Phytomedicine 21 (2014) 1767–1773

Contents lists available at ScienceDirect

Phytomedicine

journal homepage: www.elsevier.de/phymed

DNA barcoding Chinese medicinal *Bupleurum*

Zhi Chao[a,*], Weiping Zeng[a], Jing Liao[a,b], Li Liu[a,c], Zhenbiao Liang[a,d], Xiaolei Li[a]

[a] School of Traditional Chinese Medicine, Southern Medical University, Guangzhou 510515, China
[b] Beijing Royal Intergrative Medicine Hospital, Beijing 102209, China
[c] Research & Development Center, Infinitus (China) Company Limited, Guangzhou 510623, China
[d] Department of Pharmacy, Zhongshan People's Hospital, Zhongshan 528403, China

ARTICLE INFO

Article history:
Received 21 July 2014
Accepted 3 September 2014

Keywords:
DNA barcoding
Bupleurum
Bupleuri radix
ITS2
Identification
Traditional Chinese medicine

ABSTRACT

We tested 4 markers, namely nuclear internal transcribed spacer 2 (ITS2), psbA-trnH, matK, and rbcL, to evaluate these candidate DNA barcodes for distinguishing Bupleuri radix (Chaihu) from its adulterants. 51 plant samples of *Bupleurum* representing 19 species were collected from different areas in China. Amplification and sequencing were attempted for all the 4 candidate barcode regions, whose validity was assessed in terms of the success rate of PCR amplification and sequencing, differential intra- and inter-specific divergences, DNA barcoding gap and the ability to discriminate species. The results showed that ITS2 had the best performance in identifying *Bupleurum* with an identification efficiency of 73.68%, which, after combining with psbA-trnH, increased to 83.33%. We further evaluated the efficiency of ITS2 for discriminating the species of *Bupleurum* using a large database from GenBank, which archived data of 223 samples from 74 species, and ITS2 successfully discriminated 64.13% of the samples at the species level. In conclusion, the ITS2 can serve as a potentially useful barcode for *Bupleurum* species, with psbA-trnH as a supplementary locus.

Introduction

Chaihu (Bupleuri radix), the dried roots of *Bupleurum chinense* DC. and *B. scorzonerifolium* Willd., is a famous Traditional Chinese medicinal herb with a history of medical use for more than 2000 years. In Traditional Chinese Medicine (TCM), Chaihu is thought to regulate the exterior and interior metabolisms, disperse evil heat from superficies, sooth the liver, and promote *yang* and *qi* (representing "life energy" or "life force" in TCM theory). Deficiency in *yang* and *qi* leads to insufficiency in the body's functionality resulting in a lowered metabolism, and malfunctioning of the entire body or certain organs. Symptoms like low spirit, tired and weak, reduced appetite, chronic diarrhea, prolapses (such as proctoptosis, gastroptosis, hysteroptosis), are all manifestations of deficiency and sinking of *yang* and *qi*. Chaihu is an important component in prescriptions to treat such symptoms through promoting *yang* and *qi*. According to Chinese Pharmacopoeia, Chaihu is effective in treating common cold with fever, influenza, hepatitis, malaria, menoxenia, hysteroptosis and proctoptosis. (National Pharmacopoeia Committee, 2010).

Bupleurum L. (Apiaceae) comprises approximately 150 species worldwide. In China, about 44 species, 17 varieties and 7 forma have been identified, distributed widely across the country in 27 provinces (municipalities or autonomous regions) (She and Watson 2005; Pan 2006). Although only the 2 species mentioned above are officially approved for clinical use in Chinese Pharmacopoeia, as many as 25 species, 6 varieties, and 1 form of *Bupleurum* are found in use for medicinal purpose, including *B. longiradiatum* Turcz. with toxic ingredients and *B. hamiltonii* Balak with almost undetectable saikosaponin contents. In some areas, even 5 or 6 different species of local *Bupleurum* are used under the same name of Chaihu. This confusion (intentional or not) results in rather variable qualities of the raw materials for Chaihu in Chinese market.

Accurate identification of the crude drugs is crucial for safe clinical applications of Chaihu in TCM practices and for its chemical and pharmacological researches. However, the conventional anatomical and chemotaxonomical studies provide limited value in determining the botanical origins and controlling the quality of Chaihu (Pan 1996), and a more reliable method for that purpose is urgently needed. Efforts had been made to discuss the feasibility and applicability of internal transcribed spacer (ITS) sequences for determining the identity of *Bupleurum* medicinal plants (Yang et al. 2007; Xie et al. 2009). We found that the homologous alignment of ITS sequences was lower than 75% between members of the genus *Bupleurum* and associated out-groups, while the

* Corresponding author. Tel.: +86 20 61648764.
E-mail address: chaozhi@smu.edu.cn (Z. Chao).

http://dx.doi.org/10.1016/j.phymed.2014.09.001
0944-7113/

within-group alignment was greater than 87%; samples of the same herb showed an ITS similarity of at least 99%. We thus proposed that ITS sequences be used as reliable molecular markers for the identification of Chaihu; but we also found 6 major groups of different species with ITS sequence homology greater than 99%, i.e., these species could not be distinguished by ITS sequence (Xie et al. 2009).

