Plant Omics Journal POJ

POJ 6(5):325-332 (2013) ISSN:1836-3644

Applying DNA barcoding to identify *Nervilia fordii* and six congeneric species

Qionglin HUANG, Lingling LIANG, Rui HE*, Xinye MA, Ruoting ZHAN, Weiwen CHEN*

Research Center of Chinese Medicinal Resource Science and Engineering, Guangzhou University of Chinese Medicine (GZUCM), Guangzhou 510006, P.R. China
Key Laboratory of Chinese Medicinal Resource from *Lingnan*, Ministry of Education, Guangzhou 510006, P.R. China

***Corresponding author: perfecthql@163.com (W. W. C.)/rayhe618@hotmail.com (R. H.)**

Abstract

DNA barcoding is a novel technique which enables species identification and taxonomic classification using a short and standard DNA sequence. In this study, traditional herbal medicine *Nervilia fordii* and its congeneric species included all three distributed in mainland China and three acquired from Genbank were sampled. Four popular candidate DNA barcodes, ITS2, *rbcL*, *matK*, and *LSU* D1-D3, were amplified and sequenced with corresponding pairs of universal primers. The PCR amplification and sequencing efficiency, intra- and interspecific variation and barcoding gap were used to evaluate each locus, and the success rate of correct identification was assessed by the BLAST 1 method. The sequencing efficiency of *rbcL* and *LSU* D1-D3 was 100%, while that of *matK* and ITS2 was 95.8% and 91.4%, respectively. Although relatively lower contribution in sequencing efficiency and interspecific divergence, *matK* exhibited significant barcoding gap and discriminated all the given species in our study. ITS2 had the largest variation in both intra- and interspecific level, which enabled classification at the genus level. Our findings showed that DNA barcoding is an efficient biomarker for identification of *N. fordii* and its congeners and that *matK* gene is a potential DNA barcode for *Nervilia*.

Keywords: *Nervilia fordii*; congeners; DNA barcode; *matK*.
Abbreviation: CBOL- the Consortium for the Barcode of Life; ITS2-the second internal transcribed space; *LSU*-ribosomal RNA large subunit gene; *matK*-maturase K; NCBI-National Center for Biotechnology Information; *rbcL*- ribulose-1, 5-bisphosphate carboxylase/oxygenase large subunit.

Introduction

Morphological traits are classically the principle factors for species authentication. However, traditional morphological identification may require the expertise of experienced professionals (Song et al., 2009). In addition, some discriminating features may only be present during a particular growth period, such as during flowering or seed setting. DNA barcode is a short standard DNA sequence from the genome that can be used for species identification accurately and quickly (Hebert and Gretory, 2005). DNA barcoding has huge potential for species identification and can overcome the limitations of traditional taxonomy (Chen et al., 2007). In animals, mitochondrial cytochrome c oxidase I (*COI*) gene has been acknowledged as the standard DNA barcode (Hebert et al., 2003a; Hebert et al., 2003b; Hajibabael et al., 2006). In plants, several regions in plastid and nucleus and their combinations have previously been recommended. A plastid group of *rbcL* and *matK* was proposed as core land plant barcode by CBOL (CBOL Plant Working Group, 2009). Another plastid interspace *psbA-trnH* was also severed as a candidate barcode (Kress and Erickson, 2007; Chase et al., 2005; Kress et al., 2005) and widely used in species discrimination of the *Dendrobium* (Yao et al., 2009) and *Cistanche* genus (Han et al., 2010) and medicinal pteridophyta (Ma et al., 2010). Nuclear loci always occupied important position in plant barcodes selection. ITS2, a part of ribosome internal transcribed space, exhibited 92.7% of identification efficiency at the species level in more than 6,600 plant specimens from 753 genera, and was suggested to be a universal DNA barcoding for medicinal plants (Chen et al., 2010). Additionally, *LSU* D1-D3 showed the strongest discriminatory power among six loci include 18S, ITS1, 5.8S, ITS2, *LSU* D1-D3 and *matK* in 159 medicinal plants from South China, and the combination of *LSU* D1-D3 and *matK* was advised as a potential DNA barcode for Angiosperm (Lab data, unpublished). However, a consensus of suitable DNA barcode for plant materials remains unformed due to less species number and low identification efficiency in closely related species. Therefore, DNA barcodes chosen to identify plant specimens varies from species to species. *Nervilia fordii* is a famous and valuable herb that has been used as a traditional Chinese medicine called “Qingtiankui” in South China for its significant effect in curing pulmonary disease (Mei, 2011). Rigid requirements of germination and growth limited the cultivation and production of *N. fordii*, and the wild resource of *N. fordii* becoming gradually endangered (Wen et al., 1993). Driven by economic interest, several species from *Nervilia* that show very similar morphology to *N. fordii* are intentionally or unintentionally used in place of *N. fordii* in cultivation, circulation and clinical administration, such as *N. aragoana*. The key distinguishing features of *N. fordii* and *N. aragoana* are flower characteristics; however, florescence is extremely short in this genus, and some years plants do not flower at all. Even worse, *N. aragoana* was chaotically cultivated with *N. fordii* by herbalists. Another species *N. plicata* has already been mistaken as the herbal medicine Qingtiankui due to

Table 1. Primers and reaction condition used in this study.

