



DNA barcoding of populations of *Fallopia multiflora*, an indigenous herb in China

X.Q. Sun^{1*}, M.M. Bai^{1*}, H. Yao², J.L. Guo¹, M.M. Li¹ and Y.Y. Hang¹

¹Jiangsu Province Key Laboratory for Plant *Ex Situ* Conservation, Institute of Botany, Jiangsu Province and Chinese Academy of Sciences, Nanjing, China

²Institute of Medicinal Plant Development, Chinese Academy of Medical Sciences, Peking Union Medical College, Beijing, China

*These authors contributed equally to this study.

Corresponding author: Y.Y. Hang

E-mail: hangyueyu@yahoo.com.cn

Genet. Mol. Res. 12 (3): 4078-4089 (2013)

Received March 12, 2013

Accepted August 1, 2013

Published September 27, 2013

DOI <http://dx.doi.org/10.4238/2013.September.27.9>

ABSTRACT. *Fallopia multiflora*, locally known as Heshouwu, is one of the most important and widely used Chinese medicinal herbs. However, there is still considerable confusion concerning its different provenances. DNA barcoding is a recent aid to taxonomic identification and uses a short standardized DNA region to discriminate plant species. We assessed the applicability of 4 candidate DNA barcodes (*matK*, *rbcL*, *psbA-trnH*, and ITS2) to identify populations of *F. multiflora*. To our knowledge, this is the first attempt involving the plant kingdom to apply DNA barcoding at a level lower than species. Four DNA loci (*matK*, *rbcL*, *psbA-trnH*, and ITS2) of 105 samples, including the wild *F. multiflora* distributed in 17 provinces of China and 4 cultivated *F. multiflora* lines, were amplified by PCR and sequenced. The 4 loci were evaluated by PCR amplification for sequence quality, extent of genetic divergence, DNA barcoding gap, and the ability to discriminate between populations by BLAST1 and Nearest Distance. We found that *psbA-trnH* was the best barcode, with significant inter-population variability and best potential for identifying *F. multiflora*. The combination of loci

gave better performance for distinguishing populations than a single locus. We recommend using *matK* + *rbcL* + *psbA-trnH* + ITS2 or *psbA-trnH* alone for this species. This research demonstrates the utility of DNA barcoding for geoherbals identifications.

Key words: *Fallopia multiflora* (Thunb.) Harald.; DNA barcoding; *matK*; *rbcL*; *psbA-trnH*; ITS2

INTRODUCTION

Fallopia multiflora (Thunb.) Harald, belonging to the genus *Fallopia* Adanson of Polygonaceae, is a perennial herb with root hypertrophy and twining stems, and the roots and canes are used as traditional Chinese herbal medicines (Li et al., 2003a). Wild *F. multiflora* is distributed in regions south of the Yellow River in China, such as Anhui, Fujian, Guangdong, Gansu, Guangxi, Guizhou, Hubei, Henan, Hunan, Jiangsu, Jiangxi, Sichuan, Shandong, Shanghai, Shaanxi, Yunan, and Zhejiang. Cultivated *F. multiflora* Siheyihao and Shizhuyihao are distributed in Anhui and Chongqing, respectively.

“Tu Jing Ben Cao” of the Sung Dynasty cited the “Li Yuan Appendix” that “The quality of *F. multiflora* in Guangdong and Guangxi is the best”. Renshan Chen’s “Yao Wu Chu Chan Bian” states that “The quality of *F. multiflora* from Deqing of Guangdong is the best”. Modern research has also confirmed that the content of 2,3,5,4'-tetrahydroxystilbene-2-O- β -D-glucoside (Chen et al., 2000; Li et al., 2003b; Fu et al., 2006) and anthraquinones (Bai et al., 2001; Yan et al., 2007), the active components, are significantly higher in *F. multiflora* of Deqing than of other origins. Regional differences are an important factor affecting its quality since *F. multiflora* is widely distributed in China. Therefore, it is very important to identify the different provenances of *F. multiflora*.

Traditional identification for *F. multiflora* has been by morphological appearances, microscopic characteristics, physical and chemical compositions, soluble protein gel electrophoresis, infrared spectroscopy, molecular identification, etc. A statistical analysis of the characteristics of leaf shape of *F. multiflora* from different provenances indicated that the samples with closer geographical distance usually fall into one branch (Wei et al., 2010). Zhou et al. (2010) found that the characters of 2-dimensional infrared spectroscopy can distinguish different sources of *F. multiflora*. A comparable divergence in the internal transcribed spacer has been observed in *F. multiflora* from 10 provenances, with a total of 22 variable sites including 4, 16, and 2 sites in ITS1, ITS2, and 5.8S rDNA, respectively (Yan et al., 2008a). A study on *matK* and 18S rRNA of 15 wild populations and cultivated *F. multiflora* indicated that there were 12 variable sites in the *matK* locus among different populations and it was proposed that the phylogenetic relationship of populations was well correlated with their geographical distribution; on the contrary, only few variable sites were found in some of the wild populations, leading to the impossibility of discriminating *F. multiflora* from different provenances (Yan et al., 2008b).

