AGRICULTURAL RESEARCH COMMUNICATION CENTRE
www.arccjournals.com

Reference ID. ARCC/A-564 **Date : 10-08-2020**

Dang Thanh Long,
Institute of Biotechnology,
Hue University, provincial road No. 10,
Phu Vang, TTHue, Vietnam

Acceptance of article

Dear Dr. Thanh,

We are pleased to inform you that your article has been accepted for publication in **Indian Journal of Agricultural Research.** Your submission is a well-thought out piece of writing and follows many of journal guidelines. The editors agreed that your submission showed great writing skills.

Article Title : Evaluation of Genetic diversity by DNA barcoding of local lotus populations from Thua Thien Hue province

Author(s) : Dang Thanh Long, Huynh Van Chuong, Hoang Thi Kim Hong, Le Ly Thuy Tram, Nguyen Thi Quynh Trang

Congratulations to you once again on your article acceptance in ARCC Journals, and we look forward to receiving more of your good submissions.

With Best Wishes and Seasonal Greetings,

Sincerely,

(R.D. Goel)
Managing Editor

1130, SADAR, KARNAL - 132 001 (HARYANA), INDIA
E-mail :contact@arccjournals.com / editor@arccjournals.com
Website :www.arccjournals.com
Phone: + 91 92555 40308

A-564
[1-8]

RESEARCH ARTICLE Indian Journal of Agricultural Research, Volume Issue : ()

Evaluation of Genetic Diversity by DNA Barcoding of local lotus Populations from Thua Thien Hue Province

Dang Thanh Long[1], Hoang Thi Kim Hong[2], Le Ly Thuy Tram[3], Nguyen Thi Quynh Trang[4]

ABSTRACT

Background: DNA barcoding is a relatively new method of identifying plant species using short sequences of chloroplast DNA. Although there is a large number of studies using barcoding on various plant species, there are no such studies in the genus *Nelumbo*.

Method: Three chloroplast DNA regions (*rbcL, matK, trnH-psbA*) were tested for their suitability as DNA barcoding regions of thirty three lotus samples which were collected in Thua Thien Hue province, Vietnam. Universal primers were used and sequenced products were analyzed using Minimum Evolution method in the MEGA 7.0 program.

Result: We did not observe high variability in nucleotide sequences within the *rbcL* region (0.135%). White Nelumbo, while, the most encoding *matK* (8.013%) and variable *trnH-psbA* (with different number of repeating regions TAAAA) intergenic regions was the most useful for *Nelumbo* barcoding. Individual application of the studied regions did not provide the expected results. None of the regions used in the study allowed the division of white and pink lotus varieties of *N. nucifera* specie according to the adopted classification of the genus *Nelumbo*. The results confirm that the use of *matK, rbcL* and *trnH-psbA* or combine all three regions together is insufficient for DNA barcoding in white and pink lotus varieties of *N. nucifera* specie and better discrimination within the genus *Nelumbo*. Our results also indicate the necessity of using a different region. All of the new sequences have been deposited in GeneBank under the following accession numbers: *rbcL* (MN011708 to MN068956); *matK* (MN011719 to MN068978) and *trnH-psbA* (MN011730 to MN086252).

Key words: Genetic diversity analysis, Lotus, *MatK,* Nelumbo nucifera, *RbcL, TrnH-psbA*.

INTRODUCTION

The appearance of invasive plant species is one of the main causes of native plant extinction. Biodiversity invasions are recognized as one of the most important causes of ecosystem degradation as well as local species community structure and biodiversity loss across the world (Thompson, 1997). *Nelumbo nucifera Gaertn* (Nelumbonaceae) is known by numerous common names like sacred lotus, Indian lotus, bean of India and simply lotus. *Nelumbo nucifera* is an aquatic herb with white or red coloured flowers (Maqbool *et al.* 2019). This perennial usually lives in lakes and ponds. Since ancient times, lotus flower has become familiar and close to Vietnamese's life as well as other countries such as India, China and Japan (Long *et al.* 2020). In addition, lotus is a valuable biological resource of Thua Thien Hue province, this is one of the indispensable parts of the natural heritage and provides not only local specialties but also services which are related to the ecosystem. Therefore, it is necessary to conserve and manage this species.

