Asian Journal of Plant Sciences

ISSN 1682-3974

science alert

ANSInet
an open access publisher
http://ansinet.com

OPEN ACCESS Asian Journal of Plant Sciences

ISSN 1682-3974
DOI: 10.3923/ajps.2022.716.726

Research Article
DNA Barcode Identification of New Yellow *Camellia* Species: Endangered Plant Species in North Vietnam

[1]Mai Hai Chau, [2]Hoàng Minh Trang and [2]Ha Van Huan

[1]Vietnam National University of Forestry at Dong Nai, Vietnam
[2]Vietnam National University of Forestry, Xuan Mai, Chuong My, Ha Noi, Vietnam

Abstract

Background and Objectives: DNA barcoding has been proposed as a powerful tool for identifying, confirming species and genetic relationships among species. Using DNA barcoding can overcome problems of morphological-based species identification. In this study, DNA barcode sequences (*mat*K, *rbc*L, *trn*H-*psb*A, *ycf*1b and *ITS*2) were identified for five new yellow *Camellia* species (*Camellia tienii* Ninh, *Camellia hakodae* Ninh, Tr., *Camellia crassiphylla* Ninh et Hakoda, *Camellia tamdaoensis* Hakoda et Ninh and *Camellia petelotii* (Merr.) Sealy) in North Vietnam. **Materials and Methods:** The leaf specimens of five yellow *Camellia* species were used for extraction of total DNA using a plant isolation kit. The five DNA barcode fragments were amplified by PCR technique and sequenced by the Sanger method. **Results:** The result showed that their size was 945-951, 599, 502-510, 750-762 and 397-399 bp, respectively. The genetic distance was calculated and indicated that no locus is suitable to be the specific DNA barcode for the identification of these species. However, the combination of *ycf*1b+*ITS*2 loci was the most effective to identify five yellow *camellia* species in North Vietnam, followed by the combinations of loci *rbc*L+*trn*H-*psb*A+*ITS*2, *trn*H-*psb*A+*ycf*1b+*ITS*2. **Conclusion:** The result revealed the genetic relationship among five researched yellow *Camellia* species and with other species. In which the *Camellia tienii* Ninh, *Camellia hakodae* Ninh and *Camellia petelotii* (Merr.) Sealy was on the same branch whereas *Camellia crassiphylla* Ninh et Hakoda and *Camellia tamdaoensis* Hakoda et Ninh were on the other.

Key words: DNA barcode, *ITS*2, *mat*K, *rbc*L, *trn*H-*psb*A, *ycf*1b, yellow *Camellia*

Citation: Chau, M.H., H.M. Trang and H.V. Huan, 2022. DNA barcode identification of new yellow *Camellia* species: Endangered plant species in North Vietnam. Asian J. Plant Sci., 21: 716-726.

Corresponding Author: Ha Van Huan, Vietnam National University of Forestry, Xuan Mai, Chuong My, Hanoi, Vietnam Tel: +84.2433840233 Fax: +84.2433840063

Competing Interest: The authors have declared that no competing interest exists.

Data Availability: All relevant data are within the paper and its supporting information files.

INTRODUCTION

According to the International *Camellia* Journal (International *Camellia* Society Congress 2013, No. 45), there are 62 kinds of yellow *Camellia* species all over the world, which there are 22 species in Vietnam, 40 species in China (36 in Guangxi, 2 in Guizhou, 1 in Yunnan and 1 in Sichuan). These are distributed from the southern Guangxi of China to Vietnam. They were discovered in Vietnam in the early 20th century, most of the yellow *Camellia* species have great economic value, ornamental properties and medicinal value[1]. The leaves of yellow *Camellia* contain diverse phenolic compounds such as ellagitannins, taxifolin deoxyhexose, proanthocyanidins, kaempferol derivatives, apigenin derivatives, glucosyl isorhamnetin, quercetin derivatives and platphyllosides. The phenolic compounds in yellow *Camellia* have a great value in terms of medicinal utility such as degrading the possibility of stroke, preventing cancer, fortifying the elasticity of blood vessels and regulating blood pressure[2]. However, nowadays, the overexploitation along with scattered distribution in nature, the number of yellow *Camellia* individuals has declined significantly. To conserve, exploit and develop effectively yellow *Camellia* species, taxonomic identification and their genetic relationship of them are very important. The classification and identification by the traditional method got a lot of trouble because yellow *Camellia* species have similar morphological characteristics, so that's necessary to develop another method to overcome the disadvantage of morphological classification. A new classification method was developed based on the molecular markers of the species-the DNA barcode method[3]. It is a short chain of DNA sequences that has a conservative region and evolutionary variation region[4]. Classification is based on the level of change in this sequence to assess the genetic difference in organisms, similar to the barcode scanner in the supermarket[5]. DNA barcoding is an effective tool for studies on the classification and identification of organisms, including animals, plants, fungi, microorganisms and viruses[6-9]. DNA barcodes can be segments in the nuclear genome (18S, 5,8S, ITS)[10], mitochondrial (*Cyt*b, *CO*1), chloroplast (*mat*K, *rbc*L, *trn*H-*psb*A, *ycf*1b). Each barcode has its characteristics with the discrimination potential at a different level: Family, genus, species or subspecies. However, no barcode can be used for all species. In the animal kingdom, the mitochondrial gene cytochrome c oxidase I (CO1) is used widely as a DNA barcode discriminating 98% of species[11]. In-plant kingdom, because of the low mutation rate in the mitochondrial genome, CO1 is not a suitable locus to be a DNA barcode. Instead, scientists have used several loci in chloroplast and nuclear genome but no single locus can be used as a universal DNA barcode for the plant.

In this study, we estimated the discrimination potential of 5 loci in both nuclear and chloroplast genomes, including *mat*K, *rbc*L, *trn*H-*psb*A, *ycf*1b and *ITS*2[12]. The sequences obtained would be deposited on Genbank to conduct a molecular database for yellow *Camellia*. The best choice of combinatory was also purposed as an efficient DNA barcode for them as well as built the phylogenetic tree of yellow *Camellia* species in particular and among *Camellia* genus in general.

