JOURNAL OF PURE AND APPLIED MICROBIOLOGY, Nov. 2013. Vol. **7**(Spl. Edn.), p. 525-539

Genetic Divergence and Phylogenetic Analysis of *Hedyotis diffusa* and *H. corymbosa* Based on Nuclear Ribosomal DNA ITS Sequence

Yan-Lin Sun[1]* and Soon-Kwan Hong[2]*

[1]School of Life Sciences, Ludong University, Yantai, Shandong, 264-025, China.
[2]Department of Bio-Health Technology, Kangwon National University, Chuncheon, Kangwon, 200-701, Korea.
Institute of Bioscience and Biotechnology, Kangwon National University, Chuncheon, Kangwon, 200-701, Korea.

(Received: 27 September 2013; accepted: 04 November 2013)

***Hedyotis diffusa* (*Oldenlandia diffusa*) and *H. corymbosa* (*O. corymbosa*), members of the Rubiaceae family, are mainly distributed in Asia. Both species, as medicinal herbs, have been traditionally used for treatment of cancers, viral infections, and hepatitis in some Asian countries. However, only *H. diffusa* refers to the ingredient of Chinese patent medicine by Chinese Pharmacopoeia. For clear identification, the phylogenetic utility of nuclear ribosomal DNA internal transcribed spacer (ITS) was investigated between different *H. diffusa* and *H. corymbosa* populations. The total ITS sequence of *H. diffusa* was 791 bp, and had been submitted to GenBank on NCBI (accession number: JF837601). The total ITS sequence of *H. corymbosa* was 785 bp (accession number: JF837603). As length variation caused by different region discriminations and different populations grown in different geographical conditions, there was intraspecific dissimilarity, nothing of interspecific variation. Both species were previously reported to be distinguished clearly using ITS sequence, in this study, we obtained different result: *H. corymbosa* AY438318 and AY534053 interfused into the *H. diffusa* group, while *H. diffusa* AY862575, AY438319, and JF837601 interfused into the *H. corymbosa* group. This result was suggested to be mainly caused by source authentication, ambiguous region discrimination, and various geographical conditions.**

Key words: *Hedyotis diffusa*, *Hedyotis corymbosa*, *Oldenlandia*, Molecular identification, nrDNA ITS.

Hedyotis diffusa, a member of the Rubiaceae family, also known by alternate *Oldenlandia diffusa* and named as Baihuasheshecao in Chinese, is mainly distributed in Orient and tropical regions of Asia[1,2]. *H. diffusa*, as a popular folk medicinal plant, is traditionally used for lung, liver, rectal, prostate cancer and other tumor treatments[3, 4]. *H. corymbosa*, also a member of the Rubiaceae family, known by *O. corymbosa* and called as Peh Hue Juwa Chi Cao in Taiwan, China[5], is a weedy herb mainly distributed throughout Asia. *H. corymbosa* is extensively used in modern Chinese practice for the treatment of viral infections, cancer, syndromes involving "toxic heat", acne, boils, skin ailments, appendicitis, hepatitis, eye diseases and bleeding[6]. However, according to the Chinese Pharmacopoeia, only *H. diffusa* is referred to be herbal medicine, but not *H. corymbosa*[7]. Moreover, due to their having similar morphological characters and chemical fingerprints[8], both *H. diffusa* and *H. corymbosa* are frequently adulterated with each other in the

* To whom all correspondence should be addressed.
Tel.: 86-535-6685003 (YL Sun);
82-33-2506476 (SK Hong);
Fax: 82-33-2506470 (SK Hong);
E-mail: laddiya@hotmail.com (YL Sun);
soonkwan@kangwon.ac.kr (SK Hong)

wholesale and food markets[9]. Although some scientists have reported that these two medicinal herbs have their respective, special functions and efficacies[5], to ensure the correct use and use safety, an efficient identification method using reliable markers for the authentication becomes required urgently and necessary.

Until now, many approaches have been applied to differentiate the genuine *H. diffusa* from *H. corymbosa*. Some botanists have reported that the two species look very much alike even when fresh, differing only by the variable number of flowers, length of pedicels and shape of stems[10]. However, these characters become unapparent when dried and dispensed in short fragments. Some chemists have attempted to authenticate *H. diffusa* and *H. corymbosa* by thin layer chromatography (TLC)[11] and high-performance liquid chromatography (HPLC)[12, 13]. However, these studies relied on some unidentified spots or markers[9,14]. The identification of *H. diffusa*, *H. corymbosa*, and other similar species have also been performed based on macroscopic characteristics, however, this method could only distinguish between *H. corymbosa* and *H. tenelliflora* by routine light microscopy, but not between *H. diffusa* and *H. corymbosa*[15]. To date, using molecular data to infer phylogenetic relationships has become a good choice, as some DNA regions could synchronously response to the evolutionary rate[16].

Molecular classification shares many advantages compared to traditional classification: molecular technique is independent of phenotypic and environmental variations and can help pinpointing plant materials to specific taxon; simple DNA cloning and sequencing may more rapidly provide more phylogenetic information embodied in some regions; the information masked by homoplasy of DNA sequences is exacter and more efficient. DNA barcoding has been considered as a useful tool for rapidly species identification based on DNA sequence[17, 18]. The chloroplast regions including *atpB-rbcL*, *petD*, *rps16*, and *trnL-F*, and nuclear regions including external transcribed spacer (ETS) and internal transcribed spacer (ITS) have been utilized in the phylogeny study of the tribe Spermacoceae of family Rubiaceae[19]. Molecular identification using five DNA barcoding regions such as ITS, *matK*, *petD*, *trnH-psbA*, and *rbcL* has been reported among 25 *Hedyotis* taxa, including *H. diffusa* and *H. corymbosa*[20]. Among the five DNA barcoding regions, ITS showed the best species discrimination with no shared alleles between any of 24 distinct species, except of *H. assimilis* and *H. mellii*[20]. ITS region of the nuclear ribosomal DNA (nrDNA) cistrons was found to be one of the more frequently utilized regions for phylogenetic analyses at the genus and species levels[21], and provide the highest number of informative characters and resolving power in the phylogenetic analysis[19]. Reconstructing the phylogeny is usually addressed by using different markers like ITS including ITS1 and ITS2 for low level and 18S or 28S rDNA for high level classification[22]. Particularly, ITS2 region of nrDNA cistron has been considered to be a marker suitable for taxonomic classification over a wide range of levels[21].

To distinguish *H. diffusa* from *H. corymbosa*, DNA sequence analysis of ITS1 and ITS2 region has been attempted by many investigations[23-26]. Particularly, Liu et al. focused on the positive role of ITS2 region sequence in efficient species identification[24], and designed the *H. diffusa*-specific PCR primers for ITS2 region amplification. Using the newly designed primers could successfully amplify and obtain a ~392 bp of PCR product using template DNA isolated from *H. diffusa*, but not other DNAs isolated from 9 confused *H. diffusa* samples[24]. Due to that *Hedyotis* has been revealed to be polyphyletic based on molecular studies[19, 27-29], together with the currently available limited sampling of Asian species in the recent molecular phylogenetic analysis, a broad concept of *H. diffusa* and *H. corymbosa* with known ITS sequences in GenBank data is presented here.