DNA barcoding is a taxonomic method that uses a short genetic marker to achieve rapid and accurate identification (Hebert et al., 2003). With easy operation, repeatability, reliability, and many other advantages, it has become a hot topic in species identification and diversity studies in recent years (Schindel and Miller, 2005; Marshall, 2005). However, in land plants, the most suitable barcode has not yet been decided even though multiple attempts have

been made to investigate DNA regions for their potential as barcodes for all land plant species (Kress et al., 2005; Lahaye et al., 2008). Besides *matK* and *rbcL*, *psbA-trnH* and ITS, 2 loci with faster evolutionary rates, have been recommended as candidate barcodes for land plant at the Third International Barcode of Life Conference; and ITS2 has been also proposed as a potential barcode with large application. The evaluation of 10 loci in the mosses found that *rbcL*, *rpoCI*, *psbA-trnH*, *rps4*, and *trnL-trnF* could be used as candidate barcodes (Liu et al., 2010). The discrimination ability of ITS2 was tested in more than 6600 plant samples belonging to 4800 species from 753 distinct genera of algae, fungi, and higher plants, and the identification efficiency of ITS2 was found to be 92.7% at the species level. Therefore, ITS2 has been suggested to be a barcode for fungi and green plants (Chen et al., 2010). A recent study was conducted in 6286 sampled individuals representing 1757 species in 141 genera of 75 families (42 orders), and 4 candidate barcodes (*matK*, *rbcL*, *psbA-trnH*, and ITS) were assessed for their effectiveness and universality (China Plant BOL Group et al., 2011). These analyses indicated that 3 chloroplast loci showed high universality (87.1 to 92.7%), whereas ITS had a better performance (79%) in angiosperms than in gymnosperms; ITS exhibited the highest discrimination efficiency of all 4, and the combination of 4 loci could discriminate 69.9 to 79.1% of all species.

In the present study, 3 chloroplast loci (*matK*, *rbcL*, *psbA-trnH*) and a nuclear locus (ITS2) were assessed in 17 wild and 4 cultivated *F. multiflora* encompassing all distribution regions in China to screen the applicable DNA barcodes and establish a reliable discrimination method for *F. multiflora*. Investigation of the application of DNA barcode in the identification of geoherbals and provenance of *F. multiflora* will also establish a foundation for further breeding selection and germplasm preservation. Also, we discuss whether DNA barcoding could provide a specific discrimination method for samples of Deqing, the geo-authentic producing area of *F. multiflora*.

MATERIAL AND METHODS

Plant materials

The plant samples were collected from the main distribution areas in China, including 17 wild and 4 cultivated populations in 2010 (Table S1). All specimens were examined and identified by Professor Y.Y. Hang from the Institute of Botany, Jiangsu Province and Chinese Academy of Sciences. The voucher samples were deposited in the Herbarium of the Institute of Botany, Jiangsu Province and Chinese Academy of Sciences. Five individuals of each provenance were included for further analysis.

DNA extraction, amplification, and sequencing

Genomic DNA of all the materials was extracted from fresh or silica gel-dried leaves using a modified cetyltrimethylammonium bromide (CTAB) method (Yan and Zhao, 2007).

Primers were designed according to the *F. multiflora* sequences published in GenBank, using Primer Premier 5.0. The primers for *matK* were 3F_KIM: 5'-CGT ACA GTA CTT TTG TGT TTA CGA G-3' and 1R_KIM: 5'-ACC CAG TCC ATC TGG AAA TCT TGG TTC-3' (<http://barcoding.si.edu>). The primers for *rbcL* were 1F: 5'-ATG TCA CCA CAA ACA GAA AC-3' and 724R: 5'-TCG CAT GTA CCT GCA GTA GC-3' (Fay et al., 1997). The primers

for *psbA-trnH* were *psbAF*: 5'-GTT ATG CAT GAA CGT AAT GCT C-3' and *trnH2*: 5'-CGC GCA TGG TGG ATT CAC AAT CC-3' (Tate and Simpson, 2003). The primers for ITS2 were ITS5: 5'-GGA AGT AAA AGT CGT AAC AAG G-3' and ITS4: 5'-TCC TCC GCT TAT TGA TAT GC-3' (White et al., 1990); S2F: 5'-ATG CGA TAC TTG GTG TGA AT-3' and ITS2AS: 5'-TTA TTG ATA TGC TTA AAC TCA GCG GG-3' (Chen et al., 2010); 293S: 5'-CGA AAT GCG ATA CTT GGT GT-3' and 637AS: 5'-TGC TTA AAC TCA GCG GGT AG-3'.

Polymerase chain reaction (PCR) amplification of the 4 candidate barcodes was carried out in a Perkin Elmer GeneAmp PCR System 9600. Each 20- μ L reaction mixture contained 30 ng genomic DNA, 1X DNA polymerase buffer, 0.12 mM dNTPs, 0.3 μ M of each primer, and 1 U *Taq* DNA polymerase. PCR was performed with the thermal cycle as follows. *matK* and *psbA-trnH*: 94°C for 5 min initially; 30 cycles at 94°C for 40 s, annealing at 47/58°C for 40 s, and extension at 70°C for 45 s; and one final cycle at 72°C for 5 min. *rbcL*: 94°C for 3 min initially; 34 cycles at 94°C for 30 s, annealing at 57°C for 30 s, and extension at 72°C for 1 min; and one final cycle at 72°C for 5 min. ITS2: 94°C for 5 min initially; 38 cycles at 94°C for 30 s, annealing at 52°C for 30 s, and extension at 72°C for 45 s; and one final cycle at 72°C for 10 min. PCR products were examined electrophoretically using 0.8-1.2% agarose gels, and observed by the Gel Doc image analyzer.