Recently, there has been increasing interest in biodiversity research through various molecular methods to classify species. Rapid identification and determination of a species is the basis for biodiversity conservation and is one of the keys to improve species management and conservation (Seberg and Petersen, 2009).

DNA barcoding involves the use of a short DNA sequence or sequences from a standardized locus (or loci) as species identification tools (Alam *et al.* 2020). A DNA sequence from such a standardized gene region can be obtained from a small amount of tissue taken from an unidentified organism and then compared to a library of reference sequences from known species. If the sequence from the unknown organism match to one of reference sequences means that the organism is recognize, thus providing a rapid identification. An ideal DNA barcode should

[1]Institute of Biotechnology, Hue University, provincial road No. 10, Phu Vang, TTHue, Vietnam.
[2]Hue University of Sciences, Hue University, 77 Nguyen Hue, Hue, Vietnam.
[3]University of Science and Technology, Da Nang University, 54 Nguyen Luong Bang, Da Nang, VietNam.
[4]Hue University of Education, Hue University, 34 Le Loi, Hue, Vietnam.

Corresponding Author: Dang Thanh Long, Institute of Biotechnology, Hue University, provincial road No. 10, Phu Vang, TTHue, Vietnam. Email: dtlong@hueuni.edu.vn

How to cite this article: Long, D.T., Hong, H.T.K., Tram, L.L.T. and Trang, N.T.Q., (). Evaluation of Genetic Diversity by DNA Barcoding of local lotus Populations from Thua Thien Hue Province. Indian Journal of Agricultural Research. ():

Submitted: 16-04-2020 **Accepted:** 10-08-2020 **Published:**

be present in all groups of land plants. it should be short (700-800 bp) and show enough sequence variation to discriminate among species, also it should be easy to amplify and sequenced with a single primer pair (Kress *et al*. 2009). Different regions from the plastid genome, including *trnH-psbA* intergenic spacer, *rbcL, matK*, *rpoC1* and *rpoB*, have been proposed and tested for DNA barcoding of land plants with different level of species identification success depending of the studied group taxa (Kress *et al*. 2009; Singh *et al*. 2012). The purpose of this study was to test the utility of DNA barcoding for the identification of closely related lotus varieties. In a conservation project, lotus leaves were collected from local farmers of the different locations (Thua Thien Hue province). The Lotus leaf sample were chosen from traditional lotus (varieties that has been passed through several generations of a family). In this study, we used the non-coding plastid *trnH-psbA* intergenic spacer region and two plastid coding regions *rbcL, matK*.

MATERIALS AND METHODS

In this study, there are 33 lotus leaf samples (include varieties of white and pink lotus) which were selected and collected on 33 different locations in Thua Thien Hue province (Fig 1). The lotus leaf samples are washed with distilled water and then refrigerated in the dark for further experiments. This study was conducted at the Institute of Biotechnology, Hue University, June, 2018.