MATERIALS AND METHODS

Plant materials and growth conditions

Mature seeds of one *H. diffusa* and one *H. corymbosa* population were collected from Korea, sent by Myung-Hun Yeom, Duck-Hee Kim, and Han-Gon Kim in R & D center, Amore-Pacific Co., Yonpin, 446-729, Korea. The isolates have been named as LT-L and ST-L, respectively, used for the convenience of sample management (Fig. 1). The species identification was made by Research and Development center, Amore-Pacific Co., who uses

H. diffusa and *H. corymbosa* in making natural cosmetics. The total ITS sequences of *H. diffusa* and *H. corymbosa* have been published by GenBank, NCBI, with the accession numbers of JF837601 and JF837603, respectively.

Mature seeds of *H. diffusa* LT-L and *H. corymbosa* ST-L isolate were surface-sterilized with 70% ethanol for 30 seconds and then rinsed with sterile water for five times. The sterilized seeds were grown in a pot containing a mixture of sterilized soil and vermiculite (v:v, 3:1) in a green house condition with 40-60% relative wet rate, a 16/8 h (light/dark) photoperiod, and 24±2°C. After germination, the fresh leaf tissue was used for DNA extraction.

DNA extraction and PCR amplification

DNA was extracted from fresh leaves of *H. diffusa* and *H. corymbosa* using the modified cetyltrimethylammonium bromide (CTAB) method[30]. Common ITS primer sets ITS5, 5'-GAA AGT AAA AGT CGT AAC AAG G-3' and ITS4, 5'-TCC TCC GCT TAT TGA TAT GC-3' were used to amplify the nrDNA ITS region including ITS1, 5.8S rRNA, ITS2 regions[31]. PCR amplification was conducted using this set of primers with the following program: 35 cycles of denaturation at 95°C for 1 min, annealing at 53°C for 1 min, and a final extension step at 72°C for 1.5 min. All PCR products were purified before DNA sequence analysis using a QIAquick PCR Purification Kit (QIAGEN, Korea) according to the manufacturer's instructions. Purified PCR products were then sequenced at MACRO GEN Advancing through Genomics (Korea).

Sequence analysis

Known ITS sequences of *H. diffusa* and *H. corymbosa* were searched in Nucleotide on NCBI. Six *H. diffusa* ITS sequences and night *H. corymbosa* ITS sequences were obtained, with the accession number, submitting personal information, and taxon information shown in Table 1. Analogue of our sequences was detected with BLAST on NCBI. Nucleotide alignments were obtained independently for total ITS, ITS1 and ITS2 region using DNAMAN 6.0. The phylogenetic tree of all seven *H. diffusa* and ten *H. corymbosa* populations were also constructed based on neighbor joining method using DNAMAN 6.0. In addition, other factors of ITS sequences were calculated, such as sequence length, (G+C) content (%), genetic distance (GD).

RESULTS

PCR amplification of total ITS region

Total ITS regions were successfully amplified by PCR using common primer sets, ITS4/ITS5, and about 840 bp of PCR product was obtained. The size of the ITS regions were 791 bp for *H. diffusa* and 785 bp for *H. corymbosa*. The interspecific similarity of total ITS region between *H. diffusa* and *H. corymbosa* was 88.43%, that mainly caused by ITS1 region because the interspecific similarity of ITS1 region was 80.79% (Table 1). Similar results could be found in the (G+C) content (%), that *H. corymbosa* showed a slight higher (G+C) content (%) content, especially in the ITS1 region, the difference had exceeded 1.0 %. From the homology result, the ITS1 region sequences of both species only reached 83.3% homology with each other, combined with absolutely identical 5.8S rRNA region and relatively high homologic ITS2 region, the total ITS region between *H. diffusa* and *H. corymbosa* showed 90.0% homology. The genetic distance between both species obtained the same results (Table 1).

Sequence analysis of known *H. diffusa* populations

In NCBI GenBank data, only six total ITS sequences of *H. diffusa* were found, with the accession numbers of AY438319, AY862575, EF570985, EF570986, EF570988, and FJ606747 (Table 2). From different phylogenetic trees according to ITS1, 5.8S, and ITS2 region sequences, two major clades were obtained, with AY438319, AY862575, and our sequence, JF837601 forming one, and other sequences forming the other (Fig. 2). Very high similarity was obtained among sequences of each clade, e.g. AY438319, AY862575, and our sequence having very similarity to each other, with 99%, 100%, and 100% in the ITS1, 5.8S, ITS2 phylogenetic trees, respectively (Fig. 2). However, these both clades showed relatively less homology between AY438319 and AY862575, and other sequences, which was found to be mainly caused by the length variation caused by confused ITS region discrimination.

To authors' knowledge, the known ITS region sequences from NCBI were distinguished into 18S rRNA, ITS1, 5.8S rRNA, ITS2, and 28S or 26S rRNA according to different region discrimination methods. For instance, the annotated start site of ITS1 region was very

Table 1. Sequence analysis of *H. diffusa* (accession number: JF837601) and *H. corymbosa* (accession number: JF837603) on length, (G+C) content (%), similarity, homology, genetic distance

Region	Total ITS region		ITS1 region		5.8S rRNA region		ITS2 region		18S rRNA region		28S rRNA region	
Species	*H. diffusa*	*H. corymbosa*	*H. diffusa*	*H. corymbosa*	*H. diffusa*	*H. corymbosa*	*H. diffusa*	*H. corymbosa*	*H. diffusa*	*H. corymbosa*	*H. diffusa*	*H. corymbosa*
Length (bp)	791	785	328	318	163	163	300	304	-	-	-	-
(G+C)%	56.26	56.94	54.27	55.66	53.37	53.37	60.00	60.20	45.74	58.02	53.13	48.72
Similarity (%)	88.43	80.79	100.00	90.46	17.49	17.50						
Homology (%)	90.0	83.3	100.0	91.7	48.1	35.9						
Genetic distance	0.100	0.167	0.000	0.083	0.519	0.641						

means unevaluated due to partial sequences.

Table 2. ITS region sequence information and GenBank accession number of known *H. diffusa* populations from NCBI.

GenBank accession number	Size (bp)	ITS1 region	5.8S rRNA region	ITS2 region	Submission information	Definition	Reference
AY438319	603	15-208	209-365	366-584	Liu (2003)	Partial 18S, ITS1, 5.8S, ITS2, and partial 28S	[24, 25]
AY862575	570	1-194	195-352	353-570	Yuan et al. (2007)	Partial ITS1, 5.8s, and partial ITS2	—
EF570985	566	1-189	190-347	348-566	Li et al. (2009)	ITS1, 5.8S, and ITS2	[9]
EF570986	565	1-188	189-346	347-565	Li et al. (2009)	ITS1, 5.8S, and ITS2	
EF570988	565	1-188	189-346	347-565	Li et al. (2009)	ITS1, 5.8S, and ITS2	
FJ606746	644	35-215	216-373	374-588	Daniel and Knoess (2009)	Partial 18S, ITS1, 5.8S, ITS2, and partial 26S	—

— means not published yet or no reported reference.