Bidirectional sequencing of purified PCR amplification products was done by the Beijing Genomics Institute (BGI) using the amplification primers. The nucleotide sequence data of these four regions were deposited in GenBank ([Table S1](#)).

Sequence alignment and analysis

Sequences were automatically aligned and then manually adjusted; all genetic distances were also calculated by the MEGA 4.1 software. The method of Meyer and Paulay (2005) was used to analyze and compare the intra-population variation and inter-population divergence. Wilcoxon signed rank tests were used to analyze the variability between the different loci (Lahaye et al., 2008). The barcoding gap was evaluated by comparing intra- versus inter-population variability (Meyer and Paulay, 2005). The BLAST1 similarity search algorithm and the Nearest Distance method were used to investigate the identification success rate of the sequences (Ross et al., 2008). The traffic light approach was used to identify the combination of loci (Chase et al., 2005). Additionally, the sequence character-based method (Rach et al., 2008) was utilized with DnaSP (Rozas et al., 2003), and the information from each site was treated as a character to distinguish the populations from each other.

RESULTS

Sequence information

A total of 105 sequences for each candidate barcode were obtained from 17 wild and 4 cultivated populations of *F. multiflora*. The length of matrices ranged from 249 bp for ITS2 to 803 bp for *matK* (Table 1). For the *matK* matrix, the distribution of 12 informative sites was dispersive and sparse across the matrix, without indels ([Table S2](#)). For the *rbcL* matrix, the aligned sequence length was 666 bp, and only 1 informative site was found in this matrix ([Table S2](#)). For the *psbA-trnH* matrix, the aligned sequence length was 384 bp; the distribution

of 53 informative sites and 167 variable sites was dispersive and dense across the matrix, with multiple indels included ranging from 1 to 26 bp (Table S2). For the ITS2 matrix, the distribution of 74 informative sites and 83 variable sites was dispersive and dense across the matrix, with 8 indels included (Table S3).

Table 1. Sequence information of 4 candidate genes.

Marker	Sequence length (bp)	Alignment length (bp)	Conserved sites (bp)	Variable sites (bp)	Psim-informative sites (bp)
<i>matK</i>	804-844	803	791	12	12
<i>rbcL</i>	671-716	666	665	1	1
<i>psbA-trnH</i>	243-434	384	217	167	53
ITS2	251-352	249	166	83	74

In the 4 data sets, some populations had unique character states that differentiated them from other populations of *F. multiflora*. For example, Guizhou, Sichuan, and Hubei populations each had unique diagnostic sites in the *matK* data set (Position 281: G could act as the diagnostic site for Guizhou population; position 574: G or position 775: A for Sichuan population; position 612: G for Hubei population) (Table S2). There were populations that shared character states. According to the 47-bp region (1704-1750 bp) in *psbA-trnH* (Table S2), the 21 populations of *F. multiflora* could be divided into 3 groups: the first group consisted of the Fujian, Guangdong, Hubei, Hunan, Jiangxi, Sichuan, Zhejiang, Guangxi, Yunnan, Guangdong 1, Guangdong 2 and Shizhuyihao populations (exhibiting “TATGGCCTCTTTGGTGT/GTTTG/T/AGTTGGGTGGACTTTTTTTTGATTCAT” sequence state); the second group was the Guizhou population with 16 distinct variable sites compared to the first group; and the 1704-1750-bp region was entirely missing in the last group including Anhui, Gansu, Henan, Jiangsu, Shaanxi, Shandong, Shanghai, and Siheyihao. Furthermore, the first group could be divided into 3 sub-groups according to the base differences at position 1724: the first sub-group included Guangdong, Hunan, Guangxi, Guangdong1, Guangdong2, and Shizhuyihao (position 1724: T); the second sub-group consisted of Fujian, Hubei, Jiangxi, Zhejiang (position 1724: G), while Sichuan and Yunnan exhibited A at this position in the third sub-group. In addition, the Hubei population was different from other populations as the insertion (position 1561-1567) acted as diagnostic sites.

Intra-population variation and inter-population divergence

The intra-population variation, inter-population divergence, maximum intra-population variation and minimum inter-population divergence of the 4 candidate barcodes (*matK*, *rbcL*, *psbA-trnH*, ITS2) and their 11 combinations (*matK* + *rbcL*, *matK* + *psbA-trnH*, *matK* + ITS2, *rbcL* + *psbA-trnH*, *rbcL* + ITS2, *psbA-trnH* + ITS2, *matK* + *rbcL* + *psbA-trnH*, *matK* + *rbcL* + ITS2, *matK* + *psbA-trnH* + ITS2, *rbcL* + *psbA-trnH* + ITS2, *matK* + *rbcL* + *psbA-trnH* + ITS2) were estimated using 6 parameters (Meyer and Paulay, 2005). The results showed that *psbA-trnH* had the highest intra- and inter-population divergence, followed by ITS2 and *matK*, with *rbcL* having the lowest (Table 2). The mean inter-population distances of 4 single loci and 11 combinations of loci were all higher than the mean intra-population distances. Furthermore, the minimum inter-population divergence of other single loci and combination of loci were greater than the corresponding maximum intra-population variation except for *matK*, *rbcL*, and *matK* + *rbcL*.

