Use of the psbA-trnH Region to Authenticate Medicinal Species of Fabaceae

Ting Gao, Xinye Ma, and Xunzhi Zhu

"Key Laboratory of Plant Biotechnology in Universities of Shandong Province, College of Life Sciences, Qingdao Agricultural University; Qingdao, Shandong 266109, P.R. China; Research Center of Chinese Herbal Resource Science and Engineering, Key Laboratory of Chinese Medicinal Resource from Lingnan, Ministry of Education, Guangzhou University of Chinese Medicine; Guangzhou, Guangdong 510006, P.R. China; and School of Biology and Chemical Engineering, Jiangsu University of Science and Technology; Zhenjiang, Jiangsu 212018, P.R. China.

Received August 3, 2013; accepted October 3, 2013; advance publication released online October 8, 2013

Fabaceae is a huge family that contains a large number of medicinal plants, many of which are commonly used in Chinese traditional medicine. However, traditional taxonomy has not been able to meet the complicated demands of species discrimination within Fabaceae. Thus, we employed a famous DNA barcode, the psbA-trnH region, to discriminate commonly used medicinal species of the family Fabaceae. Here, the psbA-trnH regions derived from 152 samples were amplified. These samples represented 104 Fabaceae medicinal species from 60 genera, including 25 authentic Fabaceae species listed in the Chinese pharmacopoeia and common adulterant species. The results indicate that the psbA-trnH region performed well in terms of its universality and high variability in length and composition. Species discriminative power analysis of the psbA-trnH region showed that 91.3% of species could be identified successfully by the BLAST1 method in conjunction with the nearest distance method. And, the species resolution rate of the TaxonGap method exceeded 93%. The results provide support for the use of the psbA-trnH plastid region as a sensitive marker to the authentication of Fabaceae medicinal plants.

Key words traditional medicine; Fabaceae; identification; psbA-trnH region

A wide range of medicinal plants are found within Fabaceae, third largest plant family on earth. The Pharmacopoeia of the People’s Republic of China states that 31 Fabaceae species spanning 20 genera have been officially designated as standard medical plants. These species possess important medicinal properties. For instance, the root of Astragalus membranaceus and Astragalus mongholicus, Radix Astragali, is one of the well-known traditional Chinese medicines (TCM), which regulates immune responses, and improves resistance to colds, allergies, infections and the flu. Sophora tonkinensis is used for the treatment of acute pharyngolaryngeal infections and sore throats. The herbal extract from Sophora flavescens has anti-viral activity against influenza. Taken together, members of the family Fabaceae have been reported to represent an important group of medicinal plants.

However, the adulteration of the traditional medicines products has been widely spread for a long time. The family Fabaceae is no exception in this problem of mis-identification. For example, the roots of several Indigofera species are used as Radix Sophorae Tonkinensis in the treatment of inflammatory diseases. The contents of the biologically active compounds of these adulterations are distinct from those of S. tonkinensis. Some poisonous adulterants may cause intoxication or even death. Thus, incorrect identification of these medicinal plants affects their appropriate usage, breach the safety of the medicinal market and the confidence of consumers. The classical morphological authentication approach was confronted with difficulties. Hence, a rapid and reliable method is required to correctly identify the plant species present within the product for the prevention of misuse and fits a current need in the market.

DNA barcoding is an aid to taxonomic identification that uses a short, standard DNA region for species discrimination. In a number of studies, psbA-trnH plastid region was found universally within plants and has been demonstrated to be effective in telling species apart. And, the psbA-trnH region has always been proposed as one of the popular DNA barcodes. However, DNA-based species identification using a large set of medicinal species of Fabaceae has not been carried out. Therefore, we examined the efficacy of using psbA-trnH in medicinal Fabaceae species determination across a wide taxon range.

MATERIALS AND METHODS

Plant Materials In this study, we employed the psbA-trnH barcode to identify 152 samples from 104 Fabaceae common medicinal species, their closely related species or adulterants (Supplementary Table S1). These plant materials include 49 samples belonging to 25 authentic species and 19 genera listed in the Chinese Pharmacopoeia.

This Sample Set Includes All Three Subfamilies of Fabaceae The plant samples were collected from large geographical areas in China and were identified by Prof. Yulin Lin of IMPLAD (the Institute of Medicinal Plant Development), Chinese Academy of Medicinal Sciences and deposited in the Herbarium IMPLAD, Beijing, China.