different, with TTG (Fig. 3) in AY438319, EF570985, EF570986, and EF570988, but with latter ATC in FJ606746. However, AY862575 shared highly conversed sequences with AY438319, EF570985, EF570986, and EF570988, but was not annotated the start site in the description of sequence information. This result forcefully validated their discrimination of ITS1, 5.8S rRNA, and ITS2 region was different. For our *H. diffusa* ITS sequence (accession number: JF837601), the discrimination method for ITS1, 5.8S rRNA, and ITS2 regions followed that of known *H. corymbosa* ITS region sequence with accession number of AM939501, because the analogue of our *H. diffusa* ITS sequence detected by BLAST on NCBI showed that the *H. corymbosa* AM939501 sequence has the highest query coverage (57%) and identification (99%) with our sequence. Thus, ITS1 of our *H. diffusa* ITS sequence contained 328 bp (Table 1), while ITS1 of AY438319, EF570985,

Table 3. Voucher specimen and habitat of known *H. diffusa* and *H. corymbosa* sequences and our sequence

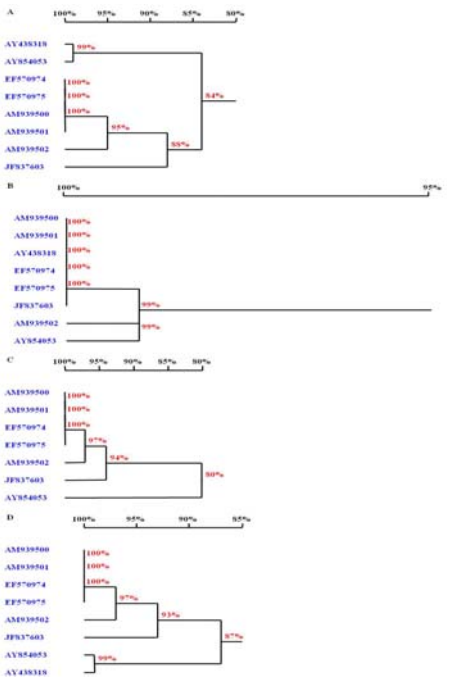
Species	Accession number	Voucher specimens	Habitat
H. diffusa	AY438319	Z. Liu	Jiangsu, China
	AY862575	C.I. Yuan, Y.C. Hsieh, & M.Y. Chiang	Taichung, Taiwan, China
	EF570985	M. Li 041	Hong Kong, China
	EF570986	M. Li 035	Hong Kong, China
	EF570988	M. Li 042	Hong Kong, China
	FJ606746	cdK142	Bonn, Germany
	JF837601	M.H. Yeom, D.H. Kim, & H.G. Kim LT-L	Korea
H. corymbosa	AM939500	Andersson 2260	Australia
	AM939501	Andersson & Nilsson 2263	Gabon
	AM939502	Dessein et al. 487	Zambia
	AY438318	Z. Liu	Jiangsu, China
	AY854053	C.I. Yuan, Y.C. Hsieh, & M.Y. Chiang	Taichung, Taiwan, China
	EF570974	S.Y. Hu & P. But 24052	Hong Kong, China
	EF570975	W.L. Chu 005	Hong Kong, China
	JF837603	M.H. Yeom, D.H. Kim, & H.G. Kim LT-L	Korea

Table 4. ITS region sequence information and GenBank accession number of known *H. corymbosa* populations from NCBI

GenBank accession number	Size (bp)	ITS1 region	5.8S rRNA region	ITS2 region	Submission information	Definition	Ref.
AF381484	191	—	—	—	Church (2003)	Partial 18S, ITS1, and partial 5.8S	[32]
AM939500	781	1-319	320-482	483-781	Karehed *et al*. (2008)	ITS1, 5.8S, and ITS2	[19]
AM939501	781	1-319	320-482	483-781	Karehed *et al*. (2008)	ITS1, 5.8S, and ITS2	
AM939502	782	1-319	320-482	483-782	Karehed *et al*. (2008)	ITS1, 5.8S, and ITS2	
AY438318	232	12-198	199-232	—	Liu (2003)	Partial 18S, ITS1, and partial 5.8S	[24 , 25]
AY438323	292	—	1-53	54-273	Liu (2003)	Partial 5.8S, ITS2, and partial	28S
AY854053	565	1-188	189-345	346-565	Yuan *et al*. (2006)	Partial ITS1, 5.8S, and partial ITS2	—
EF570974	570	1-192	193-355	356-570	Li *et al*. (2009)	Partial ITS1, 5.8S, and partial ITS2	[9]
EF570975	570	1-192	193-355	356-570	Li *et al*. (2009)	Partial ITS1, 5.8S, and partial ITS2	

— means not published yet or no reported reference

Fig. 1. Morphological characteristics of *H. diffusa* LT-L (A) and *H. corymbosa* ST-L (B) used in this study



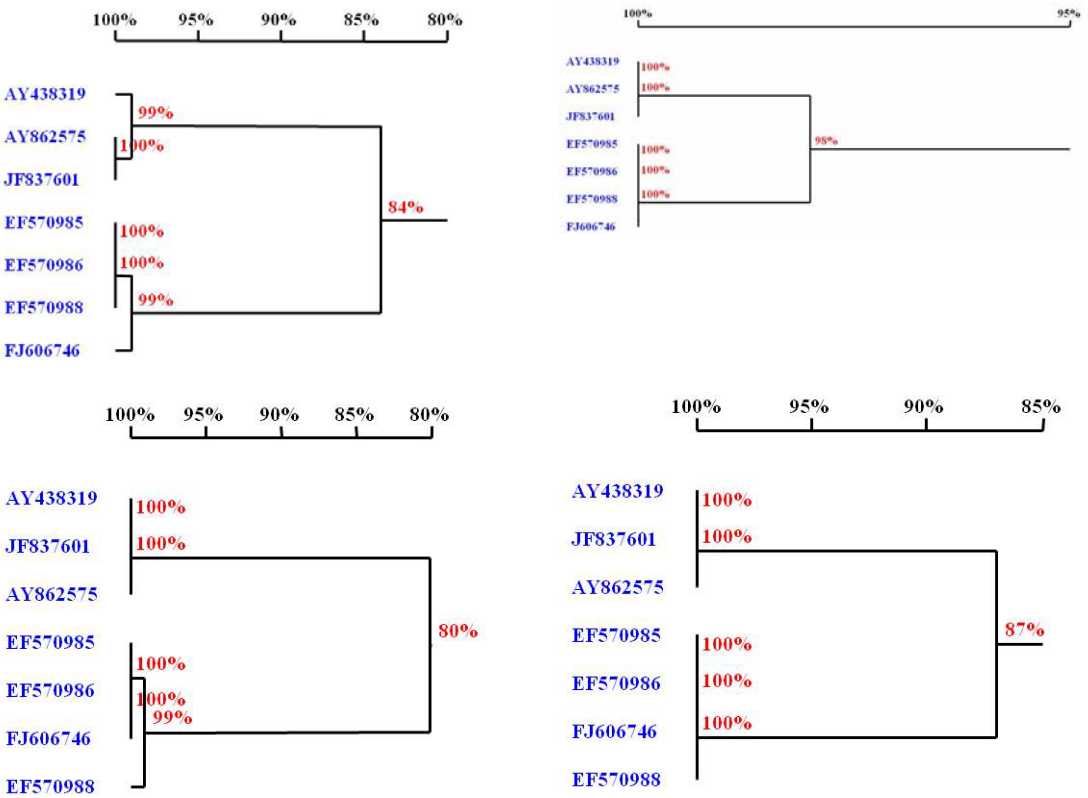


Fig. 2. Phylogenetic relationships of *H. diffusa* among ITS1 (A), 5.8S rRNA (B), ITS2 region (C), and total ITS region (D) of known sequences from NCBI and our sequence, JF837601