DNA extraction PCR amplification and sequencing

Nelumbo nucifera lotus leaves (*N. nucifera*) are stored at 4°C in the dark for about 1 to 2 days to remove part of the starch existing in leaf tissue. 100 mg samples of young leaf tissues were ground to a fine powder in liquid nitrogen. The powder was then moved in 1,5 mL microtubes containing 700 mL 2% CTAB extraction buffer (20 mM EDTA, 0.1 M Tris-HCl pH 8.0, 1.4 M NaCl, 2% CTAB, plus 0.4% β-mercaptoethanol added just before use). The solution was incubated at 65°C for 60 min, gently mixing by inversion every 15 min; 500 mL of chloroform-isoamylalcohol (24:1) was added to the tubes and gently mixed for 1 min, samples were centrifuged for 10 min. at 12,000 rpm; 500 mL of the supernatant was then transferred to a fresh tube following the addition of 500 mL chloroform-isoamylalcohol (24:1); this procedure was repeated twice. The supernatant (400 mL) was then transferred to a fresh tube with 400 mL of cold isopropanol (-20°C); samples were gently mixed by inversion and centri fuged at 12,000 rpm for 10 min and so it was possible to visualize the DNA adhered to the bottom of the tube; the liquid solution was then released and the DNA pellet washed with 400 mL of 70% ethanol to eliminate salt residues adhered to the DNA and set to dry at room temperature; the pellet was then ressuspended in 100 mL TE buffer (10 mM Tris-HCl pH 8.0 and 1 mM EDTA pH 8.0) plus 5 µL ribonuclease (RNAse 10 mg/mL) in each tube; this solution was incubated at 37°C for 1h and after stored at -20°C. The process of DNA extraction is performed by basic CTAB method as described by Sharma *et al.* (2008) (Sharma *et al.*, 2008).

The three plant DNA barcodes, *rbcL, matK* and *trnH-psbA* were amplified in a 25 µL reaction volume, using My Taq™ DNA Polymerase (Bioline Reagents Ltd. UK), 0.5 µL primers (10 pmol/µL) and 100 ng DNA template (50 ng/µL). PCR amplification was performed on a Applied Biosystems – Life Technologies (Thermo Fisher Scientific Inc.. United States). Primers for PCR and sequencing (Accesion number: KF009944.1) and PCR cycling conditions used in this study are provided in Table 1. Purified PCR products were send to Maccrogen Company, Korea and sequenced in both

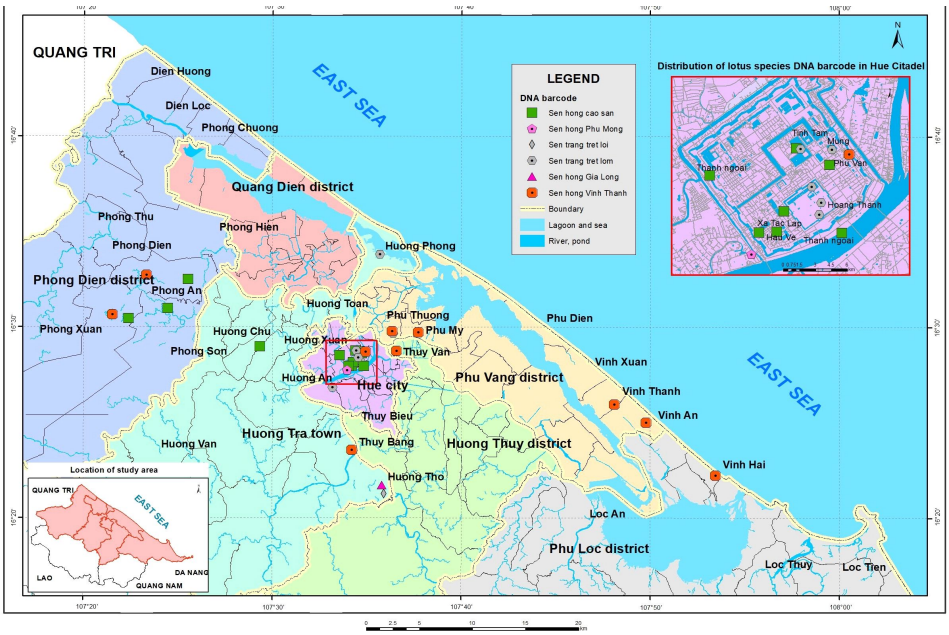

Fig 1: The sampling location collected lotus used in the study.

directions with the same primers used for PCR.