DNA barcoding is a method of species identification which involves sequencing a standard and specific DNA region. It was first proposed by Hebert et al. (2003a) and has been demonstrated as a powerful tool for species identification. The cytochrome *c* oxidase subunit 1 mitochondrial region (COI) is emerging as the standard barcode region for distinguish a wide range of animals (Hebert et al. 2003b, 2004; Ward et al. 2005; Smith et al. 2007), but its evolution rate is too slow for application in plants. In 2005, Kress proposed the combined use of psbA-trnH and ITS of nuclear ribosomal DNA for identifying plants; and Chase et al. (2005) and Newmaster et al. (2006) recommended rbcL for that purpose. The combination of rbcL and matK was proposed as a core barcode to identify land plants by CBOL Plant Working Group (2009). But as the amplification and identification efficiencies of the proposed regions vary with different plant groups, there has been an ongoing effort to screen for universal barcoding markers for plants. In a large sample volume study, Chen et al. (2010) found that nuclear internal transcribed spacer 2 (ITS2) regions had not only a high amplification efficiency, but also a high identification efficiency up to 90%, and they proposed that ITS2 be used as a potential barcode for plant identification.

In this study, we tested 4 candidate loci, namely the chloroplast genes rbcL, matK, psbA-trnH, and ITS2 regions for their validity as DNA barcodes to identify species in Chinese medicinal *Bupleurum*. We aimed to establish a DNA barcode system of Chinese *Bupleurum* archiving the overall information of the samples of crude drug Chaihu to facilitate the identification of Chaihu, and to provide better insights into the genetic relationship and phylogeny of *Bupleurum* species.

Material and methods

Taxon sampling

In total, 51 samples representing 19 *Bupleurum* species were collected from different areas of China including Yunnan, Hebei, Shanxi, and Gansu Provinces, etc. All the plant species were identified by one of the author, Chao Zhi, and Professor Pan Shengli of the School of Pharmacy, Fudan University. Commercially available crude drugs of Bupleuri radix in the form of dried roots were purchased from the local market.

The voucher specimens were deposited in School of Traditional Chinese Medicine, Southern Medical University. Detailed information about the materials used and sequences obtained in the study were provided in Table 1.

DNA extraction, amplification and sequencing

Genomic DNA was extracted from the leaf samples dried in silica gel using a Plant Genomic DNA Extraction Kit (Tiangen Biotech, Beijing, China) according to the manufacturer's protocol. The quality of the extracted DNA samples was verified to ensure their suitability for subsequent amplification. The 4 candidate barcode regions (rbcL, matK, psbA-trnH, ITS2) were tested for their feasibility for barcoding *Bupleurum*. The universal primers used for amplifying these four DNA regions were as follows: rbcL, 1F and 724R (Fay et al. 1997); matK, 390F and 1326R (Cuenoud et al. 2002); psbA-trnH, trnHf (Tate and Simpson 2003) and psbA3′f (Sang et al. 1997). The forward primer (5′-cgtagcgaaatgcgatacttggtg-3′) for ITS2 amplification were designed by the authors, and the reverse primer was ITS4 (White et al., 1990). PCR amplification was performed in 25 μl reaction mixture containing 12.5 μl 2 × Taq PCR Colorless Mix,1 μl each primer, 1 μl genomic DNA, and 9.5 μl ddH_2O. The PCR amplification conditions were as follows: initially at 93 °C for 5 min, 53 °C for 2 min, followed by 35 cycles at 93 °C for 30 s, 53 °C for 45 s, 70 °C for 45 s, with a final extension at 70 °C for 5 min. The PCR products were detected by 1.2% agarose gel electrophoresis, and visualized under ultraviolet light. After purification, the products were sequenced in both directions by Invitrogen Biotechnology (Shanghai) Co., Ltd.

For Chaihu crude drug samples (i.e. dried roots), DNA extraction, amplification and sequencing were carried out in identical procedures and conditions described above.

Sequence analysis

DNA sequences were aligned using the software Clustal X (Thompson et al. 1997) and Mega 5.0 (Tamura et al. 2007) and assembled with Codoncode Aligner version 3.7.1 (CodonCode., USA). In addition, all of the ITS2 sequences in our study were delimited based on hidden Markov models (HMM) annotion methods.

The average intra- and inter-specific distances were calculated to evaluate the intraspecifc variation and interspecific divergence (Meyer and Paulay 2005; Meier et al. 2008). Wilcoxon signed-rank tests and two-sample tests were used for analyzing the results. The DNA barcoding gaps were gained by using the program TaxonDNA (Meier et al. 2006) to evaluate the distributions of intra- and interspecific divergences of each candidate locus. Furthermore, BLAST1 and the nearest genetic distance methods (Ross et al. 2008) were used to test the species identification efficiency.

Results

DNA samples of good quality were obtained from all the test materials. The 4 candidate barcode regions were amplified and sequenced. The amplicon size was 226–227 bp for ITS2, 328–367 bp for psbA-trnH, (923∼) 932–933 bp for rbcL, and (823∼) 882–883 bp for matK. The efficiency of PCR amplification and sequencing is one of the important indicators for evaluating the applicability of DNA barcodes. The ITS2 and rbcL regions exhibited an amplification and sequencing rate of 100% for the test species (Table 2). The PCR success rate for the psbA-trnH and matK was 96.08% and 66.67%, with a sequencing efficiency of 87.76% and 94.12%, respectively.