Candidate barcode	Name of primer	Sequence of primer 5'-3'	Reaction condition	Literatures
ITS2	ITS2F ITS3R	ATGCGATACTTGGTGTGAAT GACGCTTCTCCAGACTACAA	94 °C 2 min; 94 °C 30 sec, 52 °C 30 sec, 72 °C 1 min, 32 cycles; 72 °C 5 min	Chen et al., 2010
rbcL	rbcLa-F rbcLa-R	ATGTCACCACAAACAGAGACTAAAGC GTAAAATCAAGTCCACCRCG	94 °C 3 min; 94 °C 30 sec, 56 °C 30 sec, 72 °C 50 sec, 32 cycles; 72 °C 7 min	CBOL working group, 2010
matK	M3F M1R	CGTACAGTACTTTTGTGTTTACGAG ACCCATGCCATCTGGAAATCTTGGTTC	94 °C 3 min; 94 °C 30 sec, 62 °C 30 sec, 72 °C 1 min, 33 cycles; 72 °C 5 min	CBOL working group, 2010
LSU D1-D3	Duan 5 Duan 6	TAGTAACGGCGAGCGAAC GGCATAGTTCACCATCTTTC	94 °C 2 min; 94 °C 30sec, 58 °C 30 sec, 72 °C 50 sec, 35 cycles; 72 °C 7 min	Lab data, unpublished

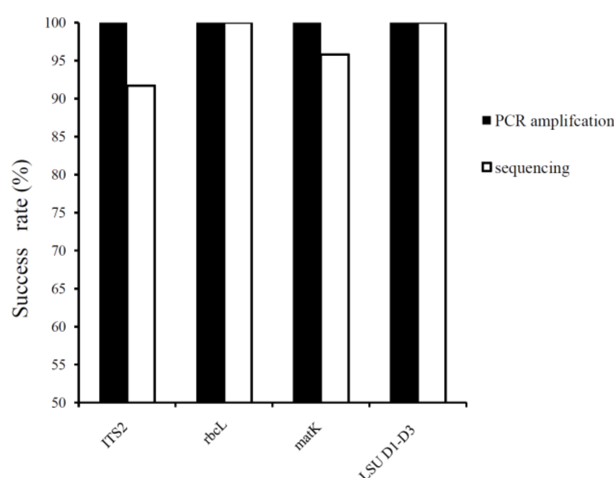

Fig 1. Amplification and sequencing efficiency of each candidate locus. The PCR and sequencing success rate of all four regions exceed 90%, which preliminary illustrated their feasibility in selection of potential biomarkers for identification between *N. fordii* and its congeners.

extraordinarily similar morphological traits of dry materials. This situation is far from ideal for resource management of *N. fordii* and causes clinical insecurity. Therefore, it is urgent and necessary to establish an easy and precise method to differentiate the *Nervilia* species.The present study aimed to establish an identification method for *N. fordii* using DNA barcoding, and provide scientific foundations for resource protection and safe usage of *N. fordii*. Species selected for this study were from the *Nervilia* genus which is intentionally or unintentionally confused with *N. fordii*. Herein, the feasibility of four popular candidate barcodes (ITS2, *rbcL*, *matK*, *LSU* D1-D3) in distinguishing plants within the *Nervilia* genus was evaluated in our study.

Results

Universality

Efficiency of amplification and sequencing were critical for evaluation of the universality of the DNA barcodes in question. All barcodes exhibited complete PCR success (100%), which indicated the suitability of the primers selected for this study (Table 1.). The sequencing success rate of the two candidate fragments *rbcL* and *LSU* D1-D3 was 100%, while the sequencing efficiency of *matK* and ITS2 were 95.8% and 91.4%, respectively (Fig 1.). The results preliminarily illustrated that the four barcodes tested were suitable candidate barcodes in *Nervilia* species.

Intraspecific variation and interspecific divergence

The intraspecific variation - that is, the within-species variance - for four candidate barcodes (ITS2, *rbcL*, *matK*, *LSU* D1-D3) was analyzed in four populations of *N. fordii* and six other single populations of *Nervilia* species (Table 2.). The acquired sequences of each region ranged in size from 242 bp for ITS2 to 785 bp for *matK*; each candidate barcode showed robust amplification of the same size fragment across all species and all populations, except for ITS2 from *N. mackinnonii*, which was 247 bp in length. For *N. fordii,* all barcode sequences showed 0.0% variability both within population and between populations, except *rbcL* which showed 0.2% variation in sequence between the four populations. In contrast, the intraspecific variation within the 242bp ITS2 region from the same population of *N. aragoana* was 0.4%, and from the single population of *N. mackinnonii*, sequence variance was 7.2% (Table 3.). At the interspecific level – that is, between the seven different species of *Nervilia*-, the highest level of average interspecific divergence was apparent in candidate barcodes from nuclear regions, compared to plastid regions. ITS2 exhibited the highest level of average interspecific divergence, and *rbcL* the lowest (Table 4.). These data were supported by the Wilcoxon signed rank test (Table 5, 6.).

