by *G. uralensis* (Figure 7, Table 4). Samples BC741 and BC744 were classified similarly to Mix 8 and Mix 1, respectively. Therefore, the most

abundant species in BC741 was found to be *G. uralensis* in both samples (>50% w/w), followed by *G. inflata* and *G. glabra*. Samples BC733 and BC734 were also classified as *G. glabra*. Sample BC734 was identified as a mixture of *G. glabra* genotypes (TG2 and TG3), whereas BC733 contained >50% w/w of *G. glabra* (TG3). Interestingly, the mislabeled sample BC736 was considered as chemically close to *G. inflata* by both statistical models and contained a remarkable amount of **7**. Notably, this sample was sold in the U.S. market without containing any of the recognized U.S. pharmacopeial *Glycyrrhiza* species, i.e., *G. glabra* and *G. uralensis*.^{17,18}

Summary and Outlook

In conclusion, this study proposes a metabolomic approach for the classification and characterization of three *Glycyrrhiza* species, which are widely accepted source plants for the production of BDSs. Complementary multivariate analyses adapted to two analytical platforms, NMR and LC, were utilized for this purpose. To ensure high specificity, only DNA-verified samples containing a single species were employed to build the classification models.

Offering an unbiased and holistic analytical overview of the licorice metabolite composition, ¹H NMR spectra were utilized for PCA and SIMCA and led to the untargeted classification of licorice samples. In parallel, UHPLC-UV was employed for LC-based classification, taking advantage of its inherently targeted characteristics, aimed at UV–visible compounds. The outcomes show that results obtained from chromatographic systems are less reproducible than those obtained by NMR, which is evidently due to the variations associated with retention time and intensity of UV absorbance. To overcome such limitations, an LC method using only the AUC ratios was developed. This led to the establishment of a CDA method that enabled a targeted classification of *Glycyrrhiza* species according to their unique F/C composition with and without considering species-specific compounds.

The newly developed chemometric models, which were built in agreement with the initial botanical authentication by DNA analysis, enabled the classification of licorice botanicals, confirmed the metabolite profiles of misidentified *Glycyrrhiza* species, and facilitated the metabolomic characterization of mixtures, species outliers, and hybrids. This approach is in line with the view of Harnly and co-workers that “the more we know about a sample, the easier it is to develop an accurate assay”.⁴⁸ Hence, the present study provides the first comprehensive method that combines the use of botanical authentication by DNA barcoding prior to extraction, with orthogonal metabolomic analyses after extraction, for the rigorous classification and characterization of *Glycyrrhiza* species.

The present results also accentuate the existence of marked chemical differences between *Glycyrrhiza* species in terms of their major secondary metabolites. Various F/C pairs were identified as species markers in addition to the accepted species-specific glabridin, licochalcone A, and glycy coumarin. As expected, glycyrrhizin (**5**) is inadequate for species distinction. From a phytochemical point of view, variations in pharmacological/biological activities and potential herbal–drug interactions can be expected, not only for each species but also for their mixtures and hybrids.^{19,20,23} When licorice is used for medicinal purposes

and in light of the distinct properties of its metabolites, it is important that the exact species is selected.^{19,23} However, studies comparing the biological properties of authenticated *Glycyrrhiza* species are rare.^{49,50} At the same time, this kind of study design is paramount for the adequate chemical and biological standardization.

In light of these observations and the results obtained herein, pharmacopeial and other monographs supporting the QC of “licorice” should consider a metabolomic characterization associated with the identification and quantitation of several key markers⁵¹ rather than the (quantitative) determination of a single ubiquitous *Glycyrrhiza* constituent, such as glycyrrhizin (**5**).

The present study also highlights the importance of building integrative botanical authentication measures. These should include DNA analysis of the raw plant material as well as metabolomic characterization of the extracts, which collectively can enhance the quality control process of botanicals. From a general point of view, the main role of DNA analysis is the botanical authentication of raw plant material, upstream of the industrial BDS production process, whereas NMR- and LC-based metabolomic analyses should primarily be implemented as a QC measure throughout the extraction and further processing chain for the assessment of the chemical composition. Hence, the QC of licorice botanicals requires that the data acquired from both botanical and analytical chemistry platforms are merged. In addition, classical methods employed for the botanical authentication of licorice raw material should be complemented by DNA analysis.

Considering the chemical diversity and complexity of *Glycyrrhiza* constituents, metabolomic and chemometric approaches that assess the compositional characteristics in terms of bioactive markers are more suitable for the QC of licorice extracts. To this end, the present study offers guidance for the implementation of a more comprehensive QC of pharmacopeial licorice materials and may inspire the development of more specific standardization protocols that enhance the integrity of widely used licorice preparations.

EXPERIMENTAL SECTION

General Experimental Procedures

All DNA preparations were checked using a Thermo Scientific NanoDrop 1000 spectrophotometer (Waltham, MA, USA). Polymerase chain reactions (PCRs) were performed with Big Dye Terminator v3.1 chemistry, and the amplicons were electrophoresed on an ABI 3730xl DNA analyzer (Life Technologies, Carlsbad, CA, USA). DNA sequence chromatograms were edited and analyzed with Sequencher v5.2 (Gene Codes, Ann Arbor, MI, USA). All ground *Glycyrrhiza* samples were extracted with an Accelerated Solvent Extraction (ASE 350) from Dionex Corporation (Sunnyvale, CA, USA). The extracts were dried using a Thermo-Fischer Savant SC250 EXP speed vacuum equipped with an RVT4104 refrigerator vapor trap. Freeze-drying was performed on a Labconco Freezone 4.5 (Kansas City, MO, USA). A precise Mettler Toledo XS105 dual range analytical balance was employed to prepare extracts for UHPLC and qHNMR analyses. The UHPLC analyses were performed on a Shimadzu UPLC, equipped with a Kinetex XB-C18 (2.1 × 5.0 mm, 1.7 μm, 00B-4498-AN, Phenomenex) column, and using a

diode array detector (DAD, Shimadzu SPD-M20-A). The autosampler temperature was set at 4 °C, and the column oven temperature at 40 °C. Postrun data analyses were done with the Shimadzu LabSolution software package. CDA was performed with the OriginPro 9.1 software (Northampton, MA, USA). Samples for NMR analyses were prepared with a Pressure-Lock gas syringe (VICI Precision Sampling Inc., Baton Rouge, LA, USA) and calibrated glass pipets (cat. no: 2-000-200, Drummond Scientific, Broomall, PA, USA). Standard NMR tubes of 3 mm, 7 in., were from Norell (part no. S-3-HT-7, Norell Inc., Landisville, NJ, USA). NMR spectra were acquired on a Bruker AVANCE 600.13 MHz spectrometer equipped with a 5 mm TXI cryoprobe. Off-line data processing was performed using the Topsin 3.0.b.8 NMR software package. Multivariate data analyses, PCA, and SIMCA were performed with the AMIX statistics 3.9.14 software, both from Bruker Biospin Corporation (Karlsruhe, Germany).