Table 2. Measures of intra- and inter-population divergence and identification efficiency of the potential and combined barcodes.

	<i>matK</i>	<i>rbcL</i>	<i>psbA-trnH</i>	ITS2	<i>matK + rbcL</i>	<i>matK + psbA-trnH</i>	<i>matK + ITS2</i>	<i>rbcL + psbA-trnH</i>
All intra-population distances	0.0003 ± 0.0008	0.0001 ± 0.0004	0.0017 ± 0.0088	0.0006 ± 0.0020	0.0002 ± 0.0006	0.0004 ± 0.0011	0.0005 ± 0.0011	0.0006 ± 0.0013
Theta	0.0003 ± 0.0006	0.0001 ± 0.0003	0.0017 ± 0.0055	0.0006 ± 0.0015	0.0002 ± 0.0004	0.0004 ± 0.0007	0.0005 ± 0.0007	0.0006 ± 0.0008
Coalescent depth	0.0006 ± 0.0014	0.0003 ± 0.0006	0.0042 ± 0.0138	0.0010 ± 0.0026	0.0005 ± 0.0010	0.0010 ± 0.0017	0.0014 ± 0.0019	0.0014 ± 0.0017
All inter-population distances	0.0022 ± 0.0017	0.0007 ± 0.0007	0.0442 ± 0.1003	0.0405 ± 0.0894	0.0017 ± 0.0010	0.0161 ± 0.0351	0.0117 ± 0.0198	0.0121 ± 0.0249
Theta prime	0.0020 ± 0.0011	0.0007 ± 0.0003	0.0443 ± 0.0753	0.0380 ± 0.0665	0.0016 ± 0.0006	0.0171 ± 0.0265	0.0111 ± 0.0144	0.0126 ± 0.0184
Minimum inter-population distance	0.0004 ± 0.0010	0.0000 ± 0.0000	0.0148 ± 0.0611	0.0206 ± 0.0678	0.0003 ± 0.0006	0.0037 ± 0.0119	0.0048 ± 0.0142	0.0035 ± 0.0130
Relative identification efficiency	32.4	9.5	59.0	22.9	100	88.6	94.3	100
BLAST1 nearest distance	18.1	0	14.3	13.3	24.8	28.6	34.3	23.8
	<i>rbcL + ITS2</i>	<i>psbA-trnH + ITS2</i>	<i>matK + rbcL + psbA-trnH</i>	<i>matK + rbcL + ITS2</i>	<i>matK + psbA-trnH + ITS2</i>	<i>rbcL + psbA-trnH + ITS2</i>	<i>matK + rbcL + psbA-trnH + ITS2</i>	
All intra-population distances	0.0006 ± 0.0013	0.0026 ± 0.0064	0.0004 ± 0.0010	0.0003 ± 0.0006	0.0020 ± 0.0048	0.0017 ± 0.0032	0.0006 ± 0.0011	
Theta	0.0006 ± 0.0010	0.0026 ± 0.0046	0.0004 ± 0.0006	0.0003 ± 0.0004	0.0020 ± 0.0034	0.0017 ± 0.0021	0.0006 ± 0.0007	
Coalescent depth	0.0010 ± 0.0016	0.0060 ± 0.0099	0.0009 ± 0.0014	0.0007 ± 0.0009	0.0045 ± 0.0069	0.0043 ± 0.0051	0.0013 ± 0.0016	
All inter-population distances	0.0112 ± 0.0210	0.0488 ± 0.0921	0.0087 ± 0.0147	0.0073 ± 0.0117	0.0249 ± 0.0341	0.0220 ± 0.0371	0.0145 ± 0.0221	
Theta prime	0.0106 ± 0.0156	0.0478 ± 0.0690	0.0089 ± 0.0109	0.0069 ± 0.0086	0.0258 ± 0.0250	0.0221 ± 0.0277	0.0142 ± 0.0164	
Minimum inter-population distance	0.0049 ± 0.0155	0.0180 ± 0.0637	0.0026 ± 0.0081	0.0031 ± 0.0086	0.0081 ± 0.0233	0.0078 ± 0.0251	0.0049 ± 0.0142	
Relative identification efficiency	89.5	100	100	100	100	100	100	
BLAST1 nearest distance	15.2	36.2	37.1	38.1	46.7	48.6	59.0	

Statistical comparison of divergence

The Wilcoxon signed rank tests indicated that the intra-population variation ranking from highest to lowest were *psbA-trnH*, *matK*, and *rbcL*, while there was no significant difference for *psbA-trnH* versus ITS2 and ITS2 versus *matK* (Table 3). The inter-population divergence of *psbA-trnH* was the highest, followed by ITS2, and again *matK*, and *rbcL* showing the lowest.

Table 3. Wilcoxon signed rank test of the intra- and inter-population divergence among the 4 loci.