Table 2. Collection information and their Genbank accession numbers of the samples in this study.

Species	Source	Population	Voucher number	Genbank accession number			
				ITS2	*rbcL*	*matK*	*LSU* D1-D3
Nervilia fordii (Hance) Schltr.	Jingxi, Guangxi	1	NFJX01	JX865481	JX865457	JX865503	KC181136
	Jingxi, Guangxi	1	NFJX02	JX865482	JX865458	JX865504	KC181137
	Napo, Guangxi	2	NFNP01	JX865483	JX865459	JX865505	KC181138
	Napo, Guangxi	2	NFNP02	JX865484	JX865460	JX865506	KC181139
	Mashan, Guangxi	3	NFMS21	JX865485	JX865461	JX865507	KC181140
	Mashan, Guangxi	3	NFMS22	JX865486	JX865462	JX865508	KC181141
	Mashan, Guangxi	3	NFMS23	JX865487	JX865463	JX865509	KC181142
	Mashan, Guangxi	4	NFMS31	JX865488	JX865464	JX865510	KC181143
	Mashan, Guangxi	4	NFMS32	JX865489	JX865465	JX865511	KC181144
	Mashan, Guangxi	4	NFMS33	JX865490	JX865466	JX865512	KC181145
Nervilia aragoana Gaud.	Tian’e, Guangxi	1	NAGX01	JX865491	JX865467	JX865513	KC181146
	Tian’e, Guangxi	1	NAGX02	JX865492	JX865468	JX865514	KC181147
	Tian’e, Guangxi	1	NAGX03	JX865493	JX865469	JX865515	KC181148
	Tian’e, Guangxi	1	NAGX04		JX865470	JX865516	KC181149
	Tian’e, Guangxi	1	NAGX05		JX865471		KC181150
	GZUCM	2	NAGZ01	JX865494	JX865472	JX865517	KC181151
Nervilia plicata (Andr.) Schltr.	Napo, Guangxi	1	NPNL01	JX865495	JX865473	JX865518	KC181152
	Napo, Guangxi	1	NPNL02	JX865496	JX865474	JX865519	KC181153
Nervilia mackinnonii (Duthie) Schltr.	Napo, Guangxi	1	NMNP01	JX865497	JX865475	JX865520	KC181154
	Napo, Guangxi	1	NMNP02	JX865498	JX865476	JX865521	KC181155
	Napo, Guangxi	1	NMNP03	JX865499	JX865477	JX865522	KC181156
	Napo, Guangxi	1	NMNP04	JX865500	JX865478	JX865523	KC181157
	Napo, Guangxi	1	NMNP05	JX865501	JX865479	JX865524	KC181158
	Napo, Guangxi	1	NMNP06	JX865502	JX865480	JX865525	KC181159
Nervilia crociformis (Zollinger&Moritzi) Seidenfaden	Genbank			JN114607	JN005560	JN004500	
				JN114608	JN005561	JN004501	
				JN114609	JN005562	JN004502	
Nervilia gammieana (Hook. f.) Pfitzer	Genbank			JN114613	JN005568	JN004505	
				JN114614	JN005569	JN004506	
				JN114615	JN005570	JN004507	
Nervilia infundibulifolia Blatt.& McCann	Genbank				JN005572	JN004510	