The intra- and inter-specific variations of the four candidate barcodes have been evaluated by six metrics (Chen et al. 2010) (Table 3). Average inter-specific distance, average theta prime, and minimum inter-specific distance were used for inter-specific divergence assessment. Of the 4 candidate barcode regions, ITS2 had the highest inter-specific divergence in all the 3 dimensions, followed by psbA-trnH and matK, while that of rbcL was the lowest. At the intra-specific level, average intra-specific distance, theta, and average coalescent depth were used to evaluate intra-specific variation. PsbA-trnH had the highest average intra-specific distance, followed by ITS2 and matK, and that of rbcL was the lowest. As to the other two parameters (theta, and average coalescent depth), ITS2 showed the highest level of intra-specific variation, followed by psbA-trnH. All the 4 regions showed a higher genetic variability between than within species.

Wilcoxon signed rank tests on the intra- and interspecific divergence data demonstrated the identical results (Tables 4 and 5). The analysis demonstrated that ITS2 and psbA-trnH represent the most suitable DNA barcodes.

Table 1
Plant materials used in the study.

Species	Specimen voucher	Locality	Collected date	ITS2	psbA	matK	rbcL
B. angustissimum	liangzhenbiao 081804	Pingshan, Hebei	2009/8/18	+	+	+	+
B. angustissimum	chaozhi 82801	Liupanshan, Ningxia	2009/8/28	+	+	–	+
B. candollei	liuli 0908063	Huize, Yunnan	2009/8/6	+	+	+	+
B. chaishoui	liuli 0908102	Wenchuan, Sichuan	2009/8/10	+	+	–	+
B. chinense	chaozhi 72801	Songshan, Beijing	2009/7/28	+	+	+	+
B. chinense	liangzhenbiao 081103	Handan, Hebei	2009/8/11	+	+	+	+
B. chinense	chaozhi 82805	Longde, Ningxia	2009/8/28	+	+	+	+
B. chinense	ciwei 090813	Dalian, Liaoning	2009/8/13	+	+	+	+
B. chinense	chaozhi 090626	Feixi, Anhui	2009/6/26	+	+	+	+
B. chinense	liangzhenbiao 081803	Pingshan, Hebei	2009/8/18	+	+	+	+
B. chinense	chaozhi 090725	Chicheng, Hebei	2009/7/25	+	+	+	+
B. chinense	liuli 0908209	Xihe, Gansu	2009/8/20	+	+	+	+
B. chinense	liuli 0908151	Zhenba, Shanxi	2009/8/15	+	+	–	+
B. chinense	liuli 0908155	Zhenba, Shanxi	2009/8/15	+	–	–	+
B. chinense	liuli 0908162	Zhenba, Shanxi	2009/8/16	–	–	–	+
B. chinense	liuli 0908223	Dingxi, Gansu	2009/8/22	+	+	+	+
B. chinense	liuli 82001	Xihe, Gansu	2009/8/20	+	–	+	+
B. chinense	liuli 0908201	Xihe, Gansu	2009/8/20	+	+	–	+
B. chinense f. *chiliosciadium*	liuli 0908221	Dingxi, Gansu	2009/8/22	+	+	+	+
B. chinense f. *octoradiatum*	chaozhi 72604	Chicheng, Hebei	2009/7/26	+	+	+	+
B. chinense f. *vanheurckii*	chaozhi 72904	Songshan, Beijing	2009/7/29	+	+	+	+
B. falcatum	chaozhi 111028	Nagasaki, Japan	2011/10/28	+	+	+	+
B. falcatum	liangzhenbiao 081501	Anguo, Hebei	2009/8/15	+	+	+	+
B. hamiltonii	liuli 0908041	Kunming,Yunnan	2009/8/4	+	+	+	+
B. hamiltonii	liuli 0908021	Lijiang, Yunnan	2009/8/2	+	+	+	+
B. longicaule var. *giraldii*	liuli 0908161	Zhenba, Shanxi	2009/8/16	+	+	–	+
B. longiradiatum	ciwei 090805	Mudanjiang, Heilongjiang	2009/8/5	+	+	–	+
B. longiradiatum	ciwei 090802	Jiamusi, Heilongjiang	2009/8/2	+	–	–	+
B. longiradiatum	ciwei 0908101	Changbaishan, Jilin	2009/8/10	+	+	+	+
B. longiradiatum f. *australe*	chaozhi 0907019	Huangshan, Anhui	2009/7/19	+	+	–	+
B. malconense	liuli 0908101	Wenchuan, Sichuan	2009/8/10	+	+	–	+
B. marginatum	liuli 0908062	Huize, Yunnan	2009/8/6	+	+	+	+
B. marginatum var. *stenophyllum*	liuli 0907312	Dali, Yunnan	2009/7/31	+	+	+	+
B. marginatum var. *stenophyllum*	liuli 0907301	Dali, Yunnan	2009/7/30	+	–	–	+
B. marginatum var. *stenophyllum*	liuli 0907311	Dali, Yunnan	2009/7/31	+	+	+	+
B. petiolulatum var. *tenerum*	liuli 0908031	Lijiang, Yunnan	2009/8/3	+	–	+	+
B. polyclonum	liuli 0908061	Huize, Yunnan	2009/8/6	+	+	+	+
B. scorzonerifolium	liangzhenbiao 081101	Handan, Hebei	2009/8/11	–	+	–	+
B. scorzonerifolium	ciwei 090801	Jiamusi, Heilongjiang	2009/8/1	+	+	–	+
B. scorzonerifolium	chaozhi 090722	Quanjiao, Anhui	2009/7/22	+	+	–	+
B. scorzonerifolium	chaozhi 72601	Chicheng, Hebei	2009/7/26	+	+	–	+
B. sichuanense	liuli 0908111	Wenchuan, Sichuan	2009/8/11	+	+	+	+
B. smithii	chaozhi 72903	Songshan, Beijing	2009/7/29	+	+	+	+
B. smithii var. *parvifolium*	chaozhi 82803	Liupanshan, Ningxia	2009/8/28	+	+	–	+
B. yinchowense	liuli 090815	Zhenba, Shanxi	2009/8/15	+	–	–	+
B. yinchowense	liangzhenbiao 082204	Yizhou, Shanxi	2009/8/22	–	–	+	+
B. yinchowense	liangzhenbiao 082202	Yizhou, Shanxi	2009/8/22	+	+	+	+
B. yinchowense	liangzhenbiao 082402	Yuncheng, Shanxi	2009/8/24	+	+	+	+
B. yinchowense	liuli 0908208	Xihe, Gansu	2009/8/20	+	+	–	+
B. yinchowense	liuli 908206	Xihe, Gansu	2009/8/20	+	+	+	+
B. yinchowense	chaozhi 82804	Longde, Ningxia	2009/8/28	+	+	+	+