Intra-population			
W+	W-	Correlation between the variation of intra-population N P	Result
<i>psbA-trnH</i>	<i>rbcL</i>	W+ = 243; W- = 10; n = 210; P ≤ 1.50E-4	<i>psbA-trnH</i> > <i>rbcL</i>
<i>psbA-trnH</i>	<i>matK</i>	W+ = 255; W- = 70; n = 210; P ≤ 0.013	<i>psbA-trnH</i> > <i>matK</i>
<i>psbA-trnH</i>	ITS2	W+ = 336; W- = 225; n = 210; P ≤ 0.321	<i>psbA-trnH</i> =ITS2
<i>rbcL</i>	<i>matK</i>	W+ = 10; W- = 221; n = 210; P ≤ 2.29E-4	<i>rbcL</i> < <i>matK</i>
<i>rbcL</i>	ITS2	W+ = 171; W- = 495; n = 210; P ≤ 0.011	<i>rbcL</i> <ITS2
<i>matK</i>	ITS2	W+ = 261; W- = 519; n = 210; P ≤ 0.072	<i>matK</i> =ITS2
Inter-population			
W+	W-	Correlation between the variation of intra-population N P	Result
<i>psbA-trnH</i>	<i>rbcL</i>	W+ = 9081008; W- = 20503; n = 5250; P ≤ 0.0	<i>psbA-trnH</i> >> <i>rbcL</i>
<i>psbA-trnH</i>	<i>matK</i>	W+ = 9207683; W- = 212287; n = 5250; P ≤ 0.0	<i>psbA-trnH</i> >> <i>matK</i>
<i>psbA-trnH</i>	ITS2	W+ = 6384166; W- = 2283200; n = 5250; P ≤ 4.77E-154	<i>psbA-trnH</i> >ITS2
<i>rbcL</i>	<i>matK</i>	W+ = 895373; W- = 8295955; n = 5250; P ≤ 0.0	<i>rbcL</i> << <i>matK</i>
<i>rbcL</i>	ITS2	W+ = 1611910; W- = 4691115; n = 5250; P ≤ 3.03E-140	<i>rbcL</i> <ITS2
<i>matK</i>	ITS2	W+ = 3572832; W- = 5868853; n = 5250; P ≤ 8.04E-44	<i>matK</i> <ITS2

DNA barcoding gap assessment

The inter-population divergence of an ideal DNA barcode should be significantly larger than the intra-population variation, and there were significant differences between the 2, forming an apparent intergenic region, namely barcoding gap (Meyer and Paulay, 2005; Lahaye et al., 2008). In the investigation of the distributions of intra- versus inter-population variation of the 4 candidate DNA barcodes and 11 combinations of loci (Figure 1) we did not find any clear barcoding gaps as typical of *COI* in animals, although all the loci had a relatively separation between intra- and inter-population variation. Meanwhile, the significance between intra- and inter-population distances was evaluated by Wilcoxon 2-sample tests for each locus. The intra-population distance was found to be smaller than the inter-population distance, with the highest significances found in *matK* (Wilcoxon 2-sample test, P ≤ 3.59E-72), followed by *psbA-trnH* (Wilcoxon 2-sample test, P ≤ 7.13E-69) (Table 4).

Applicability for population discrimination

The BLAST1 similarity search algorithm and the nearest distance method (Ross et al., 2008) were used to assess the identification efficiency of the 4 single loci simultaneously, and the identification efficiency of 11 combinations was evaluated by the traffic light method (traffic light approach) (Chase et al., 2005). The results showed that the efficiency of suc-

successful identification using *psbA-trnH* was the highest in the 4 single loci, which was 59.0% (BLAST1) and 14.3% (nearest distance) (Table 2). The rates of successful identification using *matK*, *rbcL*, and ITS2 were all less than 50% (32.4, 9.5, and 22.9%, respectively) by BLAST1 and not suitable for identification of *F. multiflora* below the species level. For multi-locus barcodes, with exception of *matK + psbA-trnH*, *matK + ITS2*, and *rbcL + ITS2*, in which the efficiency of successful identification was 88.6 (BLAST1), 94.0 (BLAST1), and 89.5% (BLAST1), respectively, the efficiency rates of the locus combinations were 100% (BLAST1).