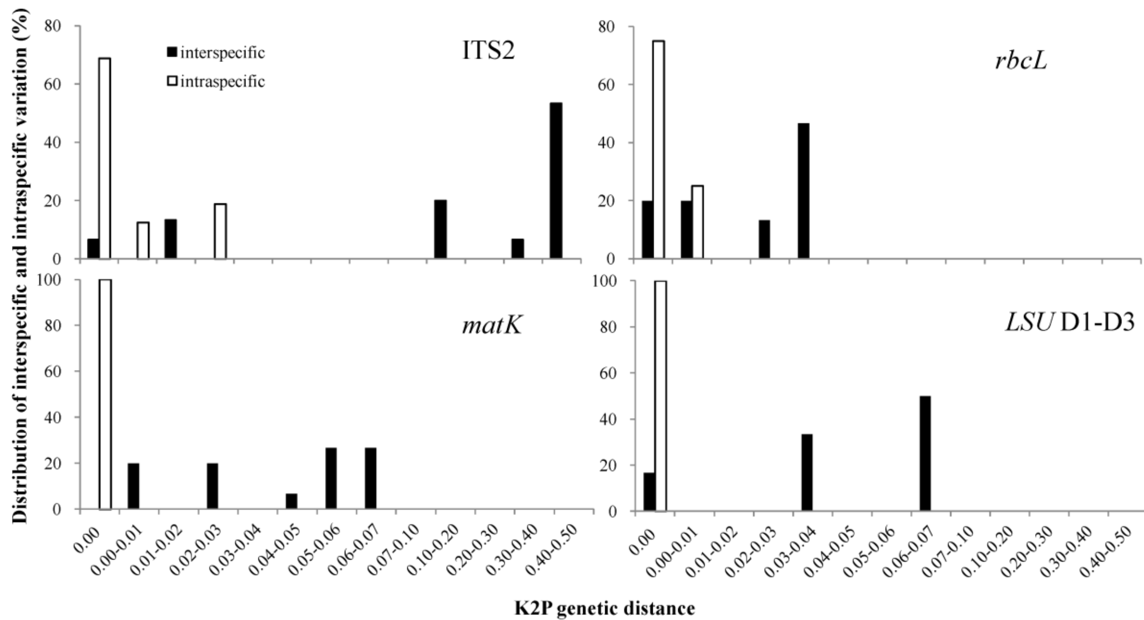

Fig 2. Distributions of intra- and interspecific variation for four candidate regions. A significant spacer between the distributions of two variations called barcoding gap was benefit for Nervilia species identification and matK gene was the sole loci exhibited barcoding gap among four regions.

Assessment of barcoding gap

A candidate locus for DNA barcoding should be present in the majority of taxa of interest, and not require species-specific PCR primers (Chen et al., 2007). The DNA sequence amplified by these primers should then ideally identify a larger variation between species (interspecific), but a small amount of variation within species (intraspecific). An appropriate DNA barcode should exhibit significant "Barcoding gap", defined as spacer region between the distribution ranges of intra- and interspecific variations (Meyer and Paulay, 2005). When an overlap between intra- versus interspecific divergence occurs, whether the variations were within specie or between species would be uncertain, and the sequences cannot fully distinguish the species sampled. Distribution of intra- and interspecific variation was investigated to see whether such a gap existed in four loci when used as DNA barcode for *Nervilia*. From fig 2., a significant barcoding gap was observed in the plastid gene *matK* for non-overlaps found in the ranges of two kinds of variation; while another gene *rbcL* located in chloroplast DNA (cpDNA) revealed the most overlap. Although the nuclear loci ITS2 and *LSU* D1-D3 contributed the most intraspecific divergence for species identification, barcoding gap were not found in both of them when used to identify *Nervilia*, the overlaps with the threshold value of distribution of intra- and interspecific divergence ranged from 0.00 to 0.03, and at 0.00, respectively.

Species identification efficiency of each potential barcode

The BLAST 1 method was used to test the identification feasibility of the candidate barcodes at the genus and species level. The method evaluated the efficiency using sequence alignment by nucleotide blast tool in NCBI. When the query sequence of a species shares the highest identify with the same species submitted in Genbank only, the sequence distinguish the species at species level; when match with different species from the same genus, the sequence distinguish the species at genus level; when match with the species from other genus, then the identification failed (Ross et al., 2008). As shown in Fig. 3, four candidate barcodes – nuclear encoded *ITS2* and *LSU*, and plastid encoded *rbcL* and *matK*- identified all the species at genus level. The plastid region *matK* was capable to fully distinguish at the species level, for all species used in this study. The nuclear gene ITS2 also identified the four species spread in Chinese mainland with larger variation, but ITS2 sequence homology exists between *N. fordii* and *N. gammieana*. The remaining regions, *rbcL* from the plastid and *LSU* D1-D3 from the nucleus, showed the lowest discriminatory power, distinguishing 28.6% and 50%, respectively.

Phylogenetic analysis

Based on interspecific variability of the *matK* gene sequence, a phylogenetic tree was constructed, clustering seven *Nervilia* species into two groups (Fig 4.). *N.mackinnonii* showed the furthest divergence from the other *Nervilia* species and formed a group alone. In the other branch, *N. fordii*, *N. gammieana* and *N. aragoana* clustered into a subgroup, *N. plicata* and *N. infundibulifolia* formed another subgroup, and then clustered with *N. crociformis*. We propose *matK* gene, which with significant barcoding gap and strong discriminatory power, to identify *N. fordii* and its congeneric species and to be a potential barcode for *Nervilia*.

Discussion

Assessment of the feasibility of four candidate barcodes

An appropriate DNA barcode should be amplified and sequenced with single pair of primers, and high-quality sequences are obtained by bidirectional sequencing. The sequence obtained must possess adequate interspecific diversity to distinguish different species but have low levels

Table 3. Length and intraspecific variation of four candidate regions.