Reagents

All chemicals and reagents including HPLC-grade solvents were obtained from Fischer Scientific (Hanover Park, IL, USA). Licorice standards glycyrrhizin, glabridin, and licochalcone A were obtained from Sigma-Aldrich (St. Louis, MO, USA). Glycycoumarin was purchased from BioBioPha Co., Ltd. (Kunming Institute of Botany, China). For NMR acquisition, DMSO-*d*₆ (99.9% D) was purchased from Cambridge Isotope Laboratories Inc. (Andover, MA, USA). The PCR-grade tubes, tips, and most biological reagents used for DNA authentication, including the DNeasy Plant mini kit, were acquired from Qiagen (Valencia, CA, USA) and/or Thermo-Fisher Scientific and Beckman Coulter (Indianapolis, IN, USA). The BeadBug microtube homogenizer was purchased from Benchmark Scientific, Inc. (Edison, NJ, USA). The bovine serum albumin, BSA, was obtained from New England Biolabs (Ipswich, MA, USA). Seakem LE agarose for gel electrophoresis was purchased from Lonza (Rockland, ME, USA). The Ambion RNase-free water was acquired from Life Technologies.

Plant Material

A total of 51 licorice samples sold as bulk root powders, sticks, or capsules were either purchased online and from Chicago metropolitan area stores or acquired through our collaborative network. Sampling of BDS capsules was restricted to dried preparations (powdered plant tissue in gelatin capsules). To protect the manufacturer's identity, the samples were reported with internal Botanical Center codes (Tables 1–4). Samples BC629, BC693–694, BC711, BC746–747, BC749, BC751–755, BC772, BC774, and BC778 were botanically/macrospectically verified prior to inclusion.

DNA Isolation

Tissue for DNA authentication was sampled and washed twice with 100% EtOH using aseptic techniques under sterile conditions, before homogenization to remove any loose debris or contaminants on the surface of the sampled tissue or powder. Total DNA was extracted with the DNeasy plant mini kit following the manufacturer's recommendations with some modifications. For each sample, total DNA was extracted from approximately 50–100 mg of tissue. Tissue was disrupted with the BeadBug microtube homogenizer with 3

mM triple-pure zirconium beads. Concentration and purity of DNA (260/280 nm) were checked using NanoDrop.

Primers

Primer sequences were obtained directly from the literature or designed from licorice GenBank accessions. New modified primers were designed to produce larger amplicons than obtained by Kondo and co-workers,²⁷ leading to a better quality of sequence for our Sanger DNA sequencing methods. Samples that did not initially amplify due to excessive DNA degradation required the use of primers that made smaller amplicons (<200 bp). Two sets of primers from Kondo and co-workers²⁷ (*rbcL* and *matK*), as well as custom-designed primers for *ITS* and *psbA-trnH*, were utilized in PCR for smaller amplicons. An initial PCR testing of the *ITS* primers designed by White and co-workers³³ (*ITS5* 5'-GGAAGTAAAAGTCGTAACAAGG-3' and *ITS4* 5'-TCCTCCGCTTATTGATATGC-3') showed some fungal interference due to rDNA sequence homology between fungi and plants. More plant-specific primers were designed based on the partial 5.8S alignment originally put forth by Oliveira de Miranda and co-workers³⁴ with extra GenBank *ITS1*–*5.8S*–*ITS2* accessions covering some of the major fungi and plant families (including licorice). The *ITS* primers designed from our alignment lay just internal to the primers by White and co-workers³³ and were *ITS-Fint* 5'-TCGATGCCTTGCAAGCAGT-3' and *ITS-Rint* 5'-AGAGCCCCAACTCAGTGGA-3'. The *ITS-Fint* and *ITS-Rint* primers make a PCR product of 585 bp in all three *Glycyrrhiza* species, whereas primers designed by White and co-workers³³ made a PCR product of 729 bp. For the three chloroplast loci *matK*, *rbcL*, and *psbA-trnH*, the forward and reverse primers used, and the expected PCR product size for all species are as follows: *matK*F1 5'-GTGTCAGATATACGAATACC-3' and *matK*R2 5'-TCCTTGAAGCCAGAATGG-3' (430 bp for *G. glabra* and *uralensis*; 424 bp for *G. inflata* TG5); *rbcL*F1 5'-CGCGCTCTACGTCTTGAGGA-3' and *rbcL*R1 5'-GCGTGAATATGATCTCCACCAGA-3' (593 bp); *psbA-trnH*F1 5'-ACGGGAATTGAACCCGCGCA-3' and *psbA-trnH*R1 5'-CATATGACTTCACAATGTA-3' (239 bp). We also used the *psbA-trnH* primers *psbA3'* 5'-GTTATGCATGAACGTAATGCTC-3' (Sang and coworkers⁵²) with *trnHf* 5'-CGCGCATGGTGGATTACAATCC-3' (434 bp, Tate & Simpson⁵³), and *trnH* (GUG) 5'-ACTGCCTTGATCCACTTGGC-3' with *psbA* 5'-CGAAGCTCCATCTACAAATGG-3' (362 bp, Hamilton⁵⁴). Primers designed for degraded DNA samples were *trnHF200* 5'-ATCCACTTGGCTACATTCGC-3' and *trnHR200* 5'-GTAATACATATGACTTCACAATG-3' (198 bp); *ITSYRM-F* 5'-CGTGAACCATCGAGTCTTTGAACG-3' and *ITSYRM-R* 5'-CTCAGTTTTGAGCCAACCGTG-3' (180 bp); *ITSY-F* 5'-CGCTGAATGCGCCAAGGAAC-3' and *ITSY-R* 5'-GTTCTTCATCGATGCAAGAGCC-3' (141 bp). Primers were purchased from Integrated DNA Technologies.