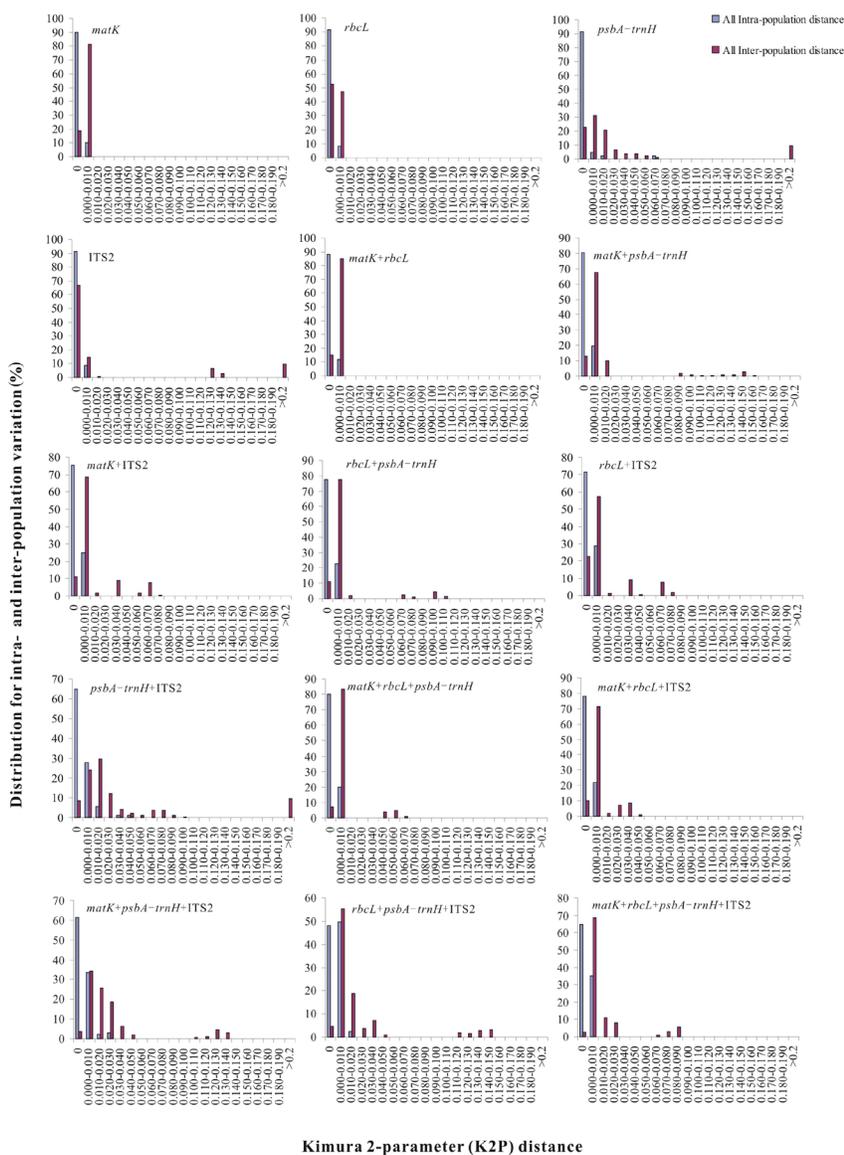


Figure 1. Relative distribution of intra-population variation and inter-population divergence.

Table 4. Wilcoxon 2-sample test based on inter-population versus intra-population specific Kimura 2-parameter distances of each barcode.

Loci	No. of intra-population distances	No. of inter-population distances	P value
<i>psbA-trnH</i>	183698.5	14724831.5	P ≤ 7.13E-69
<i>rbcL</i>	349655	14558875	P ≤ 1.44E-27
<i>matK</i>	172915.5	14735614.5	P ≤ 3.59E-72
ITS2	426502	14482028	P ≤ 3.18E-15

DISCUSSION

DNA barcoding screening standards and identification of each sequence in different populations of *F. multiflora*

An ideal DNA barcode should have high PCR amplification efficiency and separate distribution of inter-population differentiation versus intra-population variation, along with conserved regions for designing of universal primers (CBOL Plant Working Group, 2009).

PsbA-trnH is one of the fastest evolving intergenic spacer regions in the chloroplast genome. Both ends of the intergenic spacer, the *psbA* and *trnH* genes, are relatively conservative, which is beneficial to designing universal primers. The existing literature suggests that the *psbA-trnH* intergenic spacer could effectively identify 18 species of Polygonaceae (Song et al., 2009), 71 species of Labiatae (Han et al., 2009), and 19 species of *Aconitum* (He et al., 2010). For *F. multiflora*, *psbA-trnH* dominated over the other 3 single loci in sequence characteristics, intra- and inter-population divergence, barcoding gap, and identification efficiency of populations, establishing *psbA-trnH* as an ideal DNA barcode.

matK, one of the recommended candidate barcodes, has a fast evolutionary rate (Lahaye et al., 2008). The argument against *matK* as a barcode is its poor performance with regard to primer universality. The primer set of 3F_KIM and 1R_KIM (<http://barcoding.si.edu>) was successfully utilized in amplification of *F. multiflora*, while there was no distinct barcoding gap for *matK*, making it unsuitable for identifying *F. multiflora* populations.

rbcL has multiple advantages of good universality, easy amplification and sequence alignment. Newmaster et al. (2006) recommended it as an ideal candidate DNA barcode. However, the variation of *rbcL* was found primarily above the species level, not suitable for the species and lower (Kress and Erickson, 2007; Lahaye et al., 2008). *rbcL* showed the smallest inter-population differentiation among the 4 single loci of *F. multiflora*, and the minimum inter-population divergence was less than the maximum intra-population variation, while the identification success rate was only 9.5%. Therefore, *rbcL* was not suitable for DNA barcoding of *F. multiflora*.

In *F. multiflora*, ITS2 had difficulties in amplification and sequencing due to its multiple copies. Three pairs of primers were used in the amplification of *F. multiflora*, yet the identification success efficiency was only 22.9%. The intra- and inter-population genetic variation of ITS2 overlapped largely, showing no distinct barcoding gap, with an inter-population genetic distance of zero accounting for 66.571% of the total samples, which led to difficulty in discrimination between different populations, making it unsuitable for DNA barcoding of *F. multiflora*.

Multi-locus barcodes always exhibit better performance of discriminations than a single locus (CBOL Plant Working Group, 2009). In this study, when the 2 or 3 barcodes

were combined for further analysis, a larger percentage of the populations of *F. multiflora* were discriminated than when using a single locus. The results showed that the identification rate of *matK* + *rbcL* + *psbA-trnH* + ITS2 was 100 and 59% by BLAST1 and nearest distance, respectively. Thus, the 4 combined loci were better than the 2 or 3 loci.

In conclusion, although we did not find clear barcoding gaps in the distribution of the 4 loci tested (Figure 1), such as with *COI* in animals (Meyer and Paulay, 2005), the intra- and inter-population divergence in each locus was significantly different (Table 4). Suitable barcodes (*matK* + *rbcL* + *psbA-trnH* + ITS2 or *psbA-trnH* alone) were established in this study.