Table 2
Identification efficiency of 4 loci using different methods for species identification.

	ITS2	psbA-trnH	matK	rbcL
Amplicon length (bp)	226–227	328–367	(823–)882–883	(923–)932–933
PCR efficiency (%)	100	96.08	66.67	100
Sequencing efficiency (%)	100	87.76	94.12	100
Idetification efficiency (%)				
Blast1	73.68	72.22	46.15	26.32
Nearest distance	52.63	44.44	30.77	21.05

Table 3
Analysis of inter-specific divergences and intra-specific variation of Candidate barcodes.

	ITS2	matK	psbA-trnH	rbcL
All inter-specific distance	0.0616 ± 0.0426	0.0100 ± 0.0074	0.0298 ± 0.0358	0.0039 ± 0.0040
Theta prime	0.0681 ± 0.0282	0.0116 ± 0.0052	0.0351 ± 0.0265	0.0039 ± 0.0021
Minimum inter-specific distance	0.0204 ± 0.0282	0.0036 ± 0.0065	0.0098 ± 0.0205	0.0006 ± 0.0015
All intra-specific distance	0.0048 ± 0.0108	0.0043 ± 0.0030	0.0111 ± 0.0139	0.0025 ± 0.0022
Theta	0.0252 ± 0.0372	0.0076 ± 0.0062	0.0186 ± 0.0284	0.0026 ± 0.0027
coalescent depth	0.0295 ± 0.0359	0.0096 ± 0.0053	0.0249 ± 0.0289	0.0045 ± 0.0048

Table 4
Wilcoxon signed rank test for inter-specific variation.

W+	W−	Correlation between the inter-specific variation, *n*, *p*	Results
ITS2	matK	W+ = 66640, W− = 1256, *n* = 378, *p* = 1.10E − 57	*p* < 0.05 ITS2 > matK
ITS2	psbA	W+ = 246957, W− = 11164, *n* = 724, *p* = 8.47E − 100	*p* < 0.05 ITS2 > psbA
ITS2	rbcL	W+ = 466340, W− = 1688, *n* = 981, *p* = 1.48E − 157	*p* < 0.05 ITS2 > rbcL
matK	psbA	W+ = 843, W− = 47362, *n* = 314, *p* = 4.35E − 49	*p* < 0.05 matK < psbA
matK	rbcL	W+ = 75033, W− = 2388, *n* = 405, *p* = 1.86E − 58	*p* < 0.05 matk > rbcl
psbA	rbcL	W+ = 263961.W− = 11692, *n* = 762, *p* = 1.94E − 103	*p* < 0.05 psba > rbcl

Table 5
Wilcoxon signed rank test for intra-specific variation.

W+	W−	Correlation between the intra-specific variation, *n*, *p*	Results
ITS2	matK	W+ = 1264, W− = 1976, *n* = 87, *p* = 0.088	*p* > 0.05 ITS2 = matK
ITS2	psbA	W+ = 716, W− = 3200, *n* = 96, *p* = 2.37E − 7	*p* < 0.05 ITS2 < psbA
ITS2	rbcL	W+ = 6077, W− = 3514, *n* = 147, *p* = 0.006	*p* < 0.01 ITS2 > rbcL
matK	psbA	W+ = 51, W− = 1779, *n* = 64, *p* = 2.01E − 10	*p* < 0.05 matK < psbA
matK	rbcL	W+ = 3319, W− = 597, *n* = 91, *p* = 1.49E − 8	*p* < 0.05 matK > rbcL
psbA	rbcL	W+ = 3870, W− = 501, *n* = 99, *p* = 1.09E − 10	*p* < 0.05 psbA > rbcL

We examined the distributions of intra-specific versus inter-specific divergence of the 4 barcodes at a scale of 0.001 distance units (Fig. 1). In an ideal situation, the divergence bar chart should show an intra-specific variation focusing on the side with smaller numerals (left side), while inter-specific variation on the side with greater numerals (right side). As shown in Fig. 1, ITS2 exhibited distinct gaps between the distributions of intra- and inter-specific variation despite of a slight overlap, which was much obvious for rbcL. For psbA-trnH and matK, no evident barcoding gaps were found. We therefore presumed that ITS2 could be used to discriminate most species in this study.