Species (Number of individuals samples)	ITS2, Length (bp)/% variation	*rbcL,* Length (bp)/% variation	*matK,* Length (bp)/% variation	*LSU* D1-D3, Length (bp)/% variation
Nervilia. fordii (10)				
Population1 (2)	242/0.0	502/0.0	785/0.0	637/0.0
Population2 (2)	242/0.0	502/0.0	785/0.0	637/0.0
Population3 (3)	242/0.0	502/0.0	785/0.0	637/0.0
Population4 (3)	242/0.0	502/0.0	785/0.0	637/0.0
Between populations(4)	242/0.0	502/0.2	785/0.0	637/0.0
Nervilia aragoana (6)				
Population1 (5)	242/0.4	502/0.0	785/0.0	637/0.0
Between population (2)	242/0.4	502/0.0	785/0.0	637/0.0
Nervilia plicata (2)	242/0.0	502/0.0	785/0.0	637/0.0
Nervilia mackinnonii (6)	247/7.2	502/0.0	785/0.0	637/0.0
Nervilia crociformis(3)	242/0.0	502/0.0	785/0.0	---
Nervilia gammieana (3)	242/0.0	502/0.0	785/0.0	---

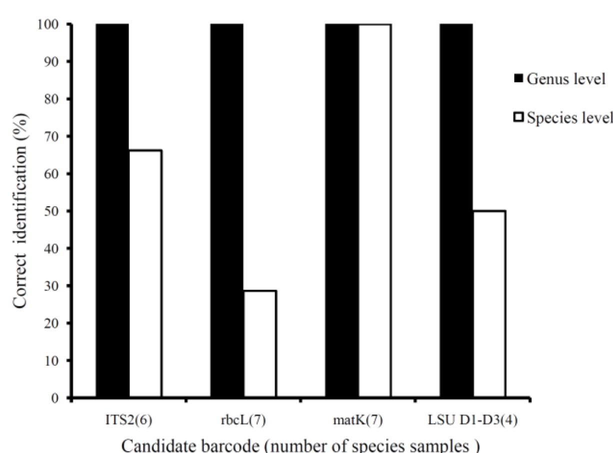

Fig 3. Identification efficiency of four loci using the BLAST 1 method. All the candidate barcodes were capable to distinguish *Nervilia* species from the other genus, however, only *matK* gene fully identified seven *Nervilia* species.

of intraspecific variation so as demonstrate homology within species (Song et al., 2009; Yao et al., 2009). The compressive analysis of amplification and sequencing efficiency, barcoding gap and success rate of correct identification demonstrated that *matK* gene had larger advantage than other loci. As a core plant barcode recommended by CBOL (CBOL Plant Working Group, 2009), *matK* gene is one of coding genes with high evolution rate in plastid DNA and characterized by universal usage, easy amplification and alignment (Gadek et al., 2000; Lahaye et al., 2008). Although relatively low contribution in sequence efficiency and interspecific divergence, *matK* exhibited significant barcoding gap (Fig 2.) and the highest discriminatory power to differentiate all given species in our study (Fig 3.). In the present study, ITS2 possessed remarkably higher intra- and interspecific variations than other loci and had the second higher discriminatory power (66.7%, followed *matK*), demonstrating that it is beneficial for species authentication in *Nervilia* to a certain degree. Another nuclear region *LSU* D1-D3 exhibited second higher average interspecific divergence, but it did not succeed in identifying the four species spread in Chinese mainland, because of the consistent sequence between *N. fordii* and *N. aragoana*. Both the nuclear genes got high evaluation in species identification with high interspecific variation in previous study (Chen et al., 2010; Lab data, unpublished); similar results were conformed in our study. The second chloroplast gene in our study, *rbcL*, showed 100% amplification efficiency, however, *rbcL* contributed least divergence and most overlap between intra- and interspecific levels. Newmaster *et al.* (2006) compared approximately 10,300 *rbcL* sequence in Genbank using a distance based method, and found that *rbcL* was incapable of distinguishing between all species, and its power was limited at species level. The identification rate at the genus level was 100%, however at the species level (28.6%) *rbcL* is considered too low to be used as a DNA barcode for identification of *Nervilia* species. *psbA-trnH* has also been proposed to be a strong candidate DNA barcodes for plant materials (Kress et al., 2005; Shaw et al., 2007). However, good-quality *psbA-trnH* sequences were not acquired from *Nervilia* specimens in our study, attributed to mononucleotide repeats interrupting the sequencing. We speculated that the primers cited from the previous study (Tate and Simpson, 2003) to amplify this locus were not specific enough for the genus. Therefore, this region was not included for further analysis and is not considered suitable choice for DNA barcoding identification in *Nervilia*.