Identification of *psbA-trnH* in different populations of *F. multiflora*

Currently, DNA barcode technology is widely used in the identification of medicinal plants, especially for identification between species, subspecies or populations, which are difficult to identify by traditional morphological methods. It is a strong complement of traditional morphological taxonomy, building a new classification system to achieve the combination of sequence variation information and existing morphological taxonomy (Will et al., 2005; Hajibabaei et al., 2007).

The morphology of different populations of *F. multiflora* is basically the same except the Guizhou population, whose plants have a greater extent of white macula on the midvein and lateral nerve, shallower sinus, shorter length of inflorescence, smaller ratio of length to breadth, shorter persistent perianth, and later anthesis. Other populations could not be distinguished on the basis of morphological characteristics, while they could be effectively distinguished by variances in *psbA-trnH* between populations. Therefore, barcodes not only provides specific molecular markers for different populations of *F. multiflora*, but also compensate for the inadequacies of morphological identification in the determination of *F. multiflora* provenances, along with the establishment of the basis for further breeding selection and germplasm preservation.

The role of DNA barcoding in the identification of *F. multiflora* geoherbalsm

The good quality of a geoherb is formed on the long-term effects of genotype and environment. Specific genes produce enzymes regulating the synthesis of active ingredients such as secondary metabolites, and thus, a particular gene is the key to the formation of geoherbalsm (Li et al., 2001). It has been reported that *F. multiflora* from Deqing, the geo-authentic producing area, could be identified by the specific sites of ITS1 (Yan et al., 2008a). However, in this study, *matK*, *rbcL*, *psbA-trnH*, and ITS2 of Deqing were consistent with the other populations, so the geoherbalsm of *F. multiflora* could not be identified here. Difficulty in the identification of geoherbalsm has been observed in other medicinal plants too such as *Andrographis paniculata*, *Achyranthes bidentata*, and so on (Wang et al., 2004; Li et al., 2007). One reason could be the long-term harvesting and cultivation of the geoherbs, leading to confusion of medicinal plants from different provenances, and the other reason lies in the 4 candidate DNA barcodes themselves, so maybe we need to choose other candidate barcodes with faster evolutionary rates in the identification of geoherbalsm.

On the other hand, in the 3 populations of Deqing, although *matK*, *rbcL*, and ITS2 of all individuals of the wild population (Guangdong) and 2 cultivated populations (Guangdong

1 and 2) were exactly the same, there was one diagnostic site (position 1769: A for wild or missing for cultivated) between the wild and 2 cultivated populations, which could effectively distinguish the wild and cultivated *F. multiflora* from Deqing.

ACKNOWLEDGMENTS

Research supported by the DNA Testing Technology Research Project of DUS, Ministry of Agriculture (#200903008-02-05).

[Supplementary material](#)

REFERENCES

- Bai Y, Huang LM, Wu FH and Gu XL (2001). Determination of anthraquinone in *Polygonum multiflorum* from Deqing. *J. Guangdong Coll. Pharm.* 17: 40-41.
- CBOL Plant Working Group (2009). A DNA barcode for land plants. *Proc. Nat. Acad. Sci. U. S. A.* 106: 12794-12797.
- Chase MW, Salamin N, Wilkinson M, Dunwell JM, et al. (2005). Land plants and DNA barcodes: short-term and long-term goals. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 360: 1889-1895.
- Chen WS, Chai YF, Zhang WD, Yuan Y, et al. (2000). Diversity of chemical composition and quality in *Polygonum multiflorum* from different regions. *J. Pharmaceut. Pract.* 18: 344-345.
- Chen S, Yao H, Han J, Liu C, et al. (2010). Validation of the ITS2 region as a novel DNA barcode for identifying medicinal plant species. *PLoS One* 5: e8613.
- China Plant BOL Group, Li DZ, Gao LM, Li HT, et al. (2011). Comparative analysis of a large dataset indicates that internal transcribed spacer (ITS) should be incorporated into the core barcode for seed plants. *Proc. Nat. Acad. Sci.* 108: 19641-19646.
- Fay MF, Swensen SM and Chase MW (1997). Taxonomic affinities of *Medusagyne oppositifolia* (Medusagynaceae). *Kew Bull.* 52: 111-120.
- Fu J, Yan HJ, Lang CQ and Lang JZ (2006). Quality evaluation for *Radix Polygoni Multiflori* from different habitats. *J. Guangdong Coll. Pharm.* 22: 253-264.
- Hajibabaei M, Singer GA, Hebert PD and Hickey DA (2007). DNA barcoding: how it complements taxonomy, molecular phylogenetics and population genetics. *Trends Genet.* 23: 167-172.
- Han J, Shi LC, Yao H, Song JY, et al. (2009). Testing potential DNA barcoding regions in the Labiatae medicinal plants. *Planta Med.* 75: 407.
- He J, Wong KL, Shaw PC, Wang H, et al. (2010). Identification of the medicinal plants in *Aconitum* L. by DNA barcoding technique. *Planta Med.* 76: 1622-1628.
- Hebert PDN, Cywinska A, Ball SL and deWaard JR (2003). Biological identifications through DNA barcodes. *Proc. Biol. Sci.* 270: 313-321.
- Kress WJ and Erickson DL (2007). A two-locus global DNA barcode for land plants: the coding *rbcL* gene complements the non-coding *trnH-psbA* spacer region. *PLoS One* 2: e508.
- Kress WJ, Wurdack KJ, Zimmer EA, Weigt LA, et al. (2005). Use of DNA barcodes to identify flowering plants. *Proc. Natl. Acad. Sci. U. S. A.* 102: 8369-8374.
- Lahaye R, van der Bank M, Bogarin D, Warner J, et al. (2008). DNA barcoding the floras of biodiversity hotspots. *Proc. Natl. Acad. Sci. U. S. A.* 105: 2923-2928.
- Li AJ, Bao BJ, Alisa EGB, McNeill J, et al. (2003a). Flora of China. Vol. 5. Science Press, Beijing, 277-350.
- Li JP, Wang PX, Cao LY, Liang RY, et al. (2007). Feasibility of the DNA sequence to internal transcribed spacer regions of 5S rRNA gene to the genuineness of *Andrographis paniculata* (Burm. F.) Ness. *Central South Pharm.* 5: 438-440.
- Li P, Cai ZH and Xing JB (2001). Preliminary attempt to identify geoherbalsim of *Flos Lonicerae* by sequence divergence of 5S rRNA gene spacer region. *Chin. Trad. Herbal Drugs* 32: 834-837.
- Li Z, Hang YY and Zhou YF (2003b). Germplasm appraisements of *Polygonum multiflorum* Thunb. from main producing areas in China by content of tetrahydroxystilbene. *Chem. Ind. For. Prod.* 23: 37-41.
- Liu Y, Yan HF, Cao T and Ge XJ (2010). Evaluation of 10 plant barcodes in Bryophyta (Mosses). *J. Syst. Evol.* 48: 36-46.
- Marshall E (2005). Taxonomy. Will DNA bar codes breathe life into classification? *Science* 307: 1037.