MATERIALS AND METHODS

This study project was conducted from January 1, 2015 to December 30, 2020, at the College of Forestry Biotechnology of Vietnam National Forestry University (VNUF).

Plant materials: The leaf specimens of five yellow *Camellia* species were collected in Tam Dao National Park, Vinh Phuc Province, Vietnam (Table 1). Each species collected three samples from different individual plants. The samples were dried with silica gel and stored at -80°C for DNA extraction.

Total DNA extraction: Total DNA was extracted from dried leaf specimens by the plant fungi^{-1} Isolation Kit of Norgen-Canada. The 100 mg of dried leaf specimens were cleaned and ground in liquid nitrogen using a mortar and pestle. They were incubated with lysis buffer L and RNase for 65°C 10 min, then added binding buffer I, mixed thoroughly by inverting and stored on ice for 5 min. Using filter column and spin column to acquire total DNA from the homogeneous mixture. Subsequent steps proceeded following the kit protocol. Total DNA then is stored at -20°C, estimated by 1% agarose electrophoresis and NanoDrop spectrophotometer.

Table 1: The information of specimens and candidate DNA barcode

Order	Scientific name of species	Symbol of specimens	Candidate DNA barcodes
1	*Camellia tienii* Ninh	C1 (C1.1, C1.2, C1.3)	*mat*K, *rbc*L, *trnH-psb*A, *ycf*1b and *ITS*2
2	*Camellia hakodae* Ninh, Tr.	C2 (C2.1, C2.2, C2.3)	
3	*Camellia crassiphylla* Ninh et Hakoda	C3 (C3.1, C3.2, C3.3)	
4	*Camellia tamdaoensis* Hakoda et Ninh	C4 (C4.1, C4.2, C4.3)	
5	*Camellia petelotii* (Merr.) Sealy	C5 (C5.1, C5.2, C5.3)	

Table 2: Primer pairs for PCR amplification of candidate DNA regions

DNA barcode locus	Primer	Primer sequence (5′-3′)	Temperature (°C)
*mat*K	mat-F	5′-TCCATGGGTTTATATGGATCCTTCCTGGTT-3′	60.7
	mat-R	5′-CCCGCCATGGATGGAAGAATTCAAAAGATA-3′	60.5
*rbc*L	rP1-F	5′-ATGTCACCACAAACAGAGACTAAAGC-3′	57.2
	rP1-R	5′- GTAAAATCAAGTCCACCRCG-3′	52.8
*trn*H-*psb*A	trnP-F1	5′-CGCGCATGGTGGATTCACAATCC-3′	61.1
	psbP-R1	5′- GTTATGCATGAACGTAATGCTC-3′	52.3
*ycf*1b	ycf1-F	5′-TCTCGACGAAAATCAGATTGTTGTGAA-3′	56.9
	ycf1-R	5′- ATACATGTCAAAGTGATGGAAAA-3′	50.6
*ITS*2	Is2P-F	5′-ATGCGATACTTGGTGTGAAT-3′	51.9
	Is2P1-R	5′- TCCTCCGCTTATTGATATGC-3′	52.1

PCR amplification and DNA sequencing: From the total DNA extracted, five candidate DNA barcode loci of all five yellow *Camellia* species were amplified and sequenced. PCR primer pairs were used in this study as in Table 2. The PCR reactions were carried out in a 2X PCR Master mix Solution (Intron-Korea). The final reaction volume was 20 µL with the component containing 10 µL 2X PCR Master mix Solution, 10 µM of each primer and approximately 30 ng of template DNA.

The reaction mixtures were performed at 94°C for 3 min and then 35 cycles of PCR as follows: denaturation at 94°C for 35 sec, annealing temperature depended on primers (from 48-57°C) for 30 sec and extension at 72°C for 1 min, final extension at 72°C for 7 min. The reactions were repeated 3 times for each DNA template and primer pair. The PCR products were detected on 1% agarose gel before being sequenced by ABI PRISM® 3100 Avant Genetic Analyzer sequencing machine, using Big Dye Terminator v3.1 Cycle Sequencing Kit.

Sequence alignments and phylogenetic trees: Multiple sequence alignment analysis using the program (MEGA6). Phylogenetic Trees Based on Maximum Likelihood Estimation Reference sequence: The sequences of five loci *mat*K, *rbc*L, *trn*H-*psb*A, *ycf*1b and *ITS*2 from other *Camellia* species were downloaded from NCBI and added to the list of sequences to analyse and build phylogenetic trees. All the selected sequences were over 300 bp in length, from identified species.

RESULTS

DNA extraction: Extracted total DNA from dried leaves of yellow *Camellia* species were separated by 1% (w/v) agarose gel electrophoresis. The results of Fig. 1 showed, DNA fagments as clearly defined bands, with not much debris and RNA, proving that the total DNA obtained was intact.

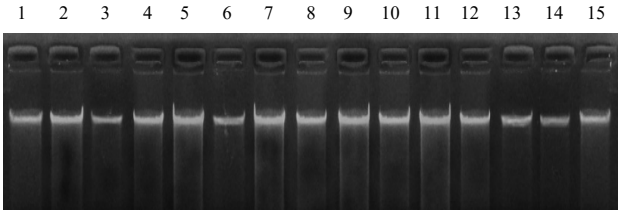


Fig. 1: Total DNA extracted from 15 specimens of five yellow *Camellia* species
Line 1-3: C1.1, C1.2, C1.3, Line 4-6: C2.1, C2.2, C2.3, Line 7-9: C3.1, C3.2, C3.3, Line: 10-12: C4.1, C4.2, C4.3; Line 13-15: C5.1, C5.2, C5.3 specimens

Diluting total DNA 10 folds and measuring by Nano Drop spectrophotometer, the result indicated that the DNA concentration was high, low contaminated and suitable for PCR amplification.

PCR amplification: Total DNA from samples of five yellow *Camellia* species was used to be templates in PCR. The primers sequences were designed following the research before and the Genbank database. All DNA barcodes loci were successfully amplified and were inspected by electrophoresis on 1% agarose (Fig. 2). The DNA bands were clear with high content, no by-product, sequences for each locus have uniform length showing that the primers were specificity. Compared to the DNA ladder, the size of *mat*K, *rbc*L, *trn*H-*psb*A, *ycf*1b and *ITS*2 sequences were about 950, 600, 500, 750 and 400 bp, respectively.