Data analysis

Sequences for each region were assembled and edited using BioEdit v7.2.5. Then, the edited sequences were aligned by ClustalW in MEGA 7.0 and the non-overlapping sequence regions at the 5′ - and 3′ -ends were trimmed (Kumar, 2016). The genetic pair wise distance for three marker was calculated using MEGA 7.0 with the Kimura 2-parameter (K2-P) model.

The evolutionary history was inferred using the Minimum Evolution method (Tamura *et al.,* 2013). Tree #1 out of 100 minimum evolution trees (sum of branch length = 0,03782652) is shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches. The tree is drawn to scale. with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the p-distance method and are in the units of the number of base differences per site. The ME tree was searched using the Close-Neighbor-Interchange (CNI) algorithm. The Neighbor-joining algorithm (Saitou and Nei, 1987) was used to generate the initial tree. All positions containing gaps and missing data were eliminated. There were a total of 2029 positions in the final dataset. Evolutionary analyses were conducted in MEGA 7.0 software (Kumar *et al.* 2016). The barcode sequences were queried against GeneBank database (NCBI) using Nucleotide BLAST algorithm.

DNA polymorphism analysis and recombinant based on eight parameters including number of separate polymorphic sites (S), total number of mutant sites (Eta), number of haplotypes (h), haplotype diversity (Hd), average number of nucleotide differences (k), nucleotide diversity (π) minimum number of recombinant processes (Rm) and number of effective populations for mutation rate at each nucleotide position per generation (Ø) are considered as a polymorphic measurement in the population. Neutrality is tested based on two methods, Tajima's D test (Tajima, 1989) and Fu and Li's D* (Fu, 1993) using DNASP 6.0 software (Rozas *et al.* 2017).

RESULTS AND DISCUSSIONS

Sequence characteristics of the barcodes

The three barcodes, *rbcL, matK* and *trnH-psbA* showed high success rates for PCR amplification and sequencing using a single primer pair. The sequences characteristics of the three regions are presented in Table 2. Of the three barcodes, the *matK* sequences had seventy-five variable sites among the thirty three samples, found in the Local Lotus varieties in Thua Thien Hue, occupies 8,013% of the total gene length, the genetic distances for the *matK* sequence ranged from 0 to 0.097 (mean = 0.027). While, *rbcL* sequences had one variable sites (site 459), occupies 0.135% of the total gene length, genetic distance ranges from 0 to 0.001 (mean = 0) and *trnH-psbA* sequence did not show any variable sites, thus these sequence were 100% conserved within the species (Table 2 and Fig. 2).

The PCR products of *rbcL, matK* and *trnH-psbA* genetic regions were sequenced on ABI PRISM® 3100 Avant Genetic Analyzer (Applied Biosystems) by dideoxy terminator method. As a result, the size of *rbcL* and *matK* genetic region was 743bp and 936bp, respectively, while this figure for *trnH-psbA* fluctuated from 351 to 410bp with different number of repeating regions (TAAAA) (Table 2). The BLAST result on NCBI was used to verify and compare with the sequences of the *N. nucifera* lotus (Accecsion number: KF009944.1), which indicated that the obtained nucleotide sequences were highly similar to *N. nucifera* lotus species. The appearing percentage of each type of nucleotide in the *rbcL* genetic region showed that Adenin (A) accounted for the highest proportion and there was no difference between the studied lotus samples (28.80%), followed by Timin (Uracin), accounting for 27.59 to 27.73% and the lowest proportion is Cystein (C) accounting for 20.86%. The *matK* genetic region containing Timin (Uracin) accounted for the highest

Table 1: Primers their sequences and PCR conditions.

Regions	Primer pairs	Sequence 5¹'→3¹	PCR conditions	References Genbank
rbcL	1F	ATGTCACCACAAACAGAGAC	95 °C/ 3 min; [35 cycles: 95 °C/ 30 s;	KF009944.1
	743R	TCACATGTACCTGCAGTAGC	50 °C/ 30 s; 72 °C/ 90 s]; 72 °C/ 7 min	
matK	385F	CGATCAATTCATTCAATATTTTC		
	1320R	ACTTCGACTTTCGTGTGCTAGA		
trnH-psbA	46F	ACTGCCTTGATCCACTTGGC		
	25R	TGAAGCTCCATCTACAAATGG		

Table 2: The characteristics of each single barcode.