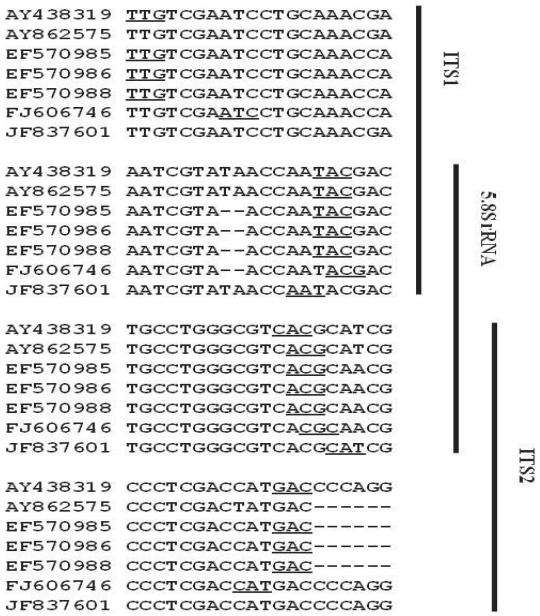


Fig. 3. Different discriminations for ITS1, 5.8S, and ITS2 of known *H. diffusa* sequences and our sequence. Letters with underlined mean the start site and the end site of only ITS2 region annotated in relevant known sequences from NCBI. In ITS1 region discrimination, AY438319 had 14 bp deletion in the start of shown sequence, and FJ606746 had 27 bp deletion

EF570986, EF570988, and FJ606746 contained only 181-197 bp (Table 2).

The difference of the 5.8S rRNA region among six known sequences and our sequence was relatively slight and receivable, showing 163 bp in our sequence and 157-158 bp in other sequences including AY862575 (Table 1, 2). The start and end sites of 5.8S rRNA region still showed certain excursion shown in Fig. 3.

The ITS2 region discrimination of *H. diffusa* exist disputes again, just like the ITS1 region discrimination described above. ITS2 regions of AY438319, EF570985, EF570986, and EF570988 contained 215-219 bp, while ITS2 region of our sequence contained 300 bp (Table 1, 2). Although the start site of ITS2 region showed slight different among six known sequences and our sequence, the end site was all annotated at GAC, except FJ606746, AY862575, and our sequence (Fig. 3).

Considering clade division in the phylogenetic tree, it did not match absolutely with the length variation caused by ITS region discrimination, indicating that there must be another factor affecting the phylogenetic analysis. To understand the phylogenetic relationship of *H. diffusa*, the influence of geographical conditions evaluated by voucher specimens and sample habitats were summarized in Table 3. The results suggested that populations from Jiangsu, and Taiwan of China were nearer to each other in phylegenetic tree, while populations from Hong Kong of China, Germany (or cultured), and Korea formed one clade, sharing high identification in the molecular levels.

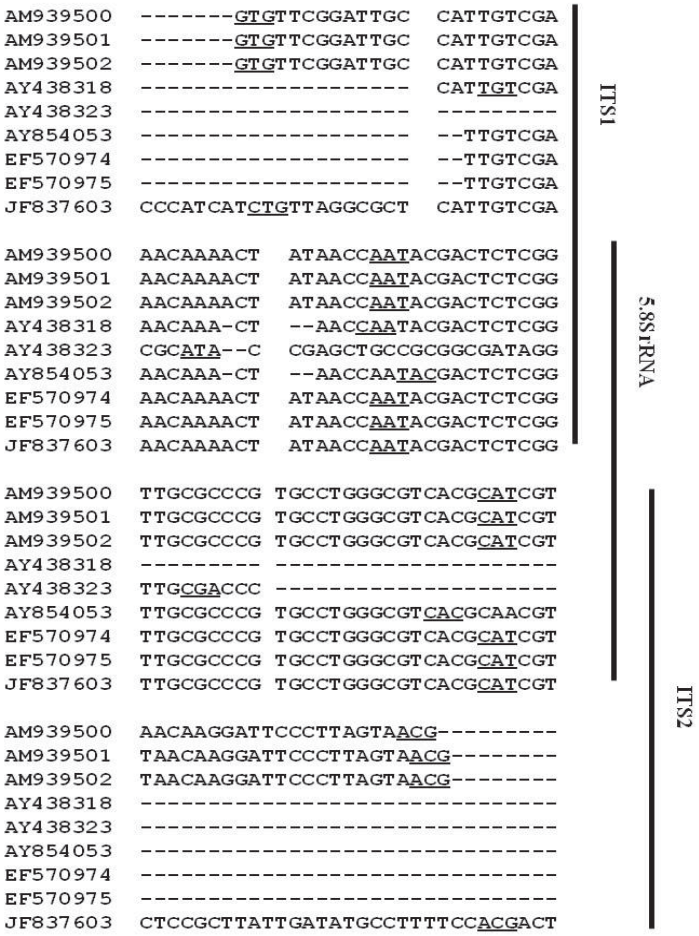


Fig. 4. Different discriminations for distinguishing ITS1, 5.8S, and ITS2 of known *H. corymbosa* sequences and our sequence. Letters with underlined mean the start site and the end site of only ITS2 region annotated in relevant known sequences from NCBI