DNA Extraction, Amplification and Sequencing Genomic DNA was extracted according to the protocol associated with the Plant Genomic DNA Kit (Tiangen Biotech Co., China) from the collected silica gel-dried leaves or from commercially available medicinal materials purchased at the market. The psbA-trnH regions were amplified and sequenced from the genomic DNA. Polymerase chain reaction (PCR)
amplification of the psbA-trnH region was carried out in a Peltier Thermal Cycler PTC0200 (BioRad Lab, Inc., U.S.A.) using approximately 30 ng of genomic DNA as a template in a 25 µL reaction mixture (1×PCR buffer without MgCl₂, 2.0 mM MgCl₂, 0.2 mM of each deoxyribonucleotide triphosphate (dNTP), 0.1 µM of each primer (synthesized by Sangon Co., China), and 1.0 U of Taq DNA Polymerase (Biocolor BioScience & Technology Co., China). Universal primers and reaction conditions for the amplification of the psbA-trnH region were as previously described.7–9 (Table 1). PCR products were run on a 1.0% agarose gel in a 0.5×TBE buffer and purified using the TIANgel Midi Purification Kit (Tiangen Biotech Co., China). The purified PCR products were sequenced using an ABI 3730XL sequencer (Applied Biosystems Inc.) using primers used for amplification. The quality control of the generated sequence traces followed as Chen et al. used.13

Sequence Analysis Consensus sequences and contig generation were performed using CodonCode Aligner V 3.5.7 (CodonCode Co., U.S.A.). Genetic distances were computed using MEGA 4.0 according to the Kimura 2-Parameter (K2P) model.15 The average intra-specific distance, theta, coalescent depth, and theta were calculated to test the intra-specific variation using the K2P model.7,16 The average inter-specific distance, the minimum inter-specific distance, and theta primer were used to represent the inter-specific divergences.7,16,17 The discriminatory power of psbA-trnH sequences for Fabaceae species was calculated using three methods, BLAST1 method, the nearest distance method, and the TaxonGap method as described previously.7,18 In BLAST1 method, species determination was based on the best hit of the query sequence and an E-value for the match less than a cutoff value. In the nearest distance method, authentication was based on all pairwise genetic distances computed among the reference sequences, and between each query and each of the reference sequences. In the TaxonGap method, the minimum distance (defined as separability) between certain species and their closest neighbor and the maximum distance (defined as heterogeneity) within the species were compared using the TaxonGAP software (version 2.4.1).8,18,19 The distance between and within species was determined by a similarity matrix which was calculated using the program AlignX (Vector NTI Suite v 9, InforMax, North Bethesda, MD, U.S.A.) with an engine based on the Clustal W algorithm.

RESULTS AND DISCUSSION

PCR Efficiency and Sequencing Success of psbA-trnH As previously proposed,7–8,20–22 psbA-trnH can be applicable to a broad range of Fabaceae plant taxa. The efficiency of PCR amplification and sequencing of the psbA-trnH region obtained from the samples of Fabaceae were 98.9% and 77.4%, respectively. There was no difficulty in the psbA-trnH sequence alignment in our experiments. While, there were approximate 11.4% poor quality psbA-trnH sequences in our study, which mainly due to the presence of poly-A/T structures. Therefore, use of the psbA-trnH region as a standard DNA barcode requires the improved sequencing technology to get more high quality sequences.8,23

Nucleotide Sequence Variations of psbA-trnH The length of the psbA-trnH region examined in the Fabaceae species ranged from 249 bp to 515 bp. As previous studies reported, owing to its relatively short length (200–400 bp), the psbA-trnH region might be applicable for the authentication of a wide variety of plants, herbarium specimens, and commercially prepared crude drugs, the amplification of this region could be much easier, even without high quality DNA.12

The inter-specific percentages of nucleotide differences ranged from 0% (in 7 species) to 37.1% (in Cassia fistula, and Cassia obtusifolia), with an average of 7.2% across all taxa. The percent differences between individuals of the same species among the 152 Fabaceae sequences ranged from 0% (in 13 species) to 3.4% (in Lotus corniculatus), with an average of 0.6% across all taxa. In addition, six parameters were used to characterize the inter-specific versus intra-specific variation (Table 2). Many researchers considered psbA-trnH provided considerable differences between species.10–14 Our findings displayed the similar trend that the psbA-trnH region showed little intra-specific variation, and large inter-specific divergence, thereby it is an ideal DNA barcode in our dataset.

Species Discriminative Power of psbA-trnH Fabaceae plants have been previously identified using the psbA-trnH region. Some results obtained using the psbA-trnH regions were not satisfactory in certain genus of Fabaceae and it was concluded that the psbA-trnH region was not the proper authentication tool.20 While, more studies support psbA-trnH as a reliable barcode.25,26 Whereas, results derived from those previous studies with different small coverage (at the genus level) may not be readily extendable to diverse Fabaceae medicinal species. In our study, the analyses come from the inclusive sampling of Fabaceae medicinal species across multiple taxonomic groups.