An ideal barcode should exhibit a high ability of species discrimination and identification (Kress et al. 2005; Lahaye et al. 2008; Hollingsworth et al. 2009). BLAST1 and the Nearest Genetic Distance methods were used to estimate the applicability of the 4 loci for species identification (Table 5). As BLAST1 had better identification accuracy and computing speed than the Nearest Distance method, we adopted the identification efficiency in BLAST1 as the criterion for evaluating the candidate barcodes in this study. Our results indicated that ITS2 had the highest authentication capability and correctly identified 73.68% of the test samples at the species level. PsbA-trnH, following ITS2, exhibited an identification efficiency of 72.22%. When ITS2 and psbA-trnH were combined, the identification efficiency increased to 84.21%. In contrast, the rates of successful species identification using matK and rbcL were lower than 50%.

We further evaluated the efficiency of ITS2 to discriminate *Bupleurum* species at a larger scale. ITS2 sequences of 173 samples (representing 73 *Bupleurum* species) were downloaded from GenBank to establish a large database (Neves and Watson 2004; Yang et al. 2007; Xie et al. 2009; Wang et al., 2008a,b, 2011). We found that ITS2 successfully discriminated 64.13% of the samples at the species level.

Discussion

Assessment of the applicability of the four candidate barcodes

An ideal DNA barcode must have high PCR amplification and sequencing efficiency, a significant inter-species divergences and a minimal intra-species variation. The CBOL plant working group proposed the combination of rbcL + matK as a plant barcode; the two fragments were characterized by good primer universality and high amplification efficiency, good sequence quality and high discrimination power. In the present study, rbcL exhibited an amplification and sequencing efficiency of 100%, but its identification efficiency was only 26.32%. As to matK, the amplification efficiency was only 66.67% with the recommended universal primer 390F-1326R, and the identification efficiency was 46.15%. Numerous studies revealed that rbcL and matK were informative to resolve phylogenetic issues at higher taxonomic ranks, but incompetent in dealing with problems at lower ranks such as species discrimination, because these regions often lack variations in closely related species, especially those that have diverged recently in evolution (Tian and Li 2002; Dong et al. 2012). The low amplication success rate of matK had also been observed by other researchers (Kress and Erickson 2007; Gao et al. 2011). The amplicon of proposed universal primers, 390F and 1326R, is a fragment of the whole matK gene. Possible variations at primer sites may lead to amplification failures in 1 out of 3 *Bupleurum*. Both of rbcL and matK showed no advantages in *Bupleurum* discrimination at the species level.

ITS2 was recently proposed as barcode region for plants by Chen et al. About 220 bp in length, ITS2 region has a high evolution rate, giving enough variations to differentiate species better. Besides, the fragment is easily amplified and sequenced to allow identification of materials with DNA degradation such as specimens and crude drugs.

In this study, we found that ITS2 had obvious advantages over the other loci as a barcode for *Bupleurum* in terms of amplification and sequencing efficiency, barcoding gaps, and success rate of identification.

Using the primers designed for ITS2, we achieved PCR amplification and sequencing efficiency both of 100%. As to the identification efficiency, ITS2 alone showed a success rate of 73.68% in *Bupleurum* discrimination; while combined with psbA-trnH sequences, the success rate was increased to 84.21%. In addition, ITS2 successfully discriminated 64.13% of the samples at the species level in a large database from GenBank, which included 173 samples from 73 species.

Based on these results, we recommend that ITS2 region be used as a barcode for differentiating *Bupleurum* species, with psbA-trnH as a complementary locus.

Advantages and limitations of ITS2 barcode in differentiating Bupleurum species

We previously demonstrated that ITS sequences could be used as reliable molecular markers for identifying *Bupleurum* species, but they failed to distinguish 6 groups of different species, i.e., (1) *B. chinese*, *B. malconense*, *B. sichuanense*; (2) *B. smithii*, *B. commelynoideum* var. *flaviflorum*, *B. sibiricum*;