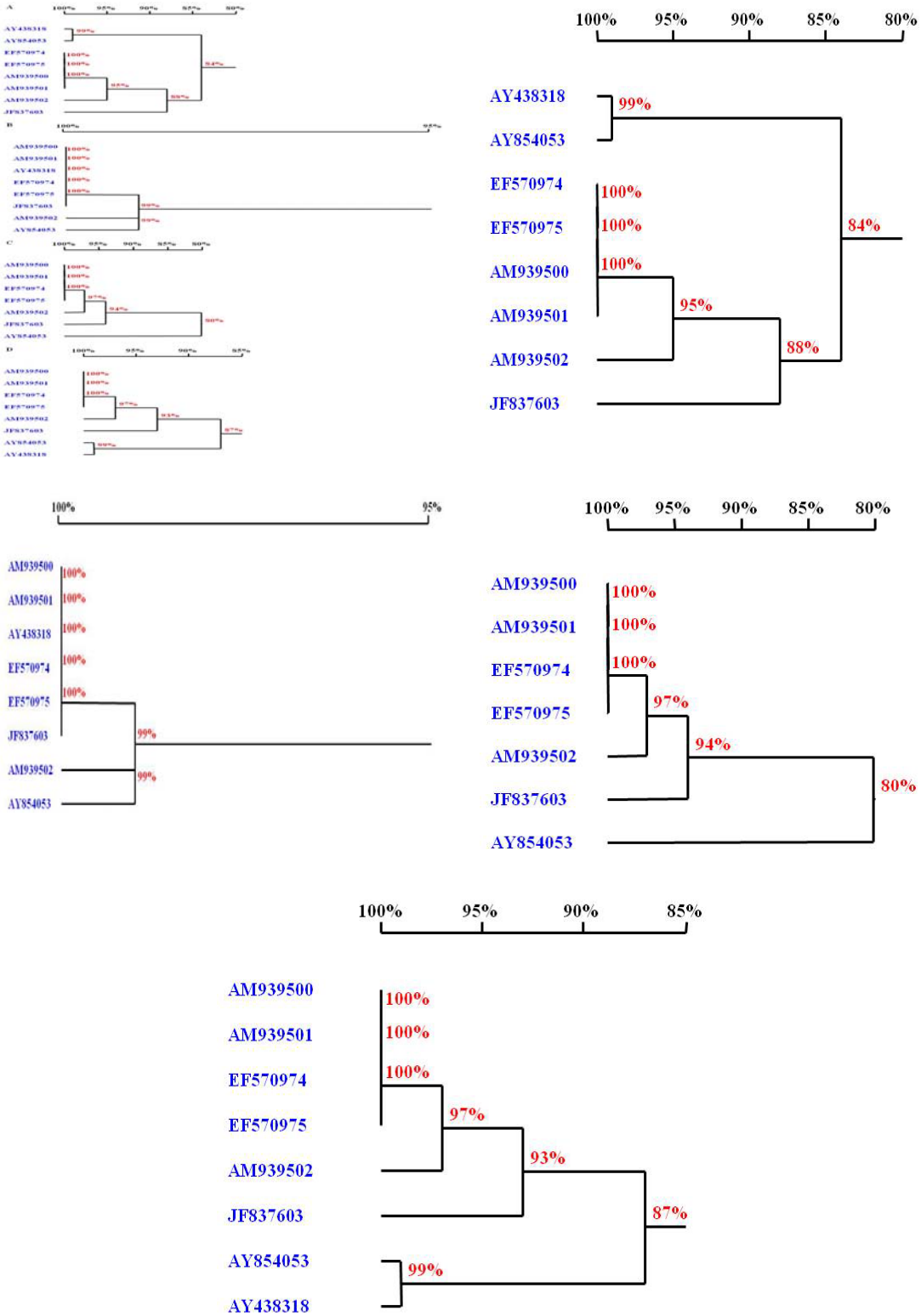


Fig. 5. Phylogenetic relationships of *H. corymbosa* among ITS1 (A), 5.8S rRNA (B), ITS2 region (C), and total ITS region (D) of known sequences from NCBI and our sequence, JF837603

Sequence analysis of known *H. corymbosa* populations

In NCBI GenBank data, only nine total ITS sequences of *H. corymbosa* were found, with the accession numbers shown in Table 4. Analyzed the lengths of different regions, ITS1, 5.8S rRNA, and ITS2 regions contained 187 or 319 bp, 157 or 163 bp, and 220 or 299-300 bp, respectively. This result suggested that the different, confused region discrimination described in *H. diffusa* ITS sequences was also appeared in *H. corymbosa* sequences. For instance, The ITS1 region in AM939500-AM939502 and our sequence, JF837603 begins from GTG and CTG with underlined, respectively, while AY438318 annotated with complete ITS1 region begins from latter TGT with underlined (Fig. 4). However, AY854053, EF570974, and EF570975 contained highly conversed TGT site that is considered as the start site of ITS1 region in AY438318, but were not annotated to be complete ITS1 region. Different region discrimination also appeared in the 5.8S rRNA and ITS2 regions. Particularly in AY438323, the start site of ITS2 region appeared later than that in other sequences, but the end site appeared earlier (Fig. 4). Furthermore, AF381484 annotated with partial 18S, ITS1, and partial 5.8S region showed obviously short length of sequence, and very low homologous arrangement when aligning DNA sequences (data not shown). Blasting of AF371484 to NCBI GenBank showed that this sequence was most similar to the complemented sequence such as partial ITS1 and 5.8S rRNA regions of AM939500, indicating that this sequence is not ready for directly DNA alignment (Fig. S1). Blasting of AY438323 to NCBI GenBank showed that this sequence was most similar to those from plants of the other family Caryophyllaceae such as *Arenaria serpyllifolia*, *Stellaria media*, indicating that the starting materials for identification may have been adulterated with other plants. Thus, AF381484 and AY438323 were ignored to use in the further sequence analysis.

To understand the phylogenetic relationship of different *H. corymbosa* populations, phylogenetic tree was separately constructed according to ITS1, 5.8S, and ITS2 region sequences. From the ITS1 phylogenetic tree, AM939500, AM939501, EF570974, and EF570975 showed very high similarity to each other, and shared 95% identification with AM939502 (Fit. 4A). These five sequences formed one monophyletic clade with our sequence, JF837603, and this clade showed 84% identification with the other monophyletic clade formed by AY438318 and AY854053 with 99 % similarity (Fig. 5A).

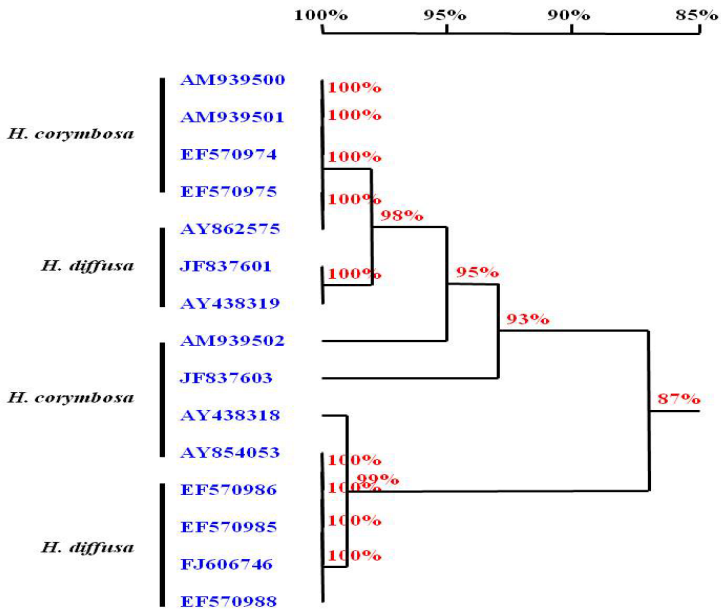


Fig. 6. Phylogenetic tree of *H. diffusa* and *H. corymbosa* using total ITS region sequences