Marker	PCR success (%)	Sequencing success (%)	Total aligned length (bp)	Number of monomorphic sites	Variable sites (%)	Intraspecific distance (mean)
rbcL	100	100	743	742	0.135	0-0.001 (0)
matK	100	100	936	861	8.013	0-0.097 (0.027)
trnH-psbA	100	100	351-410	375	0	0

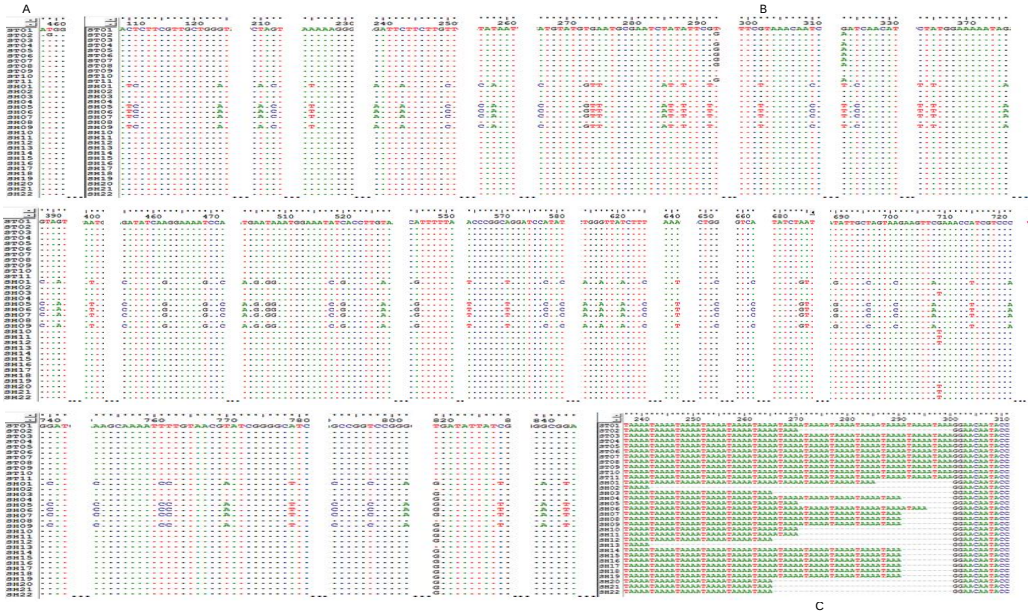

Fig 2: Variable Position of nucleotide base on the three nucleotide sequences of chloroplast genome *Nelumbo*. A. sequence *rbcL* gene; B sequence *matK* gene and C. sequence *trnH-psbA* gene.

proportion of 34.72% and fluctuated from 34.29 to 34.72% among the studied lotus samples, reaching an average of 34.61%. Guanidin (G) accounted for the lowest proportion and there was a difference between lotus samples, fluctuated from 15.81 to 16.67%. At the same time, the types of nucleotides contained in the *trnH-psbA* genetic region fluctuated from 8.54% (G) to 44.74% (A). The highest percentage (G + C) contained in the genome was 43.61% (*rbcL*), 36.86% (*matK*) and 27.07% (*trnH-psbA*), among different lotus samples, the rate also varied and fluctuated from 43.47% to 43.61% (*rbcL*), 35.79% to 36.86% (*matK*) and 23.17% to 27.07% (*trnH-psbA*) and averaged 43.48% (*rbcL*), 35.98% (*matK*) and 24.12% (*trnH-psbA*), respectively (Fig 2).