- Meyer CP and Paulay G (2005). DNA barcoding: Error rates based on comprehensive sampling. *PLoS Biol.* 3: 2229-2238.
- Newmaster SG, Fazekas AJ and Ragupathy S (2006). DNA barcoding in land plants: evaluation of *rbcL* in a multigene tiered approach. *Can. J. Bot.* 84: 335-341.
- Rach J, Desalle R, Sarkar IN, Schierwater B, et al. (2008). Character-based DNA barcoding allows discrimination of genera, species and populations in Odonata. *Proc. Biol. Sci.* 275: 237-247.
- Ross HA, Murugan S and Li WLS (2008). Testing the reliability of genetic methods of species identification via simulation. *Syst. Biol.* 57: 216-230.
- Rozas J, Sánchez-DelBarrio JC, Messeguer X and Rozas R (2003). DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* 19: 2496-2497.
- Schindel DE and Miller SE (2005). DNA barcoding a useful tool for taxonomists. *Nature* 435: 17.
- Song J, Yao H, Li Y, Li X, et al. (2009). Authentication of the family Polygonaceae in Chinese pharmacopoeia by DNA barcoding technique. *J. Ethnopharmacol.* 124: 434-439.
- Tate JA and Simpson BB (2003). Paraphyly of *Tarasa* (Malvaceae) and diverse origins of the polyploid species. *Syst. Bot.* 28: 723-737.
- Wang SM, Liang SW, Zhou KY, Liu ZQ, et al. (2004). Ribosomal rDNA ITS sequence analysis of root of *Achyranthes bidentata*. *Chin. Tradit. Herbal Drugs* 35: 559-562.
- Wei YM, Wang LH, Cao FL, Wei SQ, et al. (2010). Variation and cluster analysis on leaf characters from different provenance sources of *Polygonum multiflorum* Thunb. *J. Anhui Agric. Sci.* 11: 94-98.
- White TJ, Bruns T, Lee S and Taylor J (1990). Amplification and Direct Sequencing of Fungal Ribosomal RNA Genes for Phylogenetics. In: PCR Protocols: A Guide to Methods and Applications (Innis MA, Gelfand DH, Sninsky JJ and White TJ, eds.). Academic Press, San Diego, 315-322.
- Will KW, Mishler BD and Wheeler QD (2005). The perils of DNA barcoding and the need for integrative taxonomy. *Syst. Biol.* 54: 844-851.
- Yan HJ, Fu J, Liang CQ and Liang JZ (2007). Determination of anthraquinone in *Polygonum multiflorum* from different habitats. *Chin. Tradit. Patent Med.* 29: 1023-1026.
- Yan P and Zhao SJ (2007). Genomic DNA extraction from the *Polygonum multiflorum* thunb. *Lishizhen Med. Mater. Med. Res.* 18: 607-609.
- Yan HJ, Fang ZJ and Yu SX (2008a). ITS sequences and phylogenetic relationships of *Fallopia multiflora* from different areas based on evidences. *Acta Bot. Boreal.-Occident. Sin.* 28: 922-927.
- Yan P, Pang QH, Jiao XW, Zhao X, et al. (2008b). Genetic variation and identification of cultivated *Fallopia multiflora* and its wild relatives by using chloroplast *matK* and 18S rRNA gene sequences. *Planta Med.* 74: 1504-1509.
- Zhou HB, Fang ZJ, Yang LW, Mo JL, et al. (2010). The identification of *Radix Polygoni Multiflori* from by IR spectroscopy. *Asia-Pacific Tradit. Med.* 6: 26-27.