DNA sequencing and alignment: These PCR products were sequenced directly after being purified by PCR purification Kit (Norgen-Canada). The sequencing result showed that five DNA barcodes were 945-951, 599, 502-510, 750-762 and 397-399 bp in length, corresponding to *mat*K, *rbc*L, *trn*H-*psb*A, *ycf*1b and *ITS*2, respectively, suitable with the electrophoresis result. All five sequences were identical (100%) among three individuals with whole five species. The *matK* sequence had the maximum length (945-951 bp) with the number of

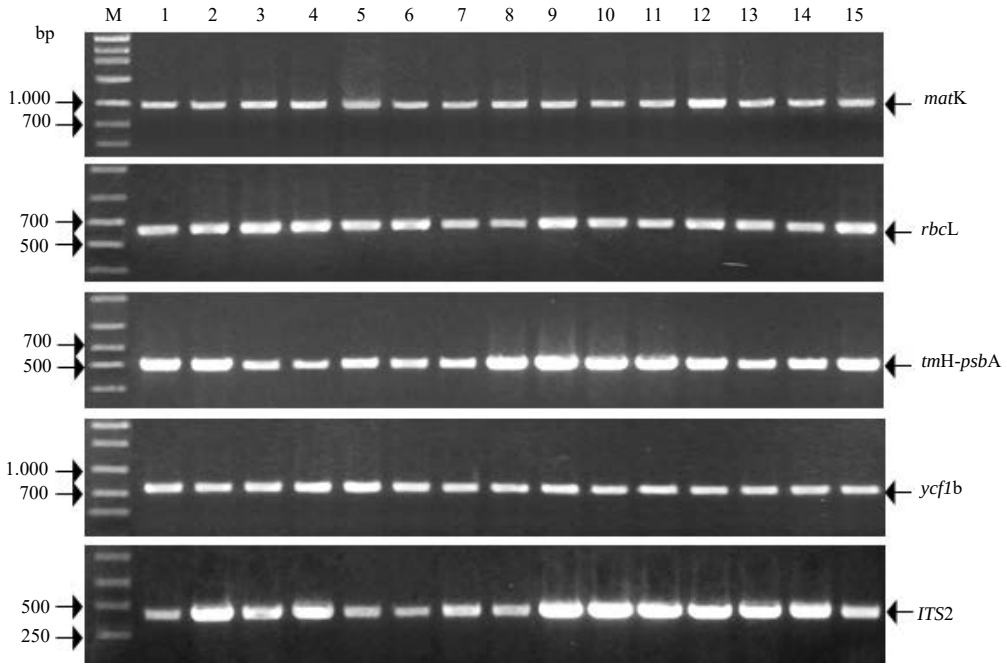


Fig. 2: PCR amplification products of five candidate DNA barcodes from 15 specimens of five yellow *Camellia* species
Line 1-3: C1.1, C1.2, C1.3; Line 4-6: C2.1, C2.2, C2.3; Line 7-9: C3.1, C3.2, C3.3; Line: 10-12: C4.1, C4.2, C4.3; Line 13-15: C5.1, C5.2, C5.3 specimens

Table 3: Evaluation of five DNA barcodes loci

	DNA barcode loci				
	matK	*rbcL*	*trnH-psbA*	*ycf*1b	*ITS*2
Number of individuals	15	15	15	15	15
PCR success (%)	100	100	100	100	100
Sequencing success (%)	100	100	100	100	100
Sequence length	945-951	599	502-510	750-762	397-399
Aligned length	951	599	510	762	408
Number of variable sites	1	1	1	2	4
Number of conservative sites	950	598	508	760	404
Number of indels	6	0	14	12	26
Number of informative sites	0	0	1	0	2

conservative nucleotides reaching 950 bp, 6 indels and only 1 variable site and no informative site. Two other sequences that had no informative site were *rbc*L and *ycf*1 b with 599 bp and 750-762 bp in length, 1 variable site, no indel and 2 variable sites, 12 indels, respectively. *trn*H-*psb*A and *ITS*2 were two sequences that had informative sites with average length (502-510 bp and 397-399 bp), the number of indels was 14 and 26 (Table 3).

The results of sequence alignment and analysis indicated that the rate of the variable site was very low (0.1-1.0%), which means the number of conservative sites was high. It led to the proportion of informative sites being trivial, occupying 0.2 and 0.5% in *trn*H-*psb*A and *ITS*2 sequence, 0% in three others. We speculated that the similar level of these sequences among researched yellow *Camellia* species was very high.

Genetic distances: Because there was no difference among three individuals for each DNA barcode loci in the same species, there was no intraspecific distance. In this research, we calculated interspecific distances to estimate the DNA barcodes and their combination discrimination (Fig. 3 and Table 4). The pairwise distance in five DNA barcode loci ranged from 0.00-0.78%. The mean ranged from 0.04% (*mat*K) to 0.42 (*ITS*2). We observed that the *mat*K loci had the lowest genetic distance (0.0-0.11%) in contrast to *ITS*2 (0.0-0.72%) but there was no locus suitable to be the specific DNA barcode discriminating five yellow *Camellia* species in this research. Instead, we combined two and more of them and measured them as a common locus. The result revealed that the minimum genetic distances had changed, for