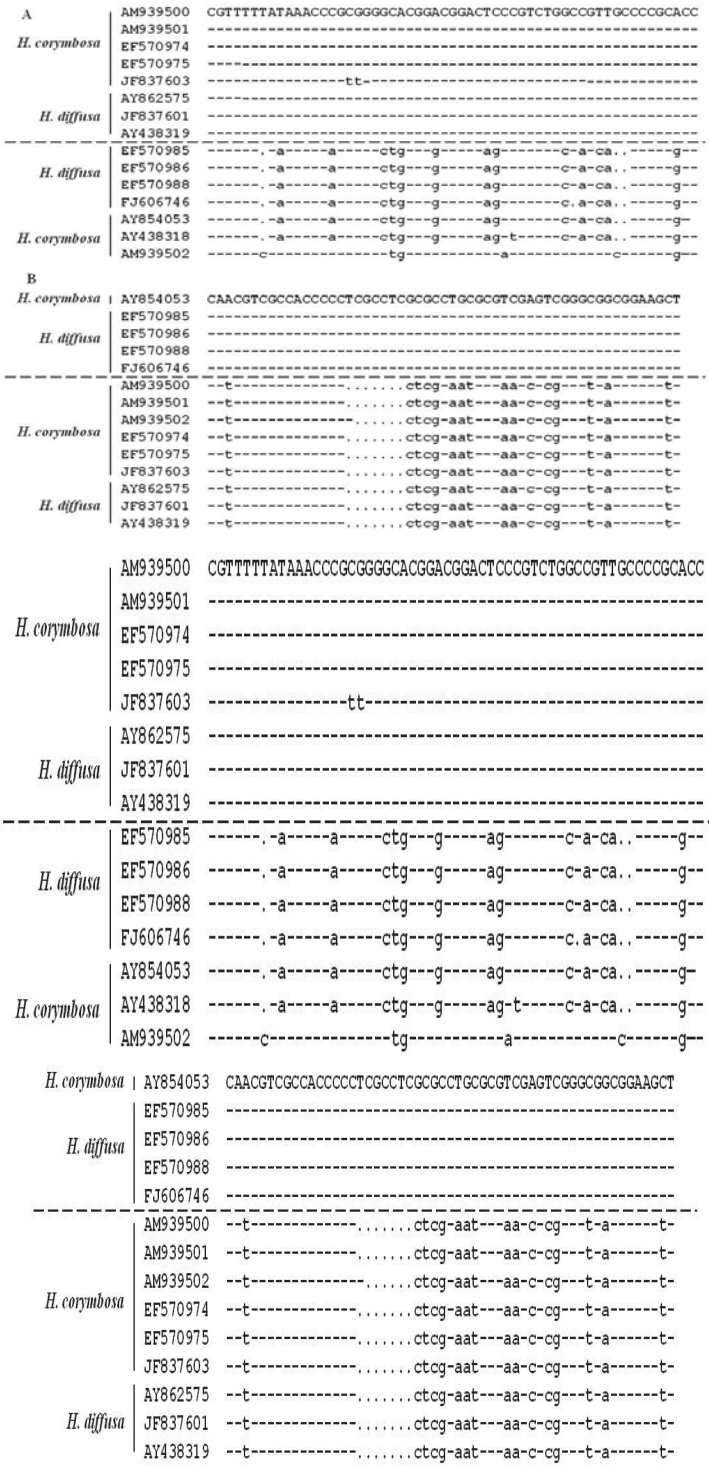


Fig. 7. DNA alignment of *H. diffusa*, *H. corymbosa* using total ITS region. Partial alignment of the total ITS regions showing notable base phlomorphisms between *H. diffusa* and *H. corymbosa* in ITS1 (A) and ITS2 (B) regions. The letters on the top line represent the base letters in sequence alignment. Sequences different from the uppermost sequence are marked as their own sequences in lowercase. '-' represents the base being identical to the uppermost sequence. '.' represents gap

In the 5.8S rRNA region, known sequences showed very high identification with each other (Fig. 5B). Except of AY854053 and AM939502, no nucleotide variation was found in the 5.8S rRNA region of other sequences. In AY854053, it not only had 6 nucleotide default, but 3 nucleotide variations, while AM939502 had 2 nucleotide variations (Fig. S2).

In the ITS2 region, AM939500, AM939501, EF570974, and EF570975 showed very high similarity of 100%, just like in the ITS1 region (Fig. 5C). AM939502, our sequence JF837603, and AY854053 had degressive homology with the group formed by AM939500, AM939501, EF570974, and EF570975 in the ITS2 phylogenetic tree, with 97%, 94%, and 80% identification, respectively (Fig. 5C). Combined with the total ITS region, the phylogenetic analysis among known *H. corymbosa* populations showed the lowest similarity of 87%, existing between AY854053 and other sequences (Fig. 5D).

In view of the influence of geographical conditions, *H. corymbosa* populations used for cloning of AM939500, AM939501, EF570974, and EF570975 sharing very high similarity, span Oceania (Australia), Africa (Gabon), and Asia (Hong Kong, China), however, the other two populations also collected from Asia (AY438318 and AY854053) showed only 87% similarity with them (Fig. 5D, Table 3). These finding suggested

```
>emb|AM939501.1| Oldenlandia corymbosa ITS1, 5.8S rRNA gene and ITS2,
specimen voucher Andersson & Nilsson 2263 (GB)

Length=781 Score = 254 bits (137), Expect = 2e-64 Identities = 140/141
(99%), Gaps = 1/141 (1%) Strand=Plus/Minus

Query 1   AGAGTCGTATTGGTTATACGATTCAGACACGTCACACCCGCCCAACGGAAACCGCCGGGG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct 330 AGAGTCGTATTGGTTATACGATTCAGACACGTCACACCCGCCCAACGGAAACCGCCGGGG 271
Query 61  GATGCGGACGATCCTTTTGTGTAGTCCTTGGCGCTTTCCGCGCCGGAAGTTTTGTTGGGT 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct 270 GATGCGGACGATCCTTTTGTGTAGTCCTTGGCGCTTTCCGCGCCGGAAGTTTTGTTGGGT 211
Query 121 GCGGGG-AACGGCCAGACGGG 140
          |||||| ||||||||||||||
Sbjct 210 GCGGGGCAACGGCCAGACGGG 190
```

Fig. S1. Blasting result of *H. corymbosa* ITS sequence, AF381484 to NCBI GenBank. Query means the sequence of AF371484, while Subject means the sequence of existing sequence, AM939501. The number following Query or Subject means the present sequence number of its relevant sequence

```
AM939500 AATACGACTCTCGGCAACGGATATCTAGGCTCTCGCATCGATGAAGAACGTAGCGAAATG
AM939501 AATACGACTCTCGGCAACGGATATCTAGGCTCTCGCATCGATGAAGAACGTAGCGAAATG
AY438318 --TACGACTCTCGGCAACGGATATCTAGGCTCTCGC------------------------
EF570974 AATACGACTCTCGGCAACGGATATCTAGGCTCTCGCATCGATGAAGAACGTAGCGAAATG
EF570975 AATACGACTCTCGGCAACGGATATCTAGGCTCTCGCATCGATGAAGAACGTAGCGAAATG
JF837603 AATACGACTCTCGGCAACGGATATCTAGGCTCTCGCATCGATGAAGAACGTAGCGAAATG
AM939502 AATACGACTCTCGGCAACGGATATCTAGGCTCTCGCATCGATGAAGAACGTAGCGAAATG
AY854053 --TACGACTCTCGGCAACGGATATCTAGGCTCTCGCATCGATGAAGAACGTAGCGAAATG

AM939500 CGATACTTGGTGTGAATTGCAGAATCCCGTGAACCATCGAGTTTTTGAACGCAAGTTGCG
AM939501 CGATACTTGGTGTGAATTGCAGAATCCCGTGAACCATCGAGTTTTTGAACGCAAGTTGCG
AY438318 ------------------------------------------------------------
EF570974 CGATACTTGGTGTGAATTGCAGAATCCCGTGAACCATCGAGTTTTTGAACGCAAGTTGCG
EF570975 CGATACTTGGTGTGAATTGCAGAATCCCGTGAACCATCGAGTTTTTGAACGCAAGTTGCG
JF837603 CGATACTTGGTGTGAATTGCAGAATCCCGTGAACCATCGAGTTTTTGAACGCAAGTTGCG
AM939502 CGATACTTGGTGTGAATTGCAGAATCCCGTGAACCATCGAGTTTTTGAACGCAAGTTGCG
AY854053 CGATACTTGGTGTGAATTGCAGAATCCCGTGAACCATCGAGTTTTTGAACGCAAGTTGCG

AM939500 CCCGAAGCCATTTGGCTGAGGGCACGCCTGCCTGGGCGTCACG
AM939501 CCCGAAGCCATTTGGCTGAGGGCACGCCTGCCTGGGCGTCACG
AY438318 ----------------------------------------
EF570974 CCCGAAGCCATTTGGCTGAGGGCACGCCTGCCTGGGCGTCACG
EF570975 CCCGAAGCCATTTGGCTGAGGGCACGCCTGCCTGGGCGTCACG
JF837603 CCCGAAGCCATTTGGCTGAGGGCACGCCTGCCTGGGCGTCACG
AM939502 CCCGAAGCCATTAGGCTGAGGGCACGTCTGCCTGGGCGTCACG
AY854053 CCCGAAGCCATTAGGCCGAGGGCACGTCTGCCTGGGCGT----
```