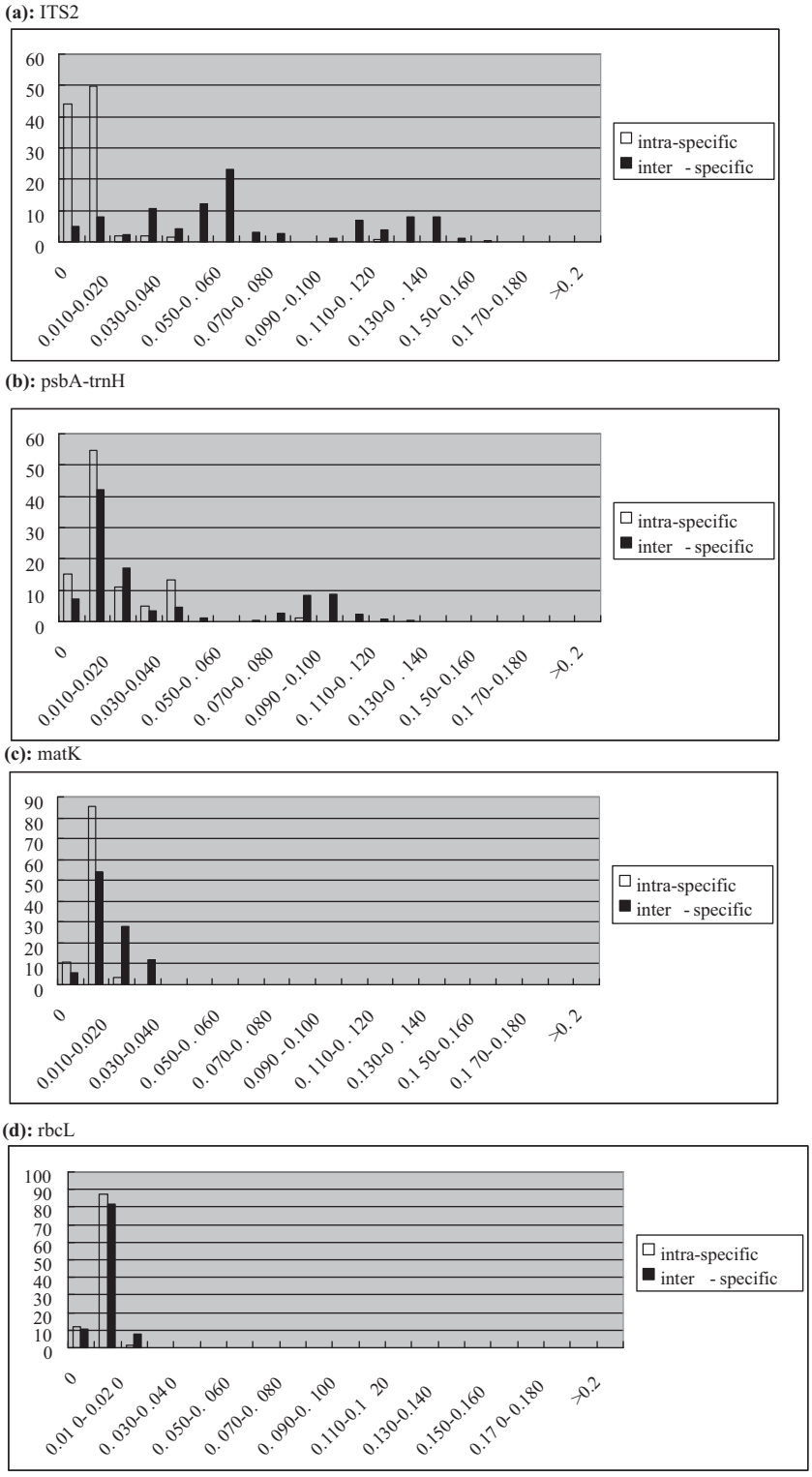

Fig. 1. *Barcoding gap*: Distribution for intra and inter-specific variation (%): (a) ITS2, (b) psbA-trnH, (c) matK, (d) rbcL

(3) *B. scorzonerifolium*, *B. angustissimum*; (4) *B. rockii*, *B. petiolulatum* var. *tenerum*; (5) *B. marginatum*, *B. marginatum* var. *stenophyllum*; (6) *B. candollei*, *B. hamiltonii* (Xie et al. 2009). The proposed barcode regions ITS2 and psbA-trnH can provide solutions to most of the species groups. For instance, ITS2 alone is capable of distinguishing *B. malconense* out of group 1, and *B. smithii* from *B. commelynoideum* var. *flaviflorum*, *B. candollei* from *B. hamiltonii*; using the combination of ITS2 and psbA-trnH, *B. sichuanense* could be distinguished from *B. chinese*, and *B. marginatum* var. *stenophyllum* from its proto-variety *B. marginatum*.

All the 4 investigated candidate barcode fragments were unable to differentiate *B. chinense* and *B. yinchowense*, which account for the majority of commercially available Chaihu crude drugs in market. The two species are closely related, share similar morphological

properties, and both have ellipsoidal pollen grains; phylogenetic analysis also revealed their close relationship (Yang et al. 2007). The high sequence homology of the 4 molecular markers between these two species (the ITS2 sequence similarity even reached 100%) causes much difficulty in distinguishing them from each other. But the ITS in full length could distinguish *B. yinchowense* from *B. chinense* (Xie et al. 2009). According to China Plant BOL Group, ITS should be incorporated into the core barcode for seed plants (China Plant BOL Group et al. 2011). In the case of *Bupleurum*, the limitations of ITS2 can be mended with ITS. We also amplified and sequenced full-length ITS (including nrITS1-5.8S-ITS2, about 600–609 bp in length) of some *Bupleurum* species, but the PCR amplification success rate was much lower than that of ITS2, which was consistent with the findings by Chase and Kress (Chase et al. 2007; Kress and Erickson 2007). We therefore excluded the whole ITS region from further consideration, and maintained the recommendation of ITS2 as the core barcode for *Bupleurum*.