PCR and Sequencing

Samples were PCR-amplified and bidirectionally sequenced for all four loci *ITS*, *rbcL*, *matK*, and *psbA-trnH*. Individual amplifications took place in a 30 μ L volume containing 3 μ L of Qiagen 10 $\times$ PCR buffer (with 15 mM MgCl₂), 0.8 μ L of Promega deoxynucleotides

(dNTPs, at 10 mM each), 0.2 μ L of Qiagen HotstarTaq polymerase (5 U/ μ L), 2.5 μ L of each primer (5 μ M), and 1 μ L of total DNA (5–50 ng/ μ L). The solutions were completed with up to 30 μ L with Ambion RNase-free water. Approximately 25 ng of total DNA was added to each PCR. The addition of 300 ng/ μ L BSA greatly enhanced amplification of samples that repeatedly yielded impure DNA based on a low 260/280 nm ratio (<1.5). A PCR reagent master mix was made for each primer pair depending on the number of samples to amplify, and water controls (no template) were included in all PCR reactions for each primer pair. The general PCR cycling conditions were as follows: 94 °C for 15 min; 45 cycles: 94 °C for 20 s, 58 °C for 30 s, 72 °C for 50 s; 72 °C for 8 min and hold at 4 °C. PCR reactions that produced single clear bands were purified with the Qiagen PCR purification kit. The concentration and purity of each PCR product were determined using NanoDrop. Amplified PCR products were diluted to 20 ng/ μ L and cycle sequenced utilizing ABI Big Dye Terminator v3.1 chemistry, purified with the Agencourt CleanSEQ kit (Beckman Coulter), and then electrophoresed on an ABI 3730xl DNA analyzer, all according to manufacturer recommendations. All PCR primers were used as cycle sequencing primers. DNA sequence chromatograms were edited and analyzed with Sequencher v5.2. The PCR conditions for *psbA-trnH* were reformatted for total DNA preparations that showed poor results with the general PCR conditions.

GenBank Accessions for Primer Design and Reference

G. uralensis ITS (AB280738), *matK* (AB280741), *rbcL* (AB012126), *G. glabra psbA-trnH* (AB280745), *rbcL* (AB012126), *G. inflata rbcL* (AB012127), *G. lepidota* ITS (U50758 and U50759) *rbcL* (AB126685).

Sample Extraction

Each licorice sample was extracted using 2 g of material for 40 mL of 95% EtOH (USP 190 proof) utilizing the ASE 350 setup at 80 °C, for 30 min of static time at 1500 psi, for a total of 45 min of extraction. The extractions were performed in duplicate, when possible. After extraction, the plant material was washed with 10 mL of the same solvent. Each extract was first concentrated, then freeze-dried. Prior to NMR analysis, samples were dried under vacuum for 24 h in the presence of P₂O₅.

UHPLC-UV Analyses

Solutions of licorice extract were prepared at 10 mg/mL in HPLC-grade MeOH (2 μ L injection). The Kinetex column was eluted with a gradient composed of H₂O/0.1% formic acid (A) and MeCN/0.1% formic acid (B) as follows: 8% to 11% B in 2 min, and during 30 s, to 13% B at 4 min, to 15.5% B at 8 min, to 36% at 13 min, and during 30 s, to 80% B at 21 min, and during 1 min, back to 8% B at 23 min (flow rate: 0.7 mL/min). Under these conditions, the retention times (*t_R*) of the licorice compounds were as follows: HBMA (**8**) 0.6 min, liquiritin (**1F**) 2.77 min, liquiritigenin 7-*O*-apiosylglucoside (**2F**) 3.08 min, liquiritin apioside (**3F**) 3.25 min, liquiritigenin (**4F**) 5.82 min, isoliquiritin (**1C**) 6.76 min, isoliquiritin apioside (**3C**) 7.11 min, licuraside (**2C**) 7.70 min, isoliquiritigenin (**4C**) 10.99 min, glycyrrhizin (**5**) 13.47 min, glabridin (**6**) 15.13 min, licochalcone A (**7**) 14.87 min, glycycomarin (**9**) 13.77 min. All Fs and Cs as well as compound **8** were identified and/or

isolated as previously described.^{20,26,40} Each extract was analyzed in duplicate during the same UHPLC sequence. Some extracts were analyzed a second time after 6 months in order to confirm the robustness of the classification method.

¹H NMR Acquisition

NMR data acquisitions were performed using 3,5-dinitrobenzoic acid (DNBA, Fluka, TraceCERT, purity P = 99.54% w/w lot no. BCBH8381 V) as internal calibrant (IC). The IC stock solution was prepared at 6.15 mM in a mixture of D₂O/DMSO-*d*₆ (1.5:8.5). NMR samples of crude extracts were prepared by precisely weighing 7–8 mg of each extract, followed by the addition of 200 μ L of DMSO-*d*₆ and 100 μ L of IC stock solution (2.05 mM final concentration). From this solution, a 200 μ L aliquot was transferred with calibrated glass pipets into 3 mm standard NMR tubes. The 1D ¹H NMR spectra were acquired at 298 K under quantitative conditions (qHNMR) using a 90° excitation pulse experiment (Bruker pulprog: zg). The 90° pulse width for each sample was determined by prorating the measured 360° pulse width ($p_{90} = 1/4 \times p_{360}$). The probe was frequency tuned and impedance matched before each acquisition. For each sample, 18 scans (ns) and 4 dummy scans (ds) were recorded with the following parameters: pulse width (P1) of typically 9.25 μ s (90°), spectral width of 28 ppm, relaxation delay (D1) of 60 s, receiver gain (RG) set to 256. The total duration of each ¹H NMR acquisition was 23 min.

¹H NMR Multivariate Data Analysis: PCA and SIMCA

Each acquired qHNMR spectrum was first processed with TopSpin 3.0.b.8. The ¹H NMR chemical shifts (δ) were expressed in ppm with reference to the residual solvent signal of DMSO-*d*₆, i.e., DMSO-*d*₅, set to 2.500 ppm. The following processing scheme was applied: a mild Lorentzian-to-Gaussian window function (line broadening = −0.3 Hz, Gaussian factor = 0.05) was applied, followed by two zero fillings before fast Fourier transformation. After manual phasing, a fifth-order polynomial baseline correction was applied. Using the AMIX software (version 3.9.14), and prior to any statistical analysis, the spectrum baselines were adjusted. Spectral intensities were reduced to integrated regions, referred to as buckets of equal width (0.04 ppm) within the region of 8.50 to 0.50 ppm and within the region of 8.50 to 4.50 ppm (downfield region). In the resulting bucket tables, all rows were scaled to the total intensity, and a Pareto scaling was applied for the columns preceding PCA. The regions between 4.00 and 3.00 ppm corresponding to the sugar region, the sucrose anomeric resonance at 5.169 ppm (doublet $J = 3.75$ Hz), the proline signals (from 1.60 to 2.12 ppm),⁴⁰ and the residual HOD (singlet at 3.149 ppm), D₂O (singlet at 3.567 ppm), EtOH (triplet at 1.043 ppm, $J = 7.01$ Hz), and DMSO-*d*₅ (2.500 ppm) signals were removed from the bucket table prior to multivariate analyses. PCA using the full-spectrum bucket table (143 buckets) was performed with a total of 83 spectra. A total of 16 principal components (PCs) were selected to explain 95.1% of the variance, and the confidence level was set at 99.99% in order to include the authenticated *G. inflata* samples. The PCA utilizing only the downfield region contained 84 buckets, and 15 PCs were selected to explain 95.7% of the total variance. The confidence level was maintained at 99.99%. From this PCA model, cross-validation and test set validation were performed to build the SIMCA classification model.