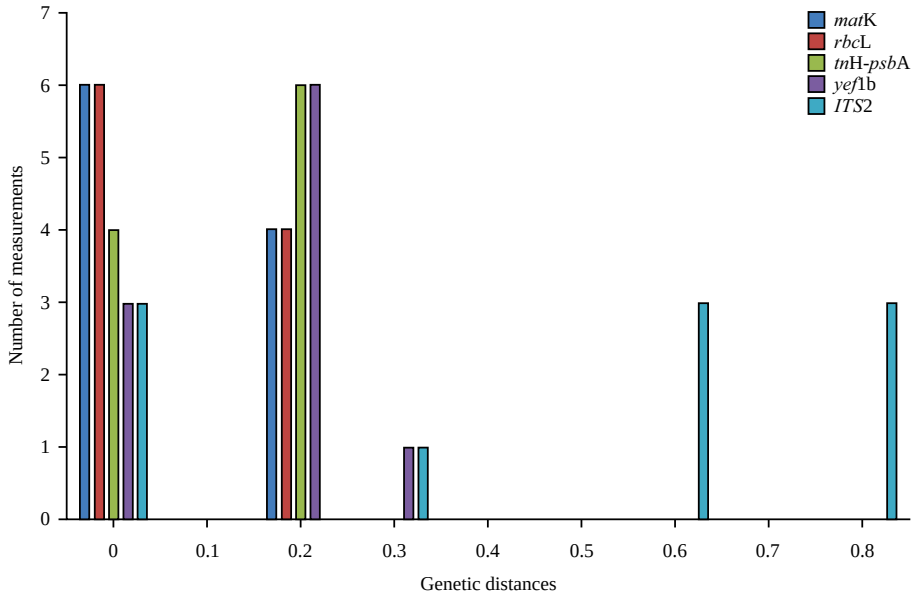


Fig. 3: The genetic distance of five DNA barcodes loci

Table 4: Genetic distance percentage of five DNA barcodes loci and their combinations using kimura 2-parameter

	Interspecific distance (%)		
Barcode loci and combinations	Minimum	Maximum	Mean
*mat*K	0.00	0.11	0.04
*rbc*L	0.00	0.17	0.07
*trn*H-*psb*A	0.00	0.2	0.12
*ycf*1b	0.00	0.27	0.11
*ITS*2	0.00	0.78	0.42
*mat*K+*rbc*L	0.00	0.13	0.05
*mat*K+*trn*H-*psb*A	0.00	0.14	0.07
*mat*K+*ycf*1b	0.00	0.12	0.07
*mat*K+*ITS*2	0.00	0.22	0.15
*rbc*L+*trn*H-*psb*A	0.00	0.18	0.09
*rbc*L+*ycf*1b	0.00	0.22	0.09
*rbc*L+*ITS*2	0.00	0.41	0.2
*trn*H-*psb*A+*ycf*1b	0.00	0.16	0.11
*trn*H-*psb*A+*ITS*2	0.00	0.45	0.25
*ycf*1b+*ITS*2	0.09	0.35	0.21
*mat*K+*rbc*L+*trn*H-*psb*A	0.00	0.15	0.07
*mat*K+*rbc*L+*ycf*1b	0.00	0.13	0.07
*mat*K+*rbc*L+*ITS*2	0.00	0.21	0.12
*mat*K+*trn*H-*psb*A+*ycf*1b	0.00	0.14	0.08
*mat*K+*trn*H-*psb*A+*ITS*2	0.00	0.22	0.14
matK+ycf1b+ITS2	0.01	0.19	0.13
*rbc*L+*trn*H-*psb*A+*ycf*1b	0.00	0.16	0.1
*rbc*L+*trn*H-*psb*A+*ITS*2	0.07	0.34	0.18
*rbc*L+*ycf*1b+*ITS*2	0.06	0.29	0.16
*trn*H-*psb*A+*ycf*1b+*ITS*2	0.06	0.31	0.18
*mat*K+*rbc*L+*trn*H-*psb*A+*ycf*1b	0.00	0.14	0.08
*mat*K+*rbc*L+*trn*H-*psb*A+*ITS*2	0.04	0.21	0.12
*mat*K+*trn*H-*psb*A+*ycf*1b+*ITS*2	0.08	0.19	0.13
*mat*K+*rbc*L+*ycf*1b+*ITS*2	0.04	0.19	0.12
*rbc*L+*trn*H-*psb*A+*ycf1*b+*ITS*2	0.04	0.27	0.15
*mat*K+*rbc*L+*trn*H-*psb*A+*ycf*1b+*ITS*2	0.09	0.19	0.12

instance *ycf*1b+*ITS*2 (0.09%), *rbc*L+*trn*H-*psb*A+*ITS*2 (0.07%), *trn*H-*psb*A+*ycf*1b+*ITS*2 (0.06%), *matK*+*trn*H-*psb*A+*ycf*1b+*ITS*2 (0.08%).

Phylogenetic tree analysis: Phylogenetic trees were constructed based on the difference in barcode loci for all five segments. The sequences were compared with sequences of *Camellia sinensis* (KJ806281.1) (*mat*K), *Camellia oleifera* (JQ975031.1) (*rbc*L), *Camellia oleifera* (JQ975031.1) (*trn*H-*psb*A), *Camellia crapnelliana* (KF753632.1) (*ycf*1b) and *Camellia hongkongensis* (KJ560908.1) (ITS2). They were presented in Fig. 4a-e and Table 5

Results showed that a phylogenetic tree built based on each candidate DNA barcode locus showed a different genetic

Table 5: Homology percentage of five DNA barcodes loci

*mat*K	mC1	mC2	mC3	mC4	mC5	*rbc*L	rC1	rC2	rC3	rC4	rC5
mC1	100					rC1	100				
mC2	100	100				rC2	99.83	100			
mC3	99.89	99.89	100			rC3	99.83	100	100		
mC4	100	100	99.89	100		rC4	99.83	100	100	100	
mC5	99.37	99.37	99.26	99.37	100	rC5	99.83	100	100	100	100
C. sinensis	100	100	99.89	100	99.37	*C. oleifera*	99.83	100	100	100	100
*trn*H-*psb*A	tpC1	tpC2	tpC3	tpC4	tpC5	*ITS*2	IC1	IC2	IC3	IC4	IC5
tpC1	100					IC1	100				
tpC2	98.62	100				IC2	100	100			
tpC3	98	98.82	100			IC3	94.61	94.61	100		
tpC4	98	98.82	100	100		IC4	94.36	94.36	99.5	100	
tpC5	98.23	99.60	98.82	98.82	100	IC5	100	100	94.61	94.36	100
C. oleifera	99	97.64	97	97	97.25	*C. hongkongensis*	90.71	90.71	92.63	92.63	90.71
*ycf*1b	yC1	yC2	yC3	yC4	yC5						
yC1	100										
yC2	99.74	100									
yC3	99.08	99.08	100								
yC4	99.08	99.08	100	100							
yC5	98.29	98.29	99.21	99.21	100						
C. crapnelliana	99.87	99.87	99.21	99.21	98.43						