Application of DNA barcodes in discrimination and genetic relationship of Nervilia species

Conventional identification methods cannot solve all species determination problems in *Nervilia*, such as distinguishing between *N. fordii* and *N. aragoana* based on morphology. Confusion in *Nervilia* species had been seriously impeded resource protection and clinic application of *N. fordii*. Our

Table 4. Length and interspecific divergence of four candidate regions.

	ITS2	*rbcL*	*matK*	*LSU* D1-D3
Length (bp)/% variation (number of species sampled)	242-247/1.3, 17.2, 43.7, 0.0,45.9, 17.8, 44.6, 1.3,46.8, 43.2, 17.2, 43.8, 43.7, 36.1, 15.9 (6)	502/0.0,1.0, 3.7, 3.1, 0.0, 1.0, 1.0, 3.7, 3.1, 0.0, 1.0, 3.5, 2.8, 1.0, 0.0, 2.6, 3.7, 3.5, 3.1, 2.8, 1.0 (7)	785/0.3, 2.3, 6.7, 5.3, 1.2, 2.1, 2.3, 6.7, 5.3, 0.5, 2.1, 5.7, 4.6, 2.3, 0.2, 6.6, 6.9, 5.5, 5.5, 4.5, 2.1 (7)	637/0.0, 3.6,6.9, 3.6, 6.9, 6.8 (4)
Mean of interspecific divergence %	22.9	2.0	3.7	4.6

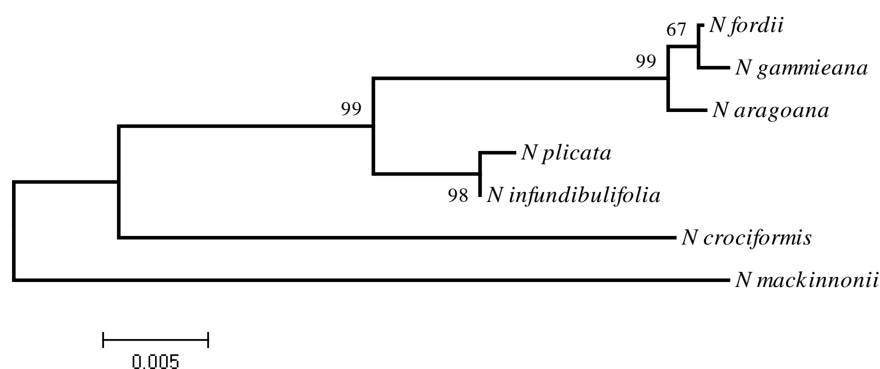

Fig 4. NJ phylogenetic tree constructed based on *matK* gene from *Nervilia*. The lengths of the bar stand for the genetic distance among species, while species clustered together means their closer genetic relationship.

Table 5. Wilcoxon signed rank test of interspecific variations among the four loci.

W_+	W_-	Inter Relative Ranks, n, P value	Result
ITS2	*rbcL*	W_+=150.00，W-=0.00，n=21，P≤0.001	ITS2>*rbcL*
ITS2	*matK*	W_+=119.00，W-=1.00，n=21，P≤0.001	ITS2>*matK*
ITS2	*LSU* D1-D3	W_+=171.00，W-=0.00，n=10，P≤0.000	ITS2>*LSU* D1-D3
matK	*rbcL*	W_+=120.00，W-=0.00，n=28，P≤0.001	*matK*> *rbcL*
LSU D1-D3	*matK*	W_+=147.00，W-=24.00，n=10，P≤0.007	*LSU* D1-D3>*matK*
LSU D1-D3	*rbcL*	W_+=120.00，W-=0.00，n=10，P≤0.001	*LSU* D1-D3>*rbcL*

Table 6. Wilcoxon signed rank test of the intraspecific divergences among the four loci.

W_+	W_-	Intra Relative Ranks, n, P value	Result
ITS2	*rbcL*	W_+=35.00，W-=10.00，n=16，P≤0.134	ITS2=*rbcL*
ITS2	*matK*	W_+=15.00，W-=0.00，n=16，P≤0.041	ITS2>*matK*
ITS2	*LSU* D1-D3	W_+=15.00，W-=0.00，n=16，P≤0.041	ITS2>*LSU* D1-D3
matK	*rbcL*	W_+=0.00，W-=10.00，n=16，P≤0.046	*rbcL* >*matK*
LSU D1-D3	*matK*	W_+=0.00，W-=0.00，n=16，P≤1.000	*LSU* D1-D3=*matK*
LSU D1-D3	*rbcL*	W_+=0.00，W-=15.00，n=16，P≤0.041	*rbcL*> *LSU* D1-D3

study found that *matK* gene and ITS2 sequence were both capable of distinguishing the four species of *Nervilia –N. fordii*, *N. aragoana*, *N. plicata*, *N. mackinnonii* – that grow in mainland China. The phylogenetic tree constructed based on *matK* gene, which exhibited high reliability in previous research (Hilu et al., 2003), illustrated that *N. fordii* had extraordinarily close relationship with *N. aragoana* and *N. plicata,* two adulterants of *N. fordii*.