The result of the TaxonGap method (Fig. 1) demonstrated that the psbA-trnH sequences generated from 97 Fabaceae species were specific enough to be separated from their neighbors, in other words, a considerable portion (93.3%) of the species sequences, had the between-species variation (if avail-

<table>
<thead>
<tr>
<th>Marker</th>
<th>Name of primers</th>
<th>Primer sequences 5′–3′</th>
<th>Reaction conditions</th>
</tr>
</thead>
<tbody>
<tr>
<td>psbA-trnH</td>
<td>fwd PA</td>
<td>GTTATGCATGAAACGTATGCTC</td>
<td>94°C 5 min, 94°C 1 min, 55°C 1 min, 72°C 1.5 min, 30 cycles, 72°C 7 min</td>
</tr>
<tr>
<td></td>
<td>rev TH</td>
<td>CGGCATGTTGGATTACAATCC</td>
<td></td>
</tr>
</tbody>
</table>

Table 2. Analysis of Inter-specific Divergence between Species and Intra-specific Variation in the psbA-trnH Sequences

<table>
<thead>
<tr>
<th>Marker</th>
<th>Dataset</th>
<th>All inter-specific distance</th>
<th>Theta prime</th>
<th>The minimum inter-specific distance</th>
<th>All intra-specific distance</th>
<th>Theta</th>
<th>Coalescent depth</th>
</tr>
</thead>
<tbody>
<tr>
<td>psbA-trnH</td>
<td></td>
<td>0.0719±0.1307</td>
<td>0.0477±0.0690</td>
<td>0.0293±0.0370</td>
<td>0.0065±0.0122</td>
<td>0.0089±0.0140</td>
<td>0.0104±0.0157</td>
</tr>
</tbody>
</table>
able), which was higher than the within-species diversity. The gene resolution rate of the \( \text{psbA-trnH} \) region was 100% at the genus level. And the results showed that 95 out of 104 species (91.3%) was accurately identified both by the BLAST1 and the nearest distance method. Therefore, the \( \text{psbA-trnH} \) regions revealed informative traits for differentiating among the common Fabaceae medicinal species. Our findings showed that the \( \text{psbA-trnH} \) region could not only discriminate 22 species

Fig. 1. Species Identification Capability of the \( \text{psbA-trnH} \) Intergenic Region among 104 Medicinal Fabaceae Species by Taxon Gap

The left panel shows the complete list of species used in this study. The right panel depicts the intra-species heterogeneity and inter-species separability for \( \text{psbA-trnH} \) as horizontal light grey and dark grey bars, respectively. The right panel also shows the names of the closest relatives identified by the similarity method.
of Fabaceae in the Chinese Pharmacopoeia, but also recognize 75 common medicinal species of Fabaceae, their sister species or adulterants. Therefore, the \textit{psbA-trnH} region can confidently serve as a potential DNA barcode for common Fabaceae medicinal species especially the species in the Chinese Pharmacopoeia.

Overall, in addition to traditional classification controversial species (\textit{Tadehagi triquetrum} vs. \textit{T. pseudotriquetrum} and \textit{Caragana rosea} vs. \textit{C. sinica}), \textit{27,28}) a small amount of species in our dataset could not be effectively identified (\textit{Caragana rosea} vs. \textit{Caragana sinica}, \textit{Glycyrrhiza glabra} vs. \textit{Glycyrrhiza inflata} and \textit{Glycyrrhiza uralensis}). Therefore, multi-locus barcode should be necessary to solve complex taxonomic problems in certain close related species.\textit{7,8,20–22})

Finally, we examined the feasibility of \textit{psbA-trnH} region applied to discriminate the herbal plants (Fabaceae) and their adulterant species. We might take an example of \textit{Radix Sophorae Tonkinensis}. The intra-specific variation among \textit{S. tonkinensis} samples examined here was relatively low (2.3%). In contrast, the inter-specific divergences among \textit{Radix Sophorae Tonkinensis} and its adulterant species, \textit{Sophora alopecuroides}, \textit{Indigofera macrophylla}, and \textit{Flemingia glutinosa} were relatively large (8.1–35.5%). Here, we determined that the \textit{psbA-trnH} region could distinguish between \textit{Radix Sophorae Tonkinensis} and its four common adulterants apart.

Another example is \textit{Radix Astragali}, which is becoming increasingly scarce. \textit{A. membranaceus} has been under the third-grade priority protection of the state in China. This scarcity resulted in the common fraudulent adulteration, \textit{Oxytropis coerulenta}, \textit{Medicago sativa}, \textit{Mellilotus suaveolens}, \textit{Caragana rosea}, \textit{Oxytropis coerulenta} species (\textbf{baceae medicinal species especially the species in the Chinese Pharmacopoeia, but also recognition of its adulterants. Therefore, the \textit{psbA-trnH} region can be built. And then, the availability of the \textit{psbA-trnH} sequences from the Fabaceae species can facilitate the development of biodiversity and phylogeny research and also the protection of endangered Fabaceae species.

**Acknowledgments**

This work was supported by the Start-up Foundation for Advanced Talents of Jiangsu University of Science and Technology under award (No. 635211204) and the Startup Foundation for Advanced Talents of Qingdao Agriculture University under award (No. 631313).

**REFERENCES**