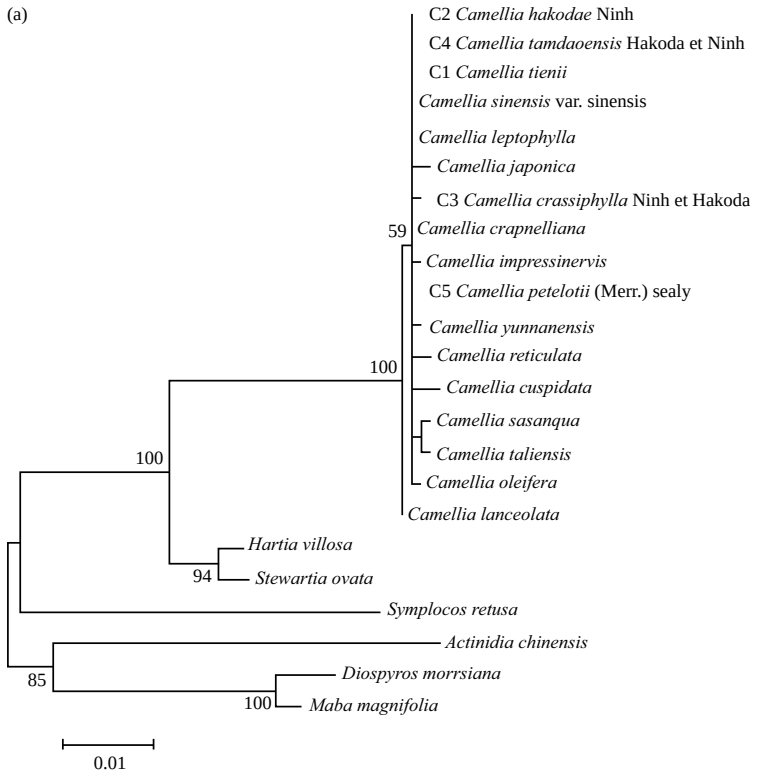
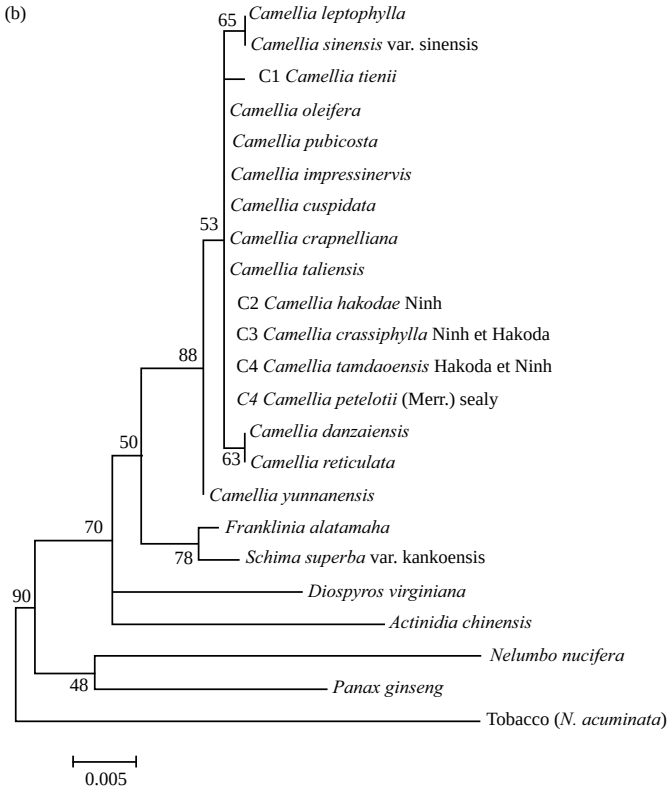