Fig. S2. DNA alignment of the 5.8S rRNA region of known *H. corymbosa* sequences and our sequence, JF837603

that the phylogenetic relationship of *H. corymbosa* was not absolutely correlated with the geographical information of samples.

Phylogeny of *H. diffusa* and *H. corymbosa*

Based on the respective phylogenetic analysis among *H. diffusa* and *H. corymbosa* populations, the intraspecific homology was understood well, but not the interspecific homology. To further understand the interspecific phylogenetic relationship, a phylogenetic tree was constructed using total ITS region sequences of *H. diffusa* and *H. corymbosa* (Fig. 6). All *H. diffusa* and *H. corymbosa* populations were divided into two group, with 87% of similarity. One group mainly covered *H. diffusa* populations, the other mainly covered *H. corymbosa* populations. However, AY862575, JF837601, and AY438319 of *H. diffusa* populations were suggested to be confused into the *H. corymbosa* group, while AY43818 and AY854053 of *H. corymbosa* populations were confused with the *H. diffusa* group (Fig. 6). Notably, AY862575 of *H. diffusa* collected from Taiwan, China showed nearly 100% of similarity to the main *H. corymbosa* group formed by AM939500, AM939501, EF570974, and EF570975. Similar result was found in AY854053 of *H. corymbosa* also collected from Taiwan, China, that AY854053 showed very high homology with the main *H. diffusa* group composed by EF570986, EF570985, FJ606746, and EF570988.

DISCUSSION

Due to the recognized authentication of *H. diffusa* and *H. corymbosa*, many investigations have been performed to attempt to distinguish them with each other, such as morphological characteristics, fluorescence microscopy, DNA regions[10, 12, 21]. According to the Flora Republicae Popularis Sinicae, *H. diffusa* having solitary flowers on thick and short pedicels growing from cylindrical stems showed unobvious difference from *H. corymbosa* having corymbose inflorescences on slender peduncles growing from tetragonal stems[10]. Based on the shape of transverse sections of the *H. corymbosa* stems, this medicinal herb could be easily differentiated from medicinal herbs *H. diffusa* and *H. tenelliflora* of the same genus[12]. On the basis of the molecular analysis of total ITS regions, *H. diffusa* and *H. corymbosa* were shown to be intraspecifically conserved but interspecifically variable enough to differentiate these two species, and distantly separated in the clustering analysis {shown in Fig. 1, 2 [20]}. However, their results were not congruent with our data. Although there were many informative sites appearing in ITS1 and ITS2 regions, the clear partition did not match exactly with species classification (Fig. 7). For instance, in the ITS1 region, five *H. corymbosa* and three *H. diffusa* sequences showed nearly 100% similarity, except of our sequence, JF837603, while other four *H. diffusa*, and three *H. corymbosa* sequences showed nearly 100% similarity, except of AM939502 (Fig. 7A). Our *H. diffusa* sequence (JF837601) did not match the sequences (EF570985, EF570986, and EF570988)[9], but our *H. corymbosa* sequence (JF837603) matched their sequences (EF570974 and EF570975). Both of *H. corymbosa* sequences (AM939500 and AM939501)[19] also shared the highly conserved with our sequences, however, the other sequence form this investigation (AM939502) showed relatively low similarity, with its own informative sites in ITS1 region (Fig. 7A).

In the ITS2 region, four *H. diffusa* and one *H. corymbosa* sequences were highly conserved, while three *H. diffusa* and six other *H. corymbosa* sequences were highly conserved (Fig. 7B). Our *H. diffusa* sequence (JF837601) in ITS2 region also did not match the sequences (EF570985, EF570986, and EF570988) [9], but showed very high similarity to the main *H. corymbosa* group formed by AM939500-AM939502[19], EF570974-EF570975 [9] and our sequence, JF973603. In addition, one *H. corymbosa* sequence (AY854053) from plants collected from Taiwan, China was inclined to be similar to the main *H. diffusa* ITS2 sequences. Based on the above results, the obvious inconsistency on molecular data among those reported and ours further confirmed that correct identification of the starting materials was very important.

In addition, the utility of ITS region to construct phylogeny should be in the premise of a clear ITS region discrimination. Owing to a feature-length illumination of our above descriptions in connection with the results, the importance of a clear ITS region discrimination must be most probably known by readers. In this study, although there were certain nucleotide variations between *H. diffusa* and *H. corymbosa*, the sequence variations were suggested not to be significantly

pivotal when compared with other known sequences, as our both sequences did not show their specific sequence characters (Fig. 6, 7A, 7B).

In conclusion, the total ITS sequences of *H. diffusa* and *H. corymbosa* populations collected from Korea were successfully amplified using common primers ITS4/ITS5. However, due to confused identities of starting materials and ITS region discrimination, *H. diffusa* and *H. corymbosa* could not yet be authenticated based on only ITS region sequences in the present study. To ensure proper use of herbal materials for efficient treatment and correct identification of two medicinal herbs, *H. diffusa* and *H. corymbosa*, additional markers and identification methods should be applied to cross check the identities of materials to achieve correct authentication.

ACKNOWLEDGMENTS

This work was supported by Nutraceutical Bio Brain Korea 21 Project Group, New talent Introduction Project, Ludong University (LY2012008), and Outstanding Young Scientist in Shandong Province Award Fund Project (No. BS2013SW021).