Validation of proposed barcode ITS2 in identification of Chaihu crude drug

In most studies about DNA barcode, the datasets were established on the DNA from leaf materials. However, the crude drug of Chaihu derives from the underground parts of *Bupleurum* plants. Thus, in order to investigate the feasibility of the barcode regions we recommended in identifying the crude drug of Chaihu, the roots of *B. chinense*, *B. scorzonerifolium*, and *B. smithii* (obtained from the Qingping market of Guangzhou and identified macroscopically by Prof. Pan Shengli and Chao Zhi) were selected for a preliminary double-blind trials. We marked the samples and asked a technician blinded to the sample identity to perform experiments following a protocol we provided, including DNA extraction, ITS2 amplification, sequencing, and blasting in *Bupleurum* barcode database. The technician correctly identified all the 3 samples.

Acknowledgements

The study was supported by National Natural Science Foundation of China (NSFC no. 30873387, and no. 81373905). We are grateful to the help of Professor Shilin Chen and Dr. Hui Yao.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.phymed.2014.09.001.

References

CBOL Plant Working Group, 2009. A DNA barcode for land plants. Proc. Natl. Acad. Sci. U.S.A. 106, 12794–12797.

Chase, M.W., Salamin, N., Wilkinson, M., Dunwell, J.M., Kesanakurthi, R.P., Haider, N., Savolainen, V., 2005. Land plants and DNA barcodes: short-term and long-term goals. Philos. Trans. R. Soc. London, Ser. B 360, 1889–1895.

Chase, M.W., Cowan, R.S., Hollingsworth, P.M., van den Berg, C., Madrinan, S., Petersen, G., Seberg, O., Jorgsensen, T., Cameron, K.M., Carine, M., Pedersen, N., Hedderson, T.A.J., Conrad, F., Salazar, G.A., Richardson, J.E., Hollingsworth, M.L., Barraclough, T.G., Kelly, L., Wilkinson, M., 2007. A proposal for a standardised protocol to barcode all land plants. Taxon 56, 295–299.

China Plant BOL GroupLi, D.Z., Gao, L.M., Li, H.T., Wang, H., Ge, X.J., Liu, J.Q., Chen, Z.D., Zhou, S.L., Chen, S.L., Yang, J.B., Fu, C.X., Zeng, C.X., Yan, H.F., Zhu, Y.J., Sun, Y.S., Chen, S.Y., Zhao, L., Wang, K., Yang, T., Duan, G.W., 2011. Comparative analysis of a large dataset indicates that internal transcribed spacer (ITS) should be incorporated into the core barcode for seed plants. Proc. Natl. Acad. Sci. U.S.A. 108, 19641–19646.

Chen, S.L., Yao, H., Han, J.P., Liu, C., Song, J.Y., Shi, L.C., Zhu, Y.J., Ma, X.Y., Gao, T., Pang, X.H., Luo, K., Li, Y., Li, X.W., Jia, X.C., Lin, Y.L., Leon, C., 2010. Validation of the ITS2 region as a novel DNA barcode for identifying medicinal plant species. PLoS One 5, e8613.

Cuenoud, P., Savolainen, V., Chatrov, L.W., Powell, M., Grayer, R.J., Chase, M.W., 2002. Molecular phylogenetics of Caryophyllales based on nuclear 18S rDNA and plastid rbcL, atpB, and matK sequences. Am. J. Bot. 89, 132–144.

Dong, W., Liu, J., Yu, J., Wang, L., Zhou, S., 2012. Highly variable chloroplast markers for evaluating plant phylogeny at low taxonomic levels and for DNA barcoding. PLoS One 7, e35071.

Fay, M.F., Swensen, S.M., Chase, M.W., 1997. Taxonomic affinities of *Medusagyne oppositifolia* (Medusagynaceae). Kew Bull. 52, 111–120.

Gao, T., Sun, Z., Yao, H., Song, J., Zhu, Y., Ma, X., Chen, S., 2011. Identification of Fabaceae plants using the DNA barcode matK. Planta Med. 77, 92–94.

Hebert, P.D.N., Cywinska, A., Ball, S.L., deWaard, J.R., 2003a. Biological identifications through DNA barcodes. Proc. R. Soc. London, Ser. B 270, 313–321.

Hebert, P.D.N., Ratnasingham, S., deWaard, J.R., 2003b. Barcoding animal life: cytochrome *c* oxidase subunit I divergences among closely related species. Proc. R. Soc. London, Ser. B 270, S96–S99.

Hebert, P.D.N., Stoeckle, M.Y., Zemlak, T.S., Francis, C.M., 2004. Identification of birds through DNA barcodes. PLoS Biol. 2, 1657–1663.

Hollingsworth, M.L., Andra Clark, A., Forrest, L.L., et al., 2009. Selecting barcoding loci for plants: evaluation of seven candidate loci with species level sampling in three divergent groups of land plants. Mol. Ecol. Resour. 9, 439–457.

Kress, W.J., Wurdack, K.J., Zimmer, E.A., Weigt, L.A., Janzen, D.H., 2005. Use of DNA barcodes to identify flowering plants. Proc. Natl. Acad. Sci. U.S.A. 102, 8369–8374.

Kress, W.J., Erickson, D.L., 2007. A two-locus global DNA barcode for land plant: the coding rbcL gene complements the noncoding trnH-psbA spacer region. PLoS One 2, e508.