Fig. 4(a-e): Continue



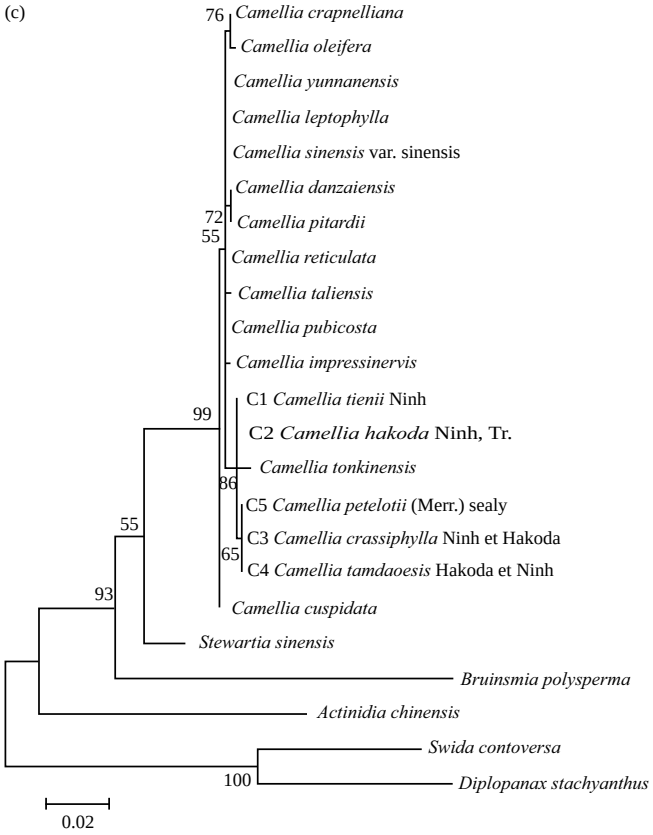


Fig. 4(a-e): Continue

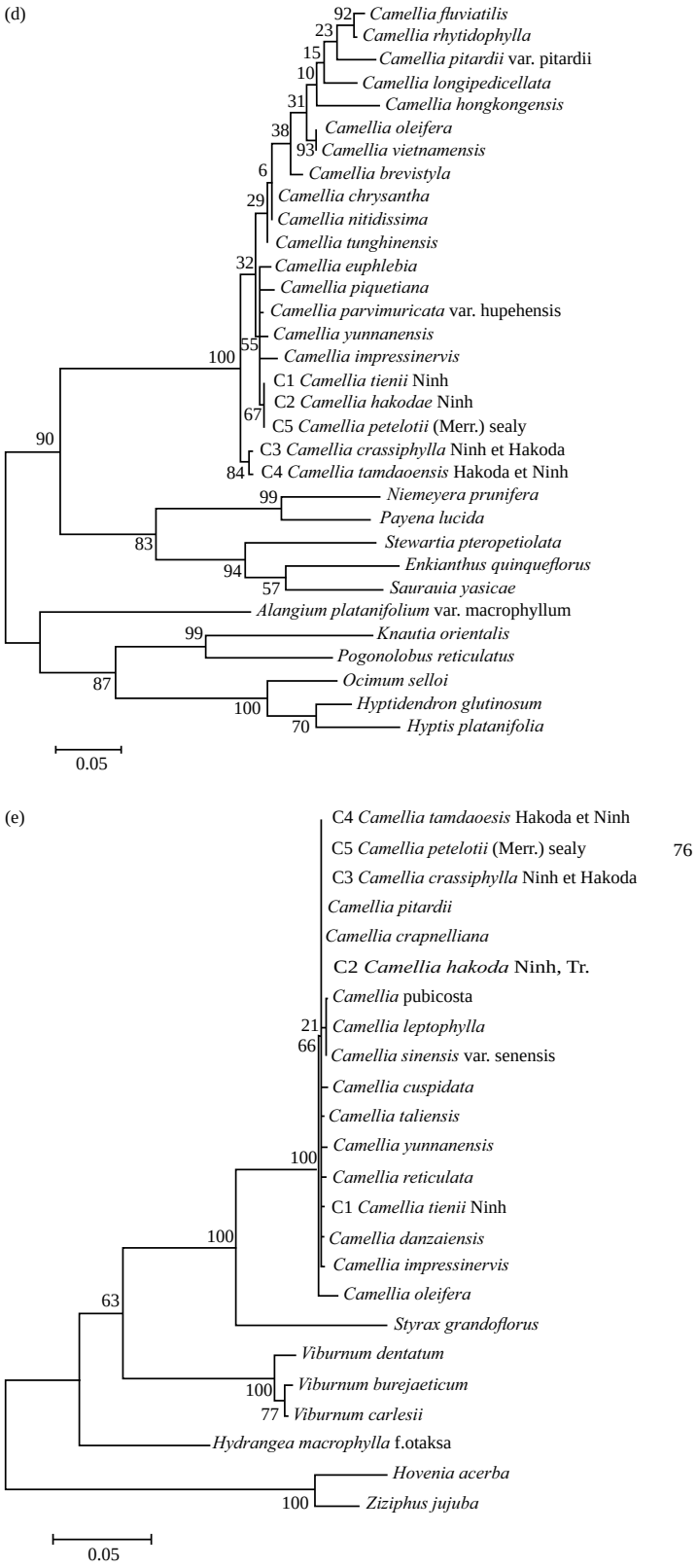


Fig. 4(a-e): Phylogenetic trees are built based on the sequence of five DNA barcodes loci (a) (*mat*K), (b) (*rbc*L), (c) (*trn*H-*psb*A), (d) (*ITS2*) and (e) (*ycf*1 b)