REFERENCES

1. Ahmad, R., Ali, A.M., Israf, D.A., Ismail, N.H., Shaari, K., Lajis, N.H. Antioxidant, radical-scavenging, anti-inflammatory, cytotoxic and antibacterial activities of methanolic extracts of some *Hedyotis* species. *Life Sci.,* 2005; **76**: 1953-1964.
2. Noiarsa, P., Ruchirawat, S., Otsuka, H., Kanchanapoom, T. Chemical constituents from *Oldenlandia corymbosa* L. of Thai origin. *J. Nat. Med.,* 2008; **62**: 249-250.
3. Chung, H.S., Jeong, H.J., Hong, S.H., Kim, M.S., Kim, S.J., Song, B.K., Jeong, I.S., Lee, E.J., Ahn, J.W., Baek, S.H., Kim H.M. Induction of nitric oxide synthase by *Oldenlandia diffusa* in mouse peritoneal macrophages. *Biol. Pharm. Bull.,* 2002; **25**: 1142-1146.
4. Willimott, S., Barker, J., Jones, L.A., Opara, E.I. Apoptotic effect of *Oldenlandia diffusa* on the leukaemic cell line HL60 and human lymphocytes. *J. Ethnopharmacol.,* 2007; **114**: 290-299.
5. Lin, C.C., Ng, L.T., Yang, J.J. Antioxidant activity of extracts of Peh-hue-juwa-chi-cao in a cell free system. *Am. J. Chin. Med.,* 2004; **32**: 339-349.
6. Lin, C.C., Chen, F.Y., Namba, T. Development of crude drug resources from Taiwan: pharmacognostical studies on a Chinese crude drug. *Shoyakugaku Zasshi,* 1987; **41**: 180-188.
7. The State Pharmacopoeia Committee of China. *The appendix of Chinese pharmacopoeia*. Beijing: The Chemical Industry Press, 2005; pp 22.
8. Zhao, Z.Z., Li, Y.S. Easily confused Chinese medicines in Hong Kong. Hong Kong: Chinese Medicine Merchants Association Ltd., 2007.
9. Li, M., Jiang, R.W., Hon, P.M., Cheng, L., Li, L.L., Zhou, J.R., Shaw, P.C., But, P.P. Authentication of the anti-tumor herb Baihuasheshecao with bioactive marker compounds and molecular sequences. *Food Chem.,* 2010; **119**: 1239-1245.
10. Ko, W.C.: Hedyotis. In: *Flora Republicae Popularis Sinicae* (Luo XR, ed). Beijing: Science Press, 1999; pp 32-77.
11. Xie, Z., Zhang, Y., Lu, R. Identification of herba *Hedyotis diffusa* and its confused material herba *Hedyotis pinifoliae*. *Zhong Yao Cai,* 1997; **20**: 287-290.
12. Liang, Z.T., Jiang, Z.H., Leung, K.S.Y., Zhao, Z.Z. Determination of iridoid glucosides for quality assessment of *Herba Oldenlandiae* by high-performance liquid chromatography. *Chem. Pharm. Bull.,* 2006; **54**: 1131-1137.
13. Liang, Z.T., He, M.F., Fong, W.F., Jiang, Z.H., Zhao, Z.Z. A comparable, chemical and pharmacological analysis of the traditional Chinese medicinal herbs *Oldenlandia diffusa* and *O. corymbosa* and a new valuation of their biological potential. *Phytomedicine,* 2008; **15**: 259-267.
14. Lee, H.Z., Bau, D.T., Kuo, C.L., Tsai, R.Y., Chen, Y.C., Chang, Y.H. Clarification of the phenotypic characteristics and anti-tumor activity of *Hedyotis diffusa*. *Am. J. Chin. Med.,* 2011; **39**: 201-213.
15. Liang, Z.T., Jiang, Z.H., Leung, K.S.Y., Peng, Y., Zhao, Z.Z. Distinguishing the medicinal herb *Oldenlandia diffusa* from similar species of the same genus using fluorescence microscopy. *Microsc. Res. Tech.,* 2006; **69**: 277-282.
16. Soltis, D.E., Soltis, P.S.: Choosing an approach and an appropriate gene for phylogenetic analysis. In: *Molecular systematic of plants II-DNA sequencing* (Soltis DE, Soltis PS, Doyle JJ, eds). Boston, Dordrecht, London: Kluwer Academic Publishers, 1998; pp 1-42.
17. Schindel, D.E., Miller, S.E. DNA barcoding a useful tool for taxonomists. *Nature,* 2005; **435**: 17.

18. Kress, W.J., Erickson, D.L. DNA barcodes: genes, genomics, and bioinformatics. *Proc. Natl. Acad. Sci. USA*, 2008; **105**: 2761-2762.
19. Kårehed, J., Groeninckx, I., Dessein, S., Motley, T.J., Bremer, B. The phylogenetic utility of chloroplast and nuclear DNA markers and the phylogeny of the Rubiaceae tribe Spermacoceae. *Mol. Phylogenet. Evol.*, 2008; **49**: 843-866.
20. Guo, X., Simmons, M.P., But, P.P.H., Shaw, P.C., Wang, R.J. Application of DNA barcodes in *Hedyotis L.* (Spermacoceae, Rubiaceae). *J. System. Evol.*, 2011; **49**: 203-212.
21. Coleman, A.W. ITS2 is a double-edged tool for eukaryote evolutionary comparisons. *Treads Genet.*, 2003; **19**: 370-375.
22. Schultz, J., Maisel, S., Gerlach, D., Müller, T., Wolf, M. A common core of secondary structure of the internal transcribed spacer 2 (ITS2) throughout the Eukaryota. *RNA*, 2005; **11**: 361-364.
23. Hao, M.G., Liu, Z.Q., Wang, J.L. Application of the sequences of rDNA ITS to identify Chinese crude drug *Hedyotis diffusa*. *J. Anhui Normal Univ. (Nat. Sci.)*, 2004; **27**: 188-191.
24. Liu, Z.Q., Hao, M.G., Wang, J.L. Application of allele-specific primer in the identification of *H. diffusa*. *Zhong Yao Cai*, 2004; **27**: 484-487.
25. Liu, Z.Q., Hao, M.G. Determination of rDNA ITS sequences in *Hedyotis diffusa*. *Shaanxi J. Trad. Chin. Med.*, 2005; **26**: 167-169.
26. Chiou, S.J., Yen, J.H., Fang, C.L., Chen, H.L., Lin T.Y. Authentication of medicinal herbs using PCR-amplified ITS2 with specific primers. *Plant. Med.*, 2007; **73**: 1421-1426.
27. Andersson, L., Rova, J.H.E., Guarin, F.A. Relationships, circumscription, and biogeography of Arcytophyllum (Rubiaceae) based on evidence from cpDNA. *Brittonia*, 2002; **54**: 40-49.
28. Robbrecht, E., Manen, J.F. The major evolutionary lineages of the coffee family (Rubiaceae, angiosperms). Combined analysis (nDNA and cpDNA) to infer the position of *Coptosapelta* and *Luculia*, and supertree construction based on *rbc*L, *rps*16, *trnL-trnF* and *atp*B-*rbc*L data. A new classification in two subfamilies, Cinchonoideae and Rubioideae. *System. Geogra. Plant.*, 2006; **76**: 85-146.
29. Groeninckx, I., Dessein, S., Ochoterena, H., Persson, C., Motley, T.J., Kårehed, J., Bremer, B., Huysmans, S., Smets, E. Phylogeny of the herbaceous tribe Spermacoceae (Rubiaceae) based on plastid DNA data. *Ann. Mo. Bot. Gard.*, 2009; **96**: 109-132.
30. Muhammad, I., Zhang, T.T., Wang, Y., Zhang, C.Y., Miao, Q., Zhang, L.J., Lin, F. Modification of CTAB protocol for maize genomic DNA extraction. *Res. J. Biotechnol.*, 2013; **8**: 41-45.
31. White, T.J., Bruns, T., Lee, S., Taylor, J.: 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, White TJ, eds). San Diego, Calif: Academic Press, 1990; pp 315-322.
32. Church, S.A. Molecular phylogenetics of Houstonia (Rubiaceae): descending aneuploidy and breeding system evolution in the radiation of the lineage across North America. *Mol. Phylogenet. Evol.*, 2003; **27**: 223-238.