UHPLC-UV Multivariate Data Analysis: CDA

For each UHPLC-UV chromatogram, the AUC values were taken at 275 nm for all major Fs (**1–4F**) and at 360 nm for all major Cs (**1–4C**). For each F/C pair, the AUC values were added, divided by the sum of all considered AUCs taken at 275 and at 360 nm (Supporting Information S7,8), and expressed as a percentage of all considered major F/C pairs. Finally, the total F/C ratio was determined by dividing the sum of all AUCs taken at 275 nm (major Fs) by the sum of all AUCs considered at 360 nm (for major Cs): [sum of AUC 275 nm]/[sum of AUC 360 nm]. The percentages of major F/C pairs **1–3** and the total F/C ratio were utilized to perform a classification by CDA. The first canonical score plot was constructed using only the ratios reflecting the F/C composition of each extract. The second score plot considered additionally the AUC ratios from the species-specific metabolites: glabridin ([AUC glabridin 275 nm]/[sum of considered AUC 275 nm]), licochalcone A ([AUC licochalcone A 360 nm]/[sum of considered AUC 360 nm]), and eventually glycycomarin ([AUC glycycomarin 360 nm]/[sum of considered AUC 360 nm]). The evaluation of the classifier was performed with the results obtained from independent acquisition time and generated from the same extracts. In both instances, prior probabilities were proportional to each group size. A linear discriminant function was selected. The error rate for the classification and cross-validation of training data was 0%.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors are grateful to A. Massarotto from Nature Med (Italy), Drs. K. Spelman from Herb Pharm (OR, USA), Dr. S. Gafner (American Botanical Council, USA), and Dr. A. Schinkovitz (Angers, France), as well as Drs. M. Farag and W. Ludger (Leibniz-Institut für Pflanzenbiochemie, Germany), and Dr. G. Topcu (Bezmialem Vakif University, Turkey) for kindly providing various licorice samples. The authors are thankful to Dr. L. Zhao (Lanzhou Institute of Chemical Physics CAS) for collecting wild *G. inflata* samples. The authors also acknowledge M. Coven from Vitality Works (NM, USA) for providing GMP-grade licorice extracts. C.S. is particularly grateful to Dr. J. Herrou for his constant support and thoughtful advice in the course of the research project. C.S. also wishes to acknowledge Dr. D. D. Soejarto and T. Burton for their guidance in microscopic analysis. The authors are grateful to the DNA service facility of the UIC Research Resources Center (UIC RRC), in particular its director, Dr. S. Green, for excellent DNA analytical support. The authors wish to acknowledge Dr. B. Ramirez for his support at the NMR facility of the UIC Center for Structural Biology (UIC, CSB). The construction of this facility was funded by NIH grant P41 GM068944, awarded by NIGMS/NIH (PI: Dr. P. G. W. Gettins). Financial support for this work was provided through grants P50AT000155 by NCCIH and ODS, as well as RC2AT005899 by NCCIH of the NIH.

References

1. Betz JM, Fisher KD, Saldanha LG, Coates PM. Anal Bioanal Chem. 2007; 389:19–25. [PubMed: 17541565]
2. Cordell GA, Colvard MD. J Nat Prod. 2012; 75:514–525. [PubMed: 22233451]
3. Smillie TJ, Khan IA. Clin Pharmacol Ther. 2010; 87:175–186. [PubMed: 20032974]
4. Techen N, Parveen I, Pan Z, Khan IA. Curr Opin Biotechnol. 2014; 25:103–110. [PubMed: 24484887]
5. Li M, Cao H, But PPH, Shaw PC. J Syst Evol. 2011; 49:271–283.
6. Chen S, Pang X, Song J, Shi L, Yao H, Han J, Leon C. Biotechnol Adv. 2014; 32:1237–1244. [PubMed: 25087935]
7. Hebert PDN, Gregory TR. Syst Biol. 2005; 54:852–859. [PubMed: 16243770]

8. Van der Kooy F, Maltese F, Choi YH, Kim HK, Verpoorte R. *Planta Med.* 2009; 75:763–775. [PubMed: 19288400]
9. Gad HA, El-Ahmady SH, Abou-Shoer MI, Al-Azizi MM. *Phytochem Anal.* 2013; 24:1–24. [PubMed: 22678654]
10. Petrakis EA, Cagliani LR, Polissiou MG, Consonni R. *Food Chem.* 2015; 173:890–896. [PubMed: 25466103]
11. Hayashi H, Sudo H. *Plant Biotechnol.* 2009; 26:101–104.
12. Asl MN, Hosseinzadeh H. *Phytother Res.* 2008; 724:709–724. [PubMed: 18446848]
13. Fiore C, Eisenhut M, Ragazzi E, Zanchin G, Armanini D. *J Ethnopharmacol.* 2005; 99:317–324. [PubMed: 15978760]
14. Zhang Q, Ye M. *J Chromatogr A.* 2009; 1216:1954–1969. [PubMed: 18703197]
15. Nomura T, Fukai T, Akiyama T. *Pure Appl Chem.* 2002; 74:1199–1206.
16. Nomura T, Fukai T. *Prog Chem Org Nat Prod.* 1998; 73:1–158.
17. World Health Organization. WHO Monographs on Selected Medicinal Plants. Vol. 1. World Health Organization; Geneva: 1999. p. 183–194.
18. United States Pharmacopeia. [accessed Jan 6, 2015] 2622: Dietary Supplements: Licorice. http://www.pharmacopeia.cn/v29240/usp29nf24s0_m45050.html
19. Kondo K, Shiba M, Nakamura R, Morota T, Shoyama Y. *Biol Pharm Bull.* 2007; 30:1271–1277. [PubMed: 17603166]
20. Simmler C, Jones T, Anderson JR, Nikoli DC, van Breemen RB, Soejarto DD, Chen SN, Pauli GF. *Phytochem Anal.* 2014; 25:378–388. [PubMed: 25859589]
21. Zhou S, Cao J, Qiu F, Kong W, Yang S, Yang M. *Phytochem Anal.* 2013; 24:527–533. [PubMed: 23427103]
22. Yang SO, Hyun SH, Kim SH, Kim HS, Lee JH, Whang WK, Lee MW, Choi HK. *Bull Korean Chem Soc.* 2010; 31:825–828.
23. Farag MA, Porzel A, Wessjohann LA. *Phytochemistry.* 2012; 76:60–72. [PubMed: 22336263]
24. Tao W, Duan J, Zhao R, Li X, Yan H, Li J, Guo S, Yang N, Tang Y. *Food Chem.* 2013; 141:1681–1689. [PubMed: 23870878]
25. Ralston L, Subramanian S, Matsuno M, Yu O. *Plant Physiol.* 2005; 137:1375–1388. [PubMed: 15778463]
26. Simmler C, Hajirahimkhan A, Lankin DC, Bolton JL, Jones T, Soejarto DD, Chen SN, Pauli GF. *J Agric Food Chem.* 2013; 61:2146–2157. [PubMed: 23427769]
27. Kondo K, Shiba M, Yamaji H, Morota T, Zhengmin C, Huixia P, Shoyama Y. *Biol Pharm Bull.* 2007; 30:1497–1502. [PubMed: 17666810]
28. Consortium for the Barcode of Life, C. Plant Barcode Region Proposals — Review Report. http://www.barcoding.si.edu/plant_working_group.html
29. Baker DA. *J AOAC Int.* 2012; 95:1023–1034. [PubMed: 22970567]
30. Ausubel JH. *Proc Natl Acad Sci U S A.* 2009; 106:12569–12570. [PubMed: 19666620]
31. Stoeckle MY, Gamble CC, Kirpekar R, Young G, Ahmed S, Little DP. *Sci Rep.* 2011; 1:42. [PubMed: 22355561]
32. Little DP. *Genome.* 2014; 57:1–4. [PubMed: 24564210]
33. White, T.; Bruns, T.; Lee, S.; Taylor, J. *PCR Protocols: A Guide to Methods and Applications.* Academic Press; San Diego: 1990. p. 315–322.
34. Oliveira de Miranda V, Martins VG, Furlan A, Bacci M. *Braz Arch Biol Technol.* 2010; 53:141–152.
35. Zhang Q, Sodmergen. *J Plant Res.* 2010; 123:201–206. [PubMed: 20052516]
36. Pan B, Zhang Y. *Sci China, Ser D: Earth Sci.* 2002; 45:174–179.
37. Zhang YM, Chen YN, Pan BR. *J Arid Environ.* 2005; 63:772–784.
38. Yamamoto Y, Tani T. *J Trad Med.* 2006; 23:27–35.
39. Pauli GF, Chen S, Simmler C, Lankin DC, Gödecke T, Jaki BU, Friesen JB, McAlpine JB, Napolitano JG. *J Med Chem.* 2014; 57:9220–9231. [PubMed: 25295852]