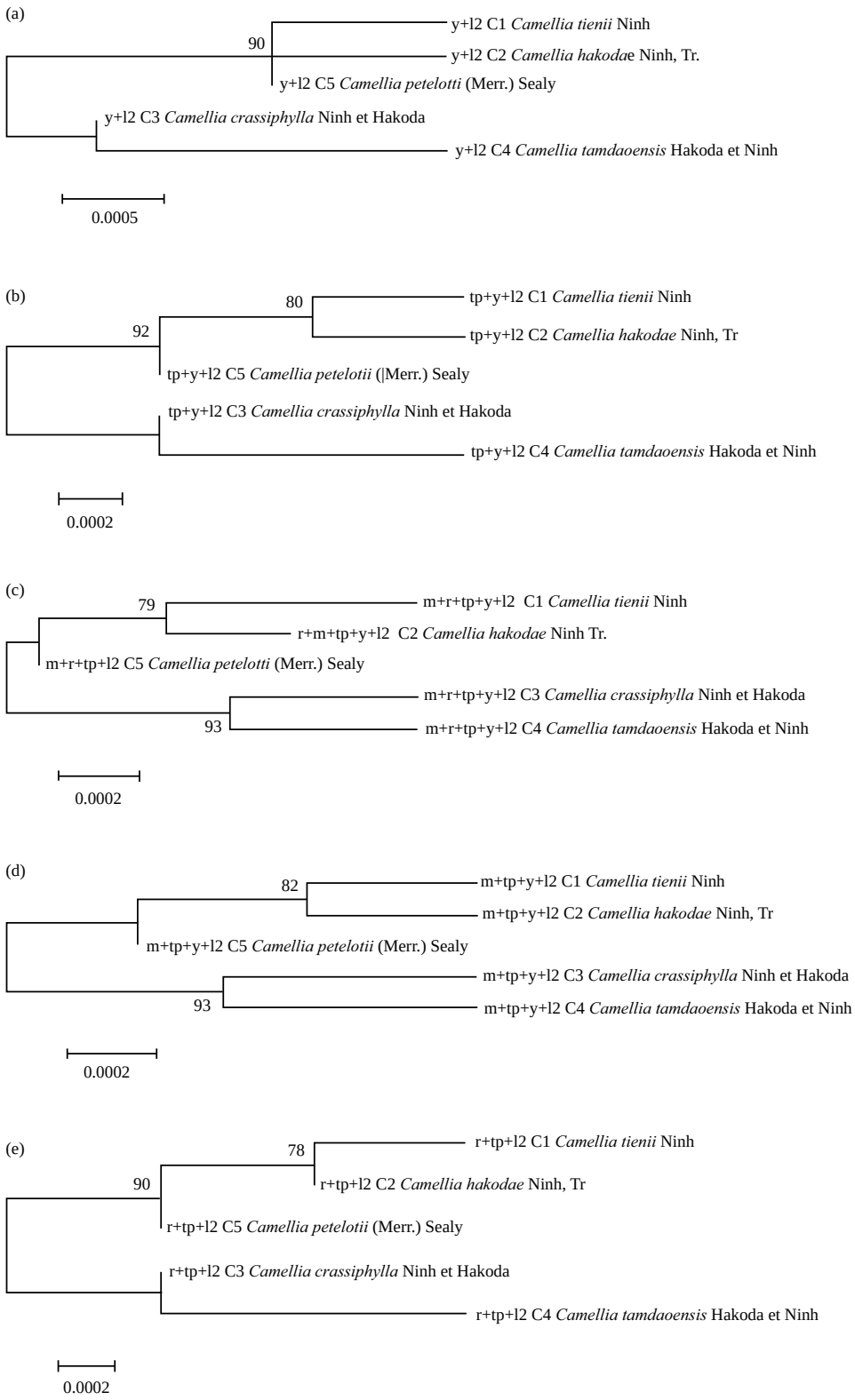


Fig. 5(a-e): Phylogenetic trees are built based on several potential combinations of five DNA barcodes loci (a) (*ycf*1b+*ITS2*), (b) (*trn*H-*psb*A+*ycf*1b+*ITS2*), (c) (*mat*K+*rbc*L+*trn*H-*psb*A+*ycf*1b+*ITS2*), (d) (*mat*K+*trn*H-*psb*A+*ycf*1b+*ITS2*) and (e) (*rbc*L+*trn*H-*psb*A+*ITS2*)

relationship between yellow *Camellia* species in this study in particular and species in the *Camellia* genus in general. This result could be explained by the variance in the similarity level among the barcode loci. However, some certain species in the genus *Camellia* were closely related to the target yellow *Camellia* species that could be seen such as *Camellia cuspidata*, *Camellia leptophylla*, *Camellia crapnelliana*, *Camellia yunnanensis*. In order to analyse the genetic relationship among five yellow species in this study, we constructed the phylogenetic trees based on the most effective discrimination barcode loci combinations, they were *ycf*1b+*ITS*2, *rbc*L+*trn*H-*psb*A+*ITS*2, *trn*H-*psb*A+*ycf*1b+*ITS*2, *mat*K+*trn*H-*psb*A+*ycf*1b+*ITS*2 and *mat*K+*rbc*L+*trn*H-*psb*A+*ycf*1b+*ITS*2. The outcome indicated that these combinations could not only discriminate all five yellow *Camellia* species but also revealed a similar relationship (Fig. 5a-e).

DISCUSSION

The idea of using DNA barcodes for species identification has been used in many types of research with plants, animals, fungi and microorganisms all over the world[13]. However, there was not much research on DNA barcode in yellow *Camellia* apart from several results of the molecular marker in *Camellia* species such as ITS, rDNA (18S, 26S, 45S) but no study focused on *Camellia* genus in general and yellow *Camellia* in particular as well as specific DNA barcodes instead of other molecular markers. In this study, five popular DNA loci were used to identify the specific DNA barcode for five yellow *Camellia* species in North Vietnam[14]. Although all of them were amplified easily from specific pairs of primers and had no problem in sequencing, the analysis showed that there was no individual able to discriminate between all five species. Some other research before had shown similar results[15]. For instance, the results of Stoeckle *et al.*[16]. about *Camellia* species had indicated that the *rbc*L and *mat*K sequences only had one or two variable sites. Even the combination of them proposed by CBOL had been considered to be the core combination barcode in the plant (with the ability to discriminate up to 72%) was also not effective (discriminated three per five species in our study). The best discriminating DNA barcode according to CBOL-*trn*H-*psb*A-had been able to distinguish 69% of 397 species, in this study just achieved 40%. The potential of the *trn*H-*psb*A locus in previous research in plant identification was not only performed as an individual but also combined with other loci, typically in the research of Pang *et al.*[17]. its combination with the *ITS*2 segment had a significant discriminating effect than the combination of *rbc*L+*mat*K. In this study, this combination also showed a considerable difference with the mean genetic distance was 0.25% (the largest among all barcode combinations). However, it was not an appropriate specific barcode for the five yellow *Camellia* species because there was no difference between *Camellia tienii* and *Camellia hakodae*. Rather than using *trn*H-*psb*A+*ITS*2, we combined all pairs of loci to find out the best discriminating combination and obtained an adequate result was *ycf*1b+*ITS*2 (with the genetic distance from 0.09-0.35%, mean 0.21%). The new locus *ycf*1b was recently proposed in a study by Dong *et al.*[18] had quickly shown the potential to discriminate 74% of species in that study[18]. This locus was even more effective when combined with other loci *rbc*L or *mat*K. Indeed, in our study, although it just discriminated against three per five species, when it was combined with the *ITS*2 segment, this combination was able to distinguish all five of them. This was the only two-loci combination satisfied. Besides, the combination of three or four sequences such as *rbc*L+*trn*H-*psb*A+*ITS*2, *mat*K+*trn*H-*psb*A+*ycf*1b+*ITS*2... got effective discrimination of five yellow *Camellia* species in this study.

Based on the five combination of five loci *ycf*1b+*ITS*2, *rbc*L+*trn*H-*psb*A+*ITS*2, *trn*H-*psb*A+*ycf*1b+*ITS*2, *mat*K+*trn*H-*psb*A+*ycf*1b+*ITS*2 and *mat*K+*rbc*L+*trn*H-*psb*A+*ycf*1b+*ITS*2, five phylogenetic trees were constructed followed maximum likelihood estimation. The result showed that these combinations not only could discriminate all five species but also reveal the relationship of the five species similarly. Although the genetic distances were different for each combination, we realized that all five phylogenetic trees separated into two main branches in which *Camellia tienii* Ninh, *Camellia hakodae* Ninh and *Camellia petelotii* (Merr.) Sealy was on one branch whereas *Camellia tamdaoensis* Hakoda et Ninh and *Camellia crassiphylla* Ninh et Hakoda were closer on the other. This is a positive result to identify, classify and systemize yellow *Camellia* species in North Vietnam and worldwide based on the DNA barcode method, especially when these species have similar figure characters that make it difficult to classify by morphological method.