Materials and methods

Collection of plant materials

Twenty-four specimens were collected from Guangxi Province or cultivated in GZUCM, and the authentication was fulfilled by Ruoting Zhan in GZUCM and Yunfeng Huang in Guangxi institute of Chinese Medicine and Pharmaceutical Science, China. All the specimens were maintained in Research Center of Chinese Medicinal Resource Science and Engineering, GZUCM (Table 2).

DNA extraction, amplification and sequencing

Genomic DNA was extracted from leaves kept in silica gel according to the manual protocols of Plant Genomic DNA Kit (Tiangen, China). The primers and reaction procedures for each region are listed in Table 1. PCR amplification was carried out in100µL reaction mixture containing 10µLof 10×*Ex Taq* Buffer, 7.5µL of 10mM dNTPs, 2.5µL of 10µM upstream primer, 2.5µL of 10µM downstream primer, 3.0U of *ExTaq* DNA polymerase (Takara, Japan) and approximately 75ng of template DNA. The PCR products

were inspected on 1.0% agarose gels and visualized using INFINITY-1000 UV imaging system (Vilber Lourmat, France), and then purified using Multifunction DNA Purification Kit (Bioteke, China). The purified products were sequenced bidirectionally on 3730XL Analyzer (Applied Biosystems, USA).

Sequence alignment and analysis

The candidate sequences of three other *Nervilia* species (*N. crociformis*, *N. gammieana*, and *N. infundibulifolia*) were downloaded from Genbank. Genbank accession numbers of candidate DNA barcodes used in this study were embodied in Table 1. Sequence quality was assessed by CodonCode Aligner V2.06 (CodonCode, USA). Sequence assembly was accomplished with DNAMAN V6.0.3.40 (Lynnon, Canada). In order to obtain clean ITS2 sequences, the conserved 5.8S and 26S regions was removed from the sequences amplified with the primes for ITS2 region based on Hidden Markov Model (Keller et al., 2009). Multiple alignments were conducted using CLUSTALX. Intra- and interspecific genetic distances were calculated based on Kimura 2-Parameter model in MEGA 5 (Tamura et al., 2007). Wilcoxon signed rank test were performed to compare intra- and interspecific divergence of each barcode by SPSS 16.0 software (Meier et al., 2006). The barcoding gap was evaluated based on the distribution of intra- and interspecific variation. The BLAST1 method was used to assess identification efficiency of each candidate barcode. To further understand the genetic relationship among *Nervilia* species, phylogenetic tree was created using neighbor-joint (NJ) model in MEGA 5.0. Bootstrap values were calculated with 1000 random addition sequence replicates.

Conclusion

In this study, we demonstrated that *matK* can be used as an effective biomarker to identify *N. fordii* and its related species. *matK* revealed the diversity and phylogenetic relationship among the species we tested, and can be used reliably to determine species identity, particularly in the absence of other defining traits. These findings will help in solving the resource depletion and safe usage of *N. fordii*. Further studies will further verify the value of *matK* as a DNA barcode for *Nervilia*.

Acknowledgements

This work was supported by the Research Fund of the Doctoral Program of Higher Education, Ministry of Education of China (NO.200805720004) granted to W.W. C. and Scientific Research Foundation for Returned Scholars, Ministry of Education of China (NO.[2009]1001) granted to R. H.. We thank Mr. Yunfeng Huang from Guangxi institute of Chinese Medicine and Pharmaceutical Science, PR China for his kind help with sample collection and authentication. We also expressed heartfelt thanks to Dr. Georgina E Drury for her critical advice.

References

CBOL Plant Working Group (2009) A DNA barcode for land plants. Proc Natl Acad Sci USA. 106: 12794-12797

Chase MW, Salamin N, Wilkinson M, Dunwell JM, Kesanakurthi RP, Haidar N, Savolainen V (2005) Land plants and DNA barcodes: short-term and long-term goals. Philos Trans R Soc Lond B Biol Sci. 360: 1889-1895

Chen SL, Yao H, Han JP, Liu C, Song JY, Shi LC, Zhu YJ, Ma XY, Gao T, Pang XH, Luo K, Li Y, Li XW, Jia XC, Lin YL, Leon C (2010) Validation of the ITS2 region as a novel DNA barcode for identifying medicinal plant species. PLoS One. 5: e8613

Chen SL, Yao H, Song JY, Li XW, Liu C, Lu JW (2007) Use of DNA barcoding to identify Chinese medicinal materials. World Sci Technol/Modernization Traditional Chinese Medicine Materia Medica. 9: 7-12

Gadek PA, Alpers DL, Heslewood MM, Quinn CJ (2000) Relationships within Cupressaceaesensulato: a combined morphological and molecular approach. Am J Bot. 87: 1044-1057