40. Simmler C, Nikoli D, Lankin D, Yu Y, Friesen J, van Breemen RB, Lecomte A, Le Quémener C, Audo G, Pauli G. *J Nat Prod*. 2014; 77:1806–1816. [PubMed: 25080313]
41. Simmler C, Napolitano JG, McAlpine JB, Chen SN, Pauli GF. *Curr Opin Biotechnol*. 2014; 25:51–59. [PubMed: 24484881]
42. Heyman H, Meyer J. *S Afr J Bot*. 2012; 82:21–32.
43. Pinheiro C, Passarinho JA, Ricardo CP. *J Plant Physiol*. 2004; 161:1203–1210. [PubMed: 15602812]
44. Vacheron J, Desbrosses G, Bouffaud ML, Touraine B, Moënné-Loccoz Y, Muller D, Legendre L, Wisniewski-Dyé F, Prigent-Combaret C. *Front Plant Sci*. 2013; 4:356. [PubMed: 24062756]
45. Yu F, Wang Q, Wei S, Wang D, Fang Y, Liu F, Zhao Z, Hou J, Wang W. *Biol Pharm Bull*. 2015; 38:75–81. [PubMed: 25744461]
46. Rauchensteiner F, Matsumura Y, Yamamoto Y, Yamaji S, Tani T. *J Pharm Biomed Anal*. 2005; 38:594–600. [PubMed: 15967286]
47. Kojoma M, Hayashi S, Shibata T, Yamamoto Y, Sekizaki H. *Biol Pharm Bull*. 2011; 34:1334–1337. [PubMed: 21804228]
48. Harnly JM, Luthria D, Chen P. *J AOAC Int*. 2012; 95:1579–1587. [PubMed: 23451372]
49. Hajirahimkhan A, Simmler C, Jeffery A, Yuan Y, Nikoli D, Chen S, Dietz BM, Pauli GF, van Breemen RB, Bolton JL. *PLoS One*. 2013; 8:e67947. [PubMed: 23874474]
50. Dunlap T, Wang S, Simmler C, Chen S-N, Pauli GF, Dietz B, Bolton JL. *Chem Res Toxicol*. 2015:150716145843000.
51. Bilia AR. *J Ethnopharmacol*. 2014; 158:487–494. [PubMed: 25017375]
52. Sang T, Crawford D, Stuessy T. *Am J Bot*. 1997; 84:1120–1136. [PubMed: 21708667]
53. Tate JA, Simpson BB. *Syst Botany*. 2003; 28:723–737.
54. Hamilton MB. *Mol Ecol*. 1999; 8:513–525. [PubMed: 10199012]

Figure 1.

DNA barcoding data used for the authentication of *Glycyrrhiza* species and the determination of hybrids and mixtures. (A) Determination of *Glycyrrhiza* genotypes according to the combination of DNA markers (alleles) at four different loci (*rbcL*, *matK*, and *psbA-trnH* in the cpDNA and *ITS* in the nrDNA). (B) Comparative analysis of nucleotide sequence at specific sites of the *ITS* and *psbA-trnH* intergenic region contributing to the identification of the alleles I-2, I-3, and T-1 to T-4, respectively. Both A and B are information obtained from Kondo's paper.²⁷ (C) Comparative nucleotide sequences at the targeted sites indicated in B for both *ITS* and *psbA-trnH*, illustrating the differences between pure species (BC754), hybrid (BC758), and a mixture of *Glycyrrhiza* species (BC742). Each nucleotide is color-coded: A for adenine in green, T for thymine in red, C for cytosine in blue, and finally G for guanine in black.

Figure 2.

Representative UHPLC-UV chromatograms and ^1H NMR spectra of extracts from individual *Glycyrrhiza* species. Comparative UHPLC-UV chromatograms and ^1H NMR (4.50–8.50 ppm) spectra of representative extracts from *G. uralensis* (A), *G. glabra* (B), and *G. inflata* (C). The ^1H NMR spectra were acquired under quantitative conditions in $\text{DMSO}-d_6$ (+5% v/v D_2O) at 600 MHz. The assignments were determined with the NMR spectra of reference standards (Supporting Information S12). The NMR spectra were directly subjected to PCA and SIMCA. The AUC values obtained for **1–4F** and **1–4C** from the chromatograms were used to determine the composition of each extract in major F/Cs in order to perform CDA. The numbers attributed to UHPLC-UV peaks and ^1H NMR signals follow those listed in Table 5 and Scheme 1.