The results of the sequencing analysis were calibrated, Align and analysed of three barcodes, *rbcL, matK* and *trnH-psbA* by using MEGA 7.0 software, obtained a conserved region of 742/743 nucleotide positions and a modified region of 1/743 nucleotide positions (*rbcL*); conserved region of 861/936 nucleotide positions and a modified region of 75/936 nucleotide positions (*matK*) and *trnH-psbA* sequence did not show any variable sites (Fig 1).

Seventy one/seventy-five separate polymorphic nucleotide sites were found in the *matK* gene region was presented in SH01, SH05, SH06, SH07 and SH09 models. This defective gene sequence could be considered as a sign to build identification and distinguish between the lotus samples SH01, SH05, SH06, SH07 and SH09 compared to the remaining lotus samples. In addition, seventy-five mutation positions (*matk*); One mutation positions (*rbcL*) and repeat mucleotide sequences of *trnH-psbA* when performing analysis with DNASP 6.0 software showed that there were seventy-five separate polymorphic positions (S) created seventy-six mutant positions (Eta) for *matk* sequences and one separate polymorphic positions (S) created one mutant positions (Eta) for *rbcL* and *trnH-psbA* sequences shown in 33 studied lotus samples classified into number of haplotypes (h), respectively, 5 types of halotypes (*matK*), one type of halotype (*rbcL*) and two types of halotype (*trnH-psbA*), with haplotype diversity coefficient (Hd) accounting for 0.822 (*matK*), 0.061 (*rbcL*) and 0.117 (*trnH-psbA*), the average number of nucleotide differences (k) is 20.563 (*matK*), 0.061 (*rbcL*) and 0.117 (*trnH-psbA*), the nucleotide diversity coefficient (ð) accounts for 21.970 x 10^{-3} (*matK*), 0.080 x 10^{-3} (*rbcL*) and 0.330 x 10^{-3} (*trnH-psbA*), the number of effective populations for the rate of mutations per nucleotide position per generation (*Ø*) accounts for 20.010 x 10^{-3} (*matK*), 0.330 x 10^{-3} (*rbcL*) and 0.700 x 10^{-3} (*trnH-psbA*), the minimum number of recombinants (Rm) to occur does not exist. All indicators were processed with statistical significance $p < 0.05$ (Table 3).

Two methods namely (Tajima's D test, Fu and Li's D*) were used to test neutrality. The results in Table 5 with D value of *rbcL* and *trnH-psbA* sequences both yield negative values with not significant $p > 0.10$, this showed that the evolution of the studied lotus population size may be increasing or we may have evidence for purifying selection at this locus. While, D value of *matK* and the combination of *rbcL* + *matK* + *trnH-psbA* sequences both yield positive values with not significant $p > 0.10$, this showed that the evolution of the studied lotus population may have suffered a recent bottleneck (or be decreaing) or we may have evidence for overdominant selection at this locus. In addition, the value of Fu and Li's D* of *trnH-psbA (Not significant: p > 0.10), matK* and and the combination of *rbcL* + *matK* + *trnH-psbA* sequences (*Statistical significance: p < 0.02*) indicated that the studied population had very few individuals showing large differences in comparision with other individuals in the population (Table 4).

Table 3: DNA Polymorphism of three markers of the barcodes region for *N. nucifera (rbcL. matK and trnH-psbA)* populations.

DNA polymorp hism	S	Eta	Haplotype (H)					k	Hd	π(x10^{-3})	Ø(x10^{-3})	Rm
			H_1	H_2	H_3	H_4	H_5					
rbcL	1	1	ST01 ST03 ST04 ST05 ST06 ST07 ST08 ST09 ST10 ST11 SH01 SH02 SH03 SH04 SH05 SH06 SH07 SH08 SH09 SH10 SH11 SH12 SH13 SH14 SH15 SH16 SH17 SH18 SH19 SH20 SH21 SH22	ST02	-	-	-	0.061	0.061	0.080	0.330	0
*matK*75	76		ST01 ST03 ST09 ST10 SH02 SH08 SH13	ST02 ST04 ST05 ST06 ST07 ST08 ST11	SH01 SH05 SH06 SH07 SH09	SH03 SH10 SH11 SH12 SH20 SH21 SH22	SH04 SH14 SH15 SH16 SH17 SH18 SH19	20.563	0.822	21.970	20.010	0
trnH-psbA	1	1	SH04 SH14 SH15 SH16 SH17 SH18 SH19	SH04 SH14 SH15 SH16 SH 17 SH 18 SH19	-	-	-	0.117	0.117	0.330	0.700	0