Lahaye, R., Van der Bank, M., Bogarin, D., Warner, J., Pupulin, F., Gigot, G., 2008. DNA barcoding the floras of biodiversity hotspots. Proc. Natl. Acad. Sci. U.S.A. 105, 2923–2928.

Meier, R., Shiyang, K., Vaidya, G., Ng, P.K.L., 2006. DNA barcoding and taxonomy in Diptera: a tail of high intraspecific variability and low identification success. Syst. Biol. 55, 715–728.

Meier, R., Zhang, G.Y., Ali, F., 2008. The use of mean instead of smallest interspecific distances exaggerates the size of the "Barcoding gap" and leads to misidentification. Syst. Biol. 57, 809–813.

Meyer, C.P., Paulay, G., 2005. DNA barcoding: error rates based on comprehensive sampling. PLoS Biol. 3, 2229–2238.

Neves, S.S., Watson, M.F., 2004. Phylogenetic relationships in *Bupleurum* (Apiaceae) based on nuclear ribosomal DNA ITS sequence data. Ann. Bot. 93, 379–398.

Newmaster, S.G., Fazekas, A.J., Ragupathy, S., 2006. DNA barcoding in land plants: evaluation of rbcL in a multigenetiered approach. Can. J. Bot. 84, 335–341.

National Pharmacopoeia Committee, 2010. Pharmacopoeia of the People's Republic of China. Chemical Industry Press, Beijing, pp. 263–264.

Pan, S.L., 1996. Investigation on resources of "Chaihu" and identification of its commodities. J. Chin. Med. Mater. 19, 231–234.

Pan, S.L., 2006. *Bupleurum* Species: Scientific Evaluation and Clinical Applications. Taylor & Francis Group, London.

Ross, H.A., Murugan, S., Li, W.L.S., 2008. Testing the reliability of genetic methods of species identification via simulation. Syst. Biol. 57, 216–230.

Sang, T., Crawford, D.J., Stuessy, T.F., 1997. Chloroplast DNA phylogeny, reticulate evolution, and biogeography of *Paeonia* (Paeoniaceae). Am. J. Bot. 84, 1120–1136.

She, M.L., Watson, M.F., 2005. Flora of China, vol.14. Science Press, Beijing & Missouri Botanical Garden Press, St. Louis, MO, pp. 60–74.

Smith, M.A., Woodley, N.E., Janzen, D.H., Hallwachs, W., Hebert, P.D., 2007. DNA barcodes affirm that 16 species of apparently generalist tropical parasitoid flies (Diptera, Tachinidae) are not all generalists. Proc. Natl. Acad. Sci. U.S.A. 104, 4967–4972.

Tamura, K., Dudley, J., Nei, M., Kumar, S., 2007. Mega 4: molecular evolutionary genetics analysis (mega) software version 4.0. Mol. Biol. Evol. 24, 1596–1599.

Tate, J.A., Simpson, B.B., 2003. Paraphyly of *Tarasa* (Malvaceae) and diverse origins of the polyploid species. Syst. Bot. 28, 723–737.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmouqin, F., Hiqqins, D.G., 1997. The Clustal X windows interface: flexible strategies for multiple sequences alignment aided by quality analysis tools. Nucleic Acids Res. 25, 4876–4882.

Tian, X., Li, D.Z., 2002. Application of DNA sequences in plant phylogenetic study. Acta Bot. Yunnan. 24, 170–184.

Wang, Q.Z., He, X.J., Zhou, S.D., Wu, Y.K., Yu, Y., Pang, Y.L., 2008a. Phylogenetic inference of the genus *Bupleurum* (Apiaceae) in Hengduan Mountains based on chromosome counts and nuclear ribosomal DNA ITS sequences. J. Syst. Evol. 46, 142–154.

Wang, Q.Z., Zhou, S.D., Liu, T.Y., Pang, Y.L., Wu, Y.K., He, X.J., 2008b. Phylogeny and classification of Chinese *Bupleurum* based on nuclear ribosomal DNA internal transcribed spacer and rps16. Acta Biol. Cracov. Bot. 50, 105–116.

Wang, C.B., Ma, X.G., He, X.J., 2011. A taxonomic re-assessment in the Chinese *Bupleurum* (Apiaceae): insights from morphology, nuclear ribosomal internal transcribed spacer, and chloroplast (trnH-psbA, matK) sequences. J. Syst. Evol. 49, 558–589.

Ward, R.D., Zemlak, T.S., Innes, B.H., Last, P.R., Hebert, P.D.N., 2005. DNA barcoding Australia's fish species. Philos. Trans. R. Soc. London, Ser. B 360, 1847–1857.

White, T.J., Bruns, T.D., Lee, S.B., Taylor, J.W., 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In: Innis, M.A., Gelfand, D.H., Sninsky, J.J., White, T.J. (Eds.), PCR Protocols—A Guide to Methods and Applications. Academic Press, San Diego, CA, pp. 315–322.

Xie, H., Huo, K.K., Chao, Z., Pan, S.L., 2009. Identification of crude drugs from Chinese medicinal plants of the genus *Bupleurum* using ribosomal DNA ITS sequences. Planta Med. 75, 89–93.

Yang, Z.Y., Chao, Z., Huo, K.K., Xie, H., Tian, Z.P., Pan, S.L., 2007. ITS sequence analysis used for molecular identification of the *Bupleurum* species from northwestern China. Phytomedicine 14, 416–422.