Figure 3.

PCA using quantitative ^1H NMR (qHNMR) spectra of DNA-authenticated *Glycyrrhiza* species. PCA score plot (A) obtained from the ^1H NMR spectra of all DNA-authenticated *Glycyrrhiza* species and considering only the downfield region from 4.50 to 8.50 ppm. The two major PCs together explained 58.2% of the total metabolomic variance. The three species were differentiated according to their major secondary metabolites. Their discriminant chemical shifts (δ in ppm) are shown on the 2D loading plot (B). The latter represents (B) a total of 84 buckets of 0.04 ppm width and explained the major trends, in terms of NMR signals, leading to samples clustering in the score plot (A). *G. glabra* samples from Iran and Afghanistan, on one hand, and samples from Egypt, on the other hand, formed the two extremities of the *G. glabra* cluster, as seen in (A).

Figure 4.

Comparative F/C composition and CDA of *Glycyrrhiza* species. Box charts (A) represent the proportion of each major F/C pair among all four considered F/C pairs and for each authenticated *Glycyrrhiza* species. The proportion was calculated from the AUC ratios at 275 nm (for Fs) and 360 nm (for Cs). Results were obtained with all EtOH extracts and expressed as mean $\pm$ standard deviation for each F/C pair. For the *G. glabra* samples, the box chart in gray represents the F/C composition of samples from Iran and Afghanistan, for which the F/C pair 2 represents 50% of all four F/C pairs. (B) The score plot 1 results from the CDA initially performed with the calculated F/C ratios and proportions of F/C pairs. The canonical score plot 2 in C was obtained after adding the AUC ratios of 6, 7, and 9 to the table used to generate the score plot 1.

Figure 5.

SIMCA and CDA characterization of *Glycyrrhiza* outliers. Classification results for *Glycyrrhiza* outliers with (A) the PCA score plot and (B) the SIMCA influence plot. The Q residuals plotted on the y-axis of B represent the calculated residual variance of the considered spectra from the model space. The influence plot revealed that samples BC686, BC755, and BC747 are outliers: they are outside the model space, above each confidence line (Supporting Information S9). (C and D) Canonical discriminant plots for the classification of *Glycyrrhiza* outliers. The targeted CDA classification considers only the F/C composition (C) and the presence of species markers (D) and, therefore, did not identify outliers, but rather confirmed the species identity of each sample.

Figure 6.

SIMCA and CDA classification of DNA-authenticated hybrids. (A, B) Classification results for *Glycyrrhiza* hybrids by ^1H NMR/SIMCA analysis. The PCA score plot indicates that BC752 and BC758 displayed a ^1H NMR profile equivalent to the samples in the *G. uralensis* group, whereas BC750 was more closely related to *G. glabra*, and BC782c/d as well as BC772a clustered in the group formed by *G. inflata* (Supporting Information S10). (C, D) Classification score plots of *Glycyrrhiza* hybrids by CDA using the F/C composition (C) and considering additionally species-specific compounds (D).

Figure 7.

SIMCA and CDA classification of mixtures. (A, B) PCA score plots from SIMCA classification results obtained for in-house (A) and commercial (B) mixtures. The canonical score plots (C) for in-house and (D) for commercial mixtures demonstrated that the F/C composition determined for each sample led to the identification of the preponderant *Glycyrrhiza* species. Samples containing >50% w/w of a given species were classified within the same group species (mix 2, 4, 8, and 10). Samples BC733, BC734, and BC742 contained >50% of *G. glabra*. Mix 5, Mix 6, and BC625 were similarly classified in both PCA and CDA score plots, suggesting that >30% w/w of the BC625 mixture is composed of *G. inflata*. According to this chemometric classification, BC741 seemed to contain >50% w/w of *G. uralensis* and 30% w/w of *G. inflata*, whereas BC744 would contain 50% w/w of *G. uralensis*, then 50% of *G. glabra*, and <30% w/w of *G. inflata* (Supporting Information S11).

Scheme 1. Major Metabolites Characterizing the EtOH Extracts of the Investigated *Glycyrrhiza* Species^a

^aLiquiritin (**1F**) and isoliquiritin (**1C**) form the F/C pair 1, liquiritigenin-7-*O*-apiosylglucoside (**2F**) and licuraside (**2C**) define the F/C pair 2, liquiritin apioside (**3F**) together with isoliquiritin apioside (**3C**) form the F/C pair 3, and finally the aglycone liquiritigenin (**4F**) and its isomer isoliquiritigenin (**4C**) are the F/C pair 4. Glycyrrhizin (**5**) is the major common saponin. The phenolic acid, *p*-hydroxybenzylmalonic acid (**8**), was recently found in extracts from all three species.⁴⁰ Glabridin (**6**), licochalcone A (**7**), and glycycomarin (**9**) are specific metabolites from *G. glabra*, *G. inflata*, and *G. uralensis*, respectively.

Table 1

Summary of DNA Authentication Results for the 51 Acquired Licorice Samples

among the 51 acquired licorice samples	BC code	claimed species	DNA authentication results
11 misidentified <i>Glycyrrhiza</i> species, 21.5% of all samples	629	<i>G. inflata</i>	<i>G. uralensis</i>
	778	<i>G. inflata</i>	<i>G. uralensis</i>
	748	<i>G. inflata</i>	<i>G. glabra</i>
	716	<i>G. glabra</i>	<i>G. uralensis</i>
	750	<i>G. uralensis</i>	<i>G. uralensis</i> × <i>G. glabra</i> ^a
	752	<i>G. inflata</i>	<i>G. uralensis</i> × <i>G. inflata</i> ^a
	758	Gan-cao	<i>G. uralensis</i> × <i>G. inflata</i> ^a
	772a	<i>G. inflata</i>	<i>G. uralensis</i> × <i>G. inflata</i> ^a
	782c	<i>G. inflata</i>	<i>G. uralensis</i> × <i>G. inflata</i> ^a
	782d	<i>G. inflata</i>	<i>G. uralensis</i> ^a × <i>G. inflata</i>
	736	<i>G. glabra</i>	mixture: <i>G. inflata</i> and <i>G. uralensis</i> and other plant species
Eight samples acquired as “licorice” or <i>G. glabra</i> , but identified as mixtures of <i>Glycyrrhiza</i> species, 15.7% of all samples	625	<i>G. glabra</i>	mixture: <i>G. inflata</i> and <i>G. uralensis</i> or <i>G. glabra</i>
	686	<i>G. glabra</i>	mixture: <i>G. glabra</i> and <i>G. uralensis</i>
	741	Licorice	mixture: <i>G. inflata</i> with <i>G. uralensis</i> or <i>G. glabra</i>
	742	Licorice	mixture: 3 <i>G.</i> species are present
	744	Licorice	mixture 3 <i>G.</i> species are present
	733	<i>G. glabra</i>	mixture: <i>G. glabra</i> with <i>G. uralensis</i>
	734	<i>G. glabra</i>	mixture: <i>G. glabra</i> with potentially <i>G. inflata</i>
	736	<i>G. glabra</i>	mixture: <i>G. inflata</i> and <i>G. uralensis</i> and other plant species

^a Genotype identified from the chloroplast DNA and corresponding to the maternal species (**highlighted in bold**).