CONCLUSION

Results concluded five candidate barcodes loci-*mat*K, *rbc*L, *trn*H-*psb*A, *ycf*1b and *ITS*2 to select the appropriate sequence to be used as a specific DNA barcode for five yellow *Camellia* species in North Vietnam. In this study, we found the appropriate combination of *ycf*1b+*ITS*2 to identify five yellow *Camellia* species in North Vietnam. The phylogenetic trees

built based on them indicated the relationship of five yellow *Camellia* species in this study so that *Camellia tienii* Ninh was more closely related to *Camellia hakodae* Ninh and *Camellia petelotii* (Merr.) Sealy than *Camellia tamdaoensis* Hakoda et Ninh and *Camellia crassiphylla* Ninh et Hakoda.

SIGNIFICANCE STATEMENT

This is the first study for five DNA barcode fragments (matK, rbcL, trnH-psbA, ycf1b and ITS2) of five new yellow *Camellia* species (*Camellia tienii* Ninh, *Camellia hakodae* Ninh, Tr., *Camellia crassiphylla* Ninh et Hakoda, *Camellia tamdaoensis* Hakoda et Ninh and *Camellia petelotii* (Merr.) Sealy). The combination of ycf1b+ITS2 loci was the most effective to identify five yellow *Camellia* species in North Vietnam.

ACKNOWLEDGMENT

This study was supported financially by the project "Creating the database of DNA barcode for some economic-value large timber trees, non-timber plant forest products" belonging to the agricultural biotechnology program of the Ministry of Agricultural and Rural Development, Vietnam. (Grant number: QD5515). The study was conducted at the Vietnam National University of Forestry (VNUF).

REFERENCES

1. Dũng, L.V., L. Anna, N.T. H nh and N.T. Li u, 2016. *Camellia* thuongiana- a new yellow *Camellia* species from Vietnam. Dalat Univ. J. Sci., 6: 338-344.
2. Lin, J.N., H.Y. Lin, N.S. Yang, Y.H. Li and M.R. Lee *et al*., 2013. Chemical constituents and anticancer activity of yellow *Camellias* against MDA-MB-231 human breast cancer cells. J. Agric. Food Chem., 61: 9638-9644.
3. Hebert, P.D., A. Cywinska, S.L. Ball and J.R. deWaard, 2003. Biological identifications through DNA barcodes. Proc. R. Soc. B: Biol. Sci., 270: 313-321.
4. Kress, W.J. and D.L. Erickson, 2007. A two-locus global DNA barcode for land plants: The coding *rbcL* gene complements the non-coding *trnH-psbA* spacer region. PLoS ONE, Vol. 2. 10.1371/journal.pone.0000508.
5. Neigel, J., A. Domingo and J. Stake, 2007. DNA barcoding as a tool for coral reef conservation. Coral Reefs, 26: 487-499.
6. Kress, W.J., K.J. Wurdack, E.A. Zimmer, L.A. Weigt and D.H. Janzen, 2005. Use of DNA barcodes to identify flowering plants. Proc. Natl. Acad. Sci. USA., 102: 8369-8374.
7. Hebert, P.D., S. Ratnasingham and J.R. de Waard, 2003. Barcoding animal life: Cytochrome c oxidase subunit 1 divergences among closely related species. Proc. R. Soc. London B: Biol. Sci., 270: S96-S99.
8. Lebonah, D.E., A. Dileep, K. Chandrasekhar, S. Sreevani, B. Sreedevi and J.P. Kumari, 2014. DNA barcoding on bacteria: A review. Adv. Biol., Vol. 2014. 10.1155/2014/541787.
9. Seifert, K.A., 2009. Progress towards DNA barcoding of fungi. Mol. Ecol. Resour., 9: 83-89.
10. Chen, S., H. Yao, J. Han, C. Liu and J. Song *et al*., 2010. Validation of the ITS2 region as a novel DNA barcode for identifying medicinal plant species. PLOS ONE, Vol. 5. 10.1371/journal.pone.0008613.
11. Waugh, J., 2007. DNA barcoding in animal species: Progress, potential and pitfalls. BioEssays, 29: 188-197.
12. Schoch, C.L., K.A. Seifert, S. Huhndorf, V. Robert and J.L. Spouge *et al*., 2012. Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for *Fungi*. Proc. Natl. Acad. Sci. U.S.A., 109: 6241-6246.
13. Hollingsworth, M.L., L.A. Clark, L.L. Forrest, J. Richardson and R.T. Pennington *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.
14. Hollingsworth, P.M., 2011. Refining the DNA barcode for land plants. Proc. Natl. Acad. Sci. USA., 108: 19451-19452.
15. Taberlet, P., E. Coissac, F. Pompanon, L. Gielly and C. Miguel *et al*., 2006. Power and limitations of the chloroplast *trn*L (UAA) intron for plant DNA barcoding. Nucleic Acids Res., 35: e14-e14.
16. Stoeckle, M.Y., C.C. Gamble, R. Kirpekar, G. Young, S. Ahmed and D.P. Little, 2011. Commercial teas highlight plant DNA barcode identification successes and obstacles. Sci. Rep., Vol. 1. 10.1038/srep00042.
17. Pang X., C. Liu, L. Shi, R. Liu,D. Liang and H. Li *et al*., 2012. Utility of the *trnH–psbA* Intergenic spacer region and its combinations as plant DNA barcodes: A meta-analysis PLOS ONE, Vol. 7. 10.1371/journal.pone.0048833.
18. Dong, W.P., C. Xu, C.H. Li, J.H. Sun and Y.J. Zuo *et al*., 2015. *ycf*1, the most promising plastid DNA barcode of land plants. Sci. Rep., Vol. 5. 10.1038/srep08348.