Hajibabaei M, Janzen DH, Burns JM, Hallwachs W, Hebert PD (2006) DNA barcodes distinguish species of tropical Lepidoptera. Proc Natl Acad Sci USA. 103: 968- 971

Han JP, Song JY, Liu C, Chen J, Qian J, Zhu YJ, Shi LC, Yao H, Chen SL (2010) Identification of *Cistanche* species (Orobanchaceae) based on sequences of the plastid *psb*A-*trn*H intergenic region. Act Pharma Sin. 1: 126-130

Hebert PD, Cywinska A, Ball SL, deWaard JR (2003a) Biological identifications through DNA barcodes. Proc Biol Sci. 270: 313-321

Hebert PD, Ratnasingham S, deWaard JR (2003b) Barcoding animal life: cytochrome oxidase subunit 1 divergences among closely related species. Proc Biol Sci. 270 (Suppl 1): S96-S99

Hebert PD, Gretory TR (2005) The promise of DNA barcoding for taxonomy. Syst Biol. 54: 852-859

Hilu KW, Borsch T, Müller K, Soltis DE, Soltis PS, Savolainen V, Chase MW, Powell MP, Alice LA, Evans R, Sauquet H, Neinhuis C, Slotta TA, Rohwer JG, Campbell CS, Chatrou LW (2003) Angiosperm phylogeny based on matK sequence information. Am J Bot. 90: 1758-1776

Keller A, Schleicher T, Schultz J, Müller T, Dandekar T, Wolf M (2009) 5.8S-28S rRNA interaction and HMM-based ITS annotation. Gene. 430: 50-57

Kress WJ, 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, Janzen DH (2005) Use of DNA barcodes to identify flowering plants. Proc Natl Acad Sci USA. 102: 8369-8374

Kurtzman CP, Robnett CJ (1998) Identification and phylogeny of ascomycetous yeasts from analysis of nuclear larger subunit (26S) ribosomal DNA partial sequences. Antonie Van Leeuwenhoek. 73: 331-371

Kuzoff RK, Sweere JA, Soltis DE, Soltis PS, Zimmer EA (1998) The phylogenetic potential of entire 26S rDNA sequences in plants. Mol Biol Evol. 15: 251-263

Lahaye R, van der Bank M, Bogarin D, Warner J, Pupulin F, Gigot G, Maurin O, Duthoit S, Barraclough TG, Savolainen V (2008) DNA barcoding the floras of biodiversity hotspots. ProcNatlAcadSci USA. 105: 2923-2928

Ma XY, Xie CX, Liu C, Song JY, Yao H, Luo K, Zhu YJ, Gao T, Pang XH, Qian J, Chen SL (2010) Species identification of medicinal pteridophytes by a DNA barcode marker, the chloroplast psba-trnH intergenic region. Biol Pharm Bull. 33: 1919-1924

Mei Q (2011) Qingtiankui. In: Mei Q (ed) Chinese Herbal medicine produced in Guangdong, 1st edn. Guangdong Technology Press, Guangzhou

Meier R, Shiyang K, Vaidya G, Ng PKL (2006) DNA barcoding and taxonomy in Diptera: a tail of high

intraspecific variability and low identification success. Syst Biol. 55: 715-728

Meyer CP, Paulay G (2005) DNA barcoding: error rates based on comprehensive sampling. PLoS Biol 3: e422

Newmaster SG, Fazekas AJ, Ragupathy S (2006) DNA barcoding in land plants: evolution of *rbcL* in multigene tiered approach. Can J Bot. 84: 335-341

Ross HA, Murugan S, Li WL (2008) Testing the reliability of genetic methods of species identification via simulation. Syst Biol. 57: 216-230.

Shaw J, Lickey EB, Schilling EE, Small RL (2007) Comparison of whole chloroplast genome sequences to choose noncoding region for phylogenetic studies in angiosperms: the tortoise and the hare III. Am J Bot. 94: 275-288

Song J, Yao H, Li Y, Li X, Lin Y, Liu C, Han J, Xie C, Chen S (2009) Authentication of the family Polygonaceae in Chinese pharmacopoeia by DNA barcoding technique. J Ethnopharmacol. 124: 434-439

Tate JA, Simpson BB (2003) Paraphyly of Tarasa (Malvaceae) and diverse origins of the polyploid species. Syst Bot. 28: 723-737.

Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA 4: Molecular Evolutionary Genetic Analysis (MEGA) software verson 4.0. Mol Biol Evol. 24: 1596-1599

Wen HQ, Xu ZR, Vill-Lobos J, Skong LE (1993) A List of threatened limestone plants in south China. Guihaia. 2: 110-127

Yao H, Song JY, Ma XY, Liu C, Li Y, Xu HX, Han JP, Duan LS, Chen SL (2009) Identification of Dendrobium species by a candidate barcode sequence: the chloroplast psbA-trnH intergenic region. Planta Med. 75: 667-669