Table 2

DNA Authentication of 37 Samples Containing Single *Glycyrrhiza* Species

<i>Glycyrrhiza</i> species	code	DNA markers					TG ^a	claimed <i>G. species</i>	origin ^b
		ITS	<i>rbcL</i>	<i>matK</i>	<i>psbA-trnH</i>				
<i>uralensis</i>	629	I-3	R-2	M-1	T-4		TG9	<i>inflata</i>	China
<i>uralensis</i>	725							<i>uralensis</i>	China
<i>uralensis</i>	751							<i>uralensis</i>	Mongolia
<i>uralensis</i>	754							<i>uralensis</i>	Mongolia
<i>uralensis</i>	755							<i>uralensis</i>	Vietnam
<i>uralensis</i>	760							Gan-caio	China
<i>uralensis</i>	778	I-3	R-2	M-1	T-1		TG8	<i>inflata</i>	China
<i>uralensis</i>	716							<i>glabra</i>	China
<i>uralensis</i>	624							<i>uralensis</i>	China
<i>uralensis</i>	689							<i>uralensis</i>	China
<i>uralensis</i>	759							“Gan-caio”	China
<i>uralensis</i>	753	I-3	R-1	M-1	T-2		TG7	“Gan-caio”	Vietnam
<i>uralensis</i>	756							“Gan-caio”	China
<i>glabra</i>	289	I-2	R-1	M-1	T-2		TG3	<i>glabra</i>	unknown
<i>glabra</i>	694								Italy
<i>glabra</i>	693								Italy
<i>glabra</i>	595								IL, US
<i>glabra</i>	727								MA, US
<i>glabra</i>	728								MA, US
<i>glabra</i>	729								MA, US
<i>glabra</i>	730								WA, US
<i>glabra</i>	731								MA, US
<i>glabra</i>	749								Afghanistan
<i>glabra</i>	743							“licorice”	EU
<i>glabra</i>	745							“Gan-caio”	Iran
<i>glabra</i>	757							<i>glabra</i> var. <i>glabra</i>	China
<i>glabra</i>	747								Turkey, Nizip

<i>Glycyrrhiza</i> species	code	DNA markers				TG ^a	claimed <i>G. species</i>	origin ^b
		<i>ITS</i>	<i>rbcL</i>	<i>matK</i>	<i>psbA-trnH</i>			
<i>glabra</i>	748						<i>inflata</i>	China, Xinjiang
<i>glabra</i>	044	I-2	R-1	M-1	T-1	TG2 ^c	<i>glabra</i>	unknown
<i>glabra</i>	726	I-2	R-1	M-1	T-1 ^d	TG2 ^c		Egypt
<i>glabra</i>	732							Egypt
<i>glabra</i>	735							Egypt
<i>glabra</i>	695							Egypt
<i>glabra</i>	746						<i>glabra</i> var. <i>glabra</i>	Turkey, Konya
<i>inflata</i>	711	I-2	R-1	M-2	T-3	TG5	<i>inflata</i>	China
<i>inflata</i>	772b	I-2	R-1	M-2	T-3	TG5	<i>inflata</i>	China, Xinjiang
<i>inflata</i>	774	I-2	R-1	M-2	T-3	TG5	<i>inflata</i>	China, Xinjiang

^aTotal genotype (TG) determined by the composition of DNA markers.

^bGeographical origin of the sample when documented.

^cThe “TG2” genotype can correspond to either *G. inflata* or *G. glabra* according to Kondo’s paper.²⁷ “Gan-Cao”: Traditional Chinese medicine containing licorice.

^dWith a novel “CTCTT” deletion.

Table 3

DNA Authentication of Six *Glycyrrhiza* Hybrids (Root Sticks)

sample code	DNA markers				TG ^c	claimed <i>G. species</i>	identified <i>G. hybrids</i> ^d	origin
	<i>ITS</i> ^{a,b}	<i>rbcL</i>	<i>matK</i>	<i>psbA-trnH</i>				
BC750	I-3/I-2 = ADD (14%)	R-1	M-1	T-2	TG7 (cp) or TG3 (cp)	<i>uralensis</i>	<i>G. uralensis</i> × <i>G. glabra</i>	China, Xinjiang
BC752	I-2/I-3 = ADD (12%)	R-1	M-2	T-3	TG5 (cp)	<i>inflata</i>	<i>G. uralensis</i> × <i>G. inflata</i>	China, Xinjiang
BC758	I-3/I-2 = ADD (13%)	R-1	M-2	T-3	TG5 (cp)	“Gan-cao”	<i>G. uralensis</i> × <i>G. inflata</i>	China
BC772a	I-2/I-3 = ADD (86%)	R-1	M-2	T-3	TG5 (cp)	<i>inflata</i>	<i>G. uralensis</i> × <i>G. inflata</i>	China, Xinjiang
BC782c	I-2/I-3 = ADD (14%)	R-1	M-2	T-3	TG5 (cp)	<i>inflata</i>	<i>G. uralensis</i> × <i>G. inflata</i>	unknown
BC782d	I-2/I-3 = ADD (7%)	R-2	M-1	T-1	TG8 (cp)	<i>inflata</i>	<i>G. uralensis</i> × <i>G. inflata</i>	unknown

^aThe occurrence of both I-2 and I-3 sequences for the ITS DNA marker is called **ADD genotype** and corresponds to “**nucleotide additivities**” often observed in hybrids between two species.²⁷

^bThe ITS ADD genotype is given as a ratio, with the primary genotype **highlighted in bold**, and the average secondary peak height ratio is given in parentheses.

^cTotal genotype (TG) determined by the composition in DNA markers.

^dGenotype identified from the chloroplast DNA and corresponding to the putative maternal species (**highlighted in bold**).

Table 4
DNA Analysis of Commercial (Bulk Powders, Capsules) and in-House *Glycyrrhiza* Mixtures

BC code	TG(s) detected ^a	claimed ^b <i>Glycyrrhiza</i> species	DNA detected species	chemically identified species and relative abundance ^c
commercial				
625	TG5 (1°), TG7 (2°), possibly TG3 (2°)	<i>G. glabra</i>	<i>G. inflata</i> , <i>G. uralensis</i> , possibly <i>G. glabra</i>	<i>G. inflata</i> <i>G. uralensis</i> ≈ <i>G. glabra</i>
686	TG3 (1°), TG8 (2°)	<i>G. glabra</i>	<i>G. glabra</i> , <i>G. uralensis</i>	<i>G. glabra</i> > <i>G. uralensis</i>
741	TG5 (1°), TG7 (2°), possibly TG3 (2°)	"licorice"	possibly all three <i>G. species</i>	<i>G. uralensis</i> > <i>G. inflata</i>
742	TG5 (1°), TG8 and/or TG7 (2°), possibly TG3 and/or TG2 (2°)		possibly all three <i>G. species</i>	<i>G. glabra</i> > <i>G. inflata</i> ≈ <i>G. uralensis</i>
744	TG5 (1°), TG8 (2°), possibly TG3 (2°)		possibly all three <i>G. species</i>	<i>G. uralensis</i> ≈ <i>G. glabra</i> > <i>G. inflata</i>
733	TG3 (1°), TG9 (2°)	<i>G. glabra</i>	<i>G. glabra</i> , <i>G. uralensis</i>	<i>G. glabra</i> > <i>G. uralensis</i>
734	TG2 [†] (1°), TG3 (2°)		<i>G. glabra</i> , or <i>G. inflata</i>	mixture of <i>G. glabra</i> genotypes
736	TG5 (1°), TG8 (2°; 200 bp primers)		<i>G. inflata</i> , <i>G. uralensis</i>	<i>G. inflata</i>
Mix 1	TG3 and TG8	TG3 50%, TG8 50%	<i>G. glabra</i> , <i>G. uralensis</i>	<i>G. glabra</i> ≈ <i>G. uralensis</i>
Mix 2	TG3 and TG8	TG3 5%, TG8 95%	<i>G. glabra</i> , <i>G. uralensis</i>	<i>G. uralensis</i>
Mix 4	TG3 and TG8	TG3 1%, TG8 99%	<i>G. glabra</i> , <i>G. uralensis</i>	<i>G. uralensis</i>
Mix 5	TG3, TG5, and TG8	TG5 33%, TG8 33%, TG3 33%	all three <i>G. species</i>	<i>G. inflata</i> <i>G. glabra</i> ≈ <i>G. uralensis</i>
Mix 6	TG3, TG5, and TG8	TG5 70%, TG8 25%, TG3 5%	all three <i>G. species</i>	<i>G. inflata</i> > <i>G. glabra</i> ≈ <i>G. uralensis</i>
Mix 8	TG3, TG5, and TG8	TG5 9%, TG8 90%, TG3 1%	all three <i>G. species</i>	<i>G. uralensis</i> > <i>G. inflata</i>
Mix 10	TG3 and TG5	TG5 10%, TG3 90%	<i>G. glabra</i> , <i>G. inflata</i>	<i>G. glabra</i> ≫ <i>G. inflata</i>
in-house ^d				

^a Primary and secondary genotypes of mixtures were determined from secondary peak height ratio of markers/alleles that specify a particular species. Some total genotypes of commercial samples were considered "possibly" present since markers are shared across total genotypes.

^b For the commercial mixtures, the claimed species corresponds to the label indication.

^c Preponderant *Glycyrrhiza* species characterizing the mixture identified by both SIMCA and CDA classifications.

^d All in-house mixtures were prepared with DNA authenticated *Glycyrrhiza* species of known genotype. The percentage corresponds to the quantity of powder from each species mixed together prior to performing the DNA analysis.

Table 5

Characteristic Downfield NMR Data of Major Detectable Metabolites in *Glycyrrhiza* Extracts (DMSO-*d*₆ with 5% v/v D₂O and 2.05 mM of Calibrant, at 600 MHz)

compound	no.	position	δ_H , mult. (J in Hz) ^a
liquiritin	1F	H-3'/5'	7.052, XX' type (8.58/2.46/0.31)
liquiritin apioside	3F	H-3'/5'	7.028, XX' type (8.58/2.46/0.31)
liquiritin and liquiritin apioside	1/3F	H-5	7.637, d (8.67/0.16)
		H-2'/6'	7.446, AA' type (8.58/2.46/0.31)
		H-6	6.511, dd (8.67/2.21)
		H-8	6.354, d (2.21)
		H-2	5.507, dd (12.84/2.91)
liquiritin-isoliquiritin	1F, 1C	1''-O-Glc	4.936, d (7.46)
liquiritin apioside and isoliquiritin apioside	3F/3C	1''-O-Glc	4.957, d (7.75)
		1''-O-Api	5.328, d (1.20)
isoliquiritin and isoliquiritin apioside	1/2C	H-6'	8.166, dd (8.93/0.53)
		H-2/6	7.838, AA' type (8.70/2.04/0.18)
		H- β	7.830, d (15.44)
		H- α	7.740, d (15.44)
		H-3/5	7.063, XX' type (8.70/2.04/0.18)
licuraside	2C	H-3'	6.296, d (2.33)
		H-6'	8.240, d (9.02)
		H-3/5	6.848, XX' type (8.51/2.72/0.15)
		H-5'	6.598, dd (9.02/2.41)
liquiritigenin-7-O-aposylglucoside	2F	H-3'	6.547, d (2.41)
		H-2'/6'	7.340 AA' type (8.14/ 1.81/0.33)
liquiritigenin	4F	H-3'/5'	6.795 XX' type (8.14/1.81/0.33)
		H-2'/6'	7.332 AA' type (8.41/2.63/0.33)
		H-3'/5'	6.822, XX' type (8.41/2.63/0.33)
		H-6	6.486, dd (8.68/2.18)
glycyrrhizin	5	H-8	6.315, d (2.18)
glabridin	6	H-12	5.381, s
		H-6'	6.848, d (8.21)
		H-5	6.827, d (8.32)
		H-4''	6.574, d (9.89)
		H-3''	5.635, d (9.89)
		H-3'	6.313, d (2.42)
		H-6	6.268, d (8.21)
licochalcone A	7	H-5'	6.183, dd (8.32/2.42)
		H-2'/6'	7.949, AA' type (8.59/2.43/0.27)
		H- β	7.870, d (15.54)
		H- α	7.557, d (15.54)
		H-3	6.526, s

compound	no.	position	δ_{H} , mult. (<i>J</i> in Hz) ^a
HBMA	8	H-3'/5'	6.887, XX' type (8.59/2.43/0.27)
		H-6	7.526, s
		H-2''	6.221, dd (17.49/ 10.62)
		H-2'/6'	6.911, AA' type (8.33/2.65/0.32)
		H-3'/5'	6.574, XX' type (8.33/2.65/0.32)

^aExact values obtained through ¹H iterative full-spin analysis (HiFSA) using Perch NMR software (v.2011.1, PERCH Solutions Ltd., Kuopio, Finland).