Phylogenetic analysis

The evolutionary history was inferred using the Minimum Evolution method. The result is shown in Fig. 3. The first cluster grouped all the lotus pink populations; while the second cluster grouped the two pink lotus and white. The tree topology is supported by a good bootstrap value. The differences between the two pink and white lotus populations were found in the three regions of barcodes *rbcL, matK* and *trnH-psbA*. Although, the two lotus populations have a different flower color, shared the same haplotype for the three markers of the barcodes region *rbcL, matK* and *trnH-psbA*, which are considered the most variable coding and non-coding regions of the plastid genome (Chase *et al.* 2007).

According to the CBOL plant working group, an ideal DNA barcode needs to have the following features: capacity of amplification with universal primers, high amplification and sequencing efficiency and genetic variation that is sufficiently high to distinguish sequences at the species level, but also sufficiently conservative among individuals of the same species (Hebert *et al.* 2003, CBOL Plant Working Group, 2009).

Evaluation of universal applicability by PCR quantification and sequencing success is the first step in determining the suitability of a given DNA fragment as a barcode. In this respect, all analyzed regions *(matK, rbcL* and *trnH-psbA*) amplified effectively, which allowed for simple and high-quality sequencing.

The amplicons obtained in our experiments were shorter (about 900 bp), which allowed for effective sequencing. Similar results were obtained for other groups of terrestrial plants, where the amplification of the *trnH-psbA* region and the sequencing quality was sufficiently high to consider it a barcode (Tripathi *et al.* 2013; Bieniek *et al.* 2015).

In this study, we isolated and analysed the sequence of non-coding plastid *trnH-psbA* intergenic spacer region and two plastid coding regions *rbcL, matK* of thirty three lotus samples which were collected in Thua Thien Hue province. As a result, the size of *rbcL* and *matK* genetic region was 743bp and 936bp, respectively, while this figure for *trnH-psbA* fluctuated from 351 to 410bp with different number of repeating regions (TAAAA), which had high similarity with species *N. nucifera* (accession number: KF009944.1). Seventy one/seventy-five separate polymorphic nucleotide sites were found in the *matK* gene region was presented in SH01, SH05, SH06, SH07 and SH09 models. This defective gene sequence could be considered as a sign to build identification and distinguish between the lotus samples SH01, SH05, SH06, SH07 and SH09 compared to the remaining lotus samples. In addition, seventy-five mutation positions (*matk*); One mutation positions (*rbcL*) and repeat mucleotide sequences of *trnH-psbA* when performing analysis with DNASP 6.0 software showed that there were seventy-five separate polymorphic positions (S) created seventy-six mutant positions (Eta) for *matk* sequences and one separate polymorphic positions (S) created one mutant positions (Eta) for *rbcL* and *trnH-psbA* sequences shown in 33 studied lotus samples classified into number of haplotypes (h), respectively, 5 types of halotypes (*matK*), one type of halotype (*rbcL*) and two types of halotype (*trnH-psbA*), with haplotype diversity coefficient (Hd) accounting for 0.822 (*matK*), 0.061 (*rbcL*) and 0.117 (*trnH-psbA*), the average number of nucleotide differences (k) is 20.563 (*matK*), 0.061 (*rbcL*) and 0.117 (*trnH-psbA*), the nucleotide diversity coefficient (ð) accounts for 21.970 x 10^{-3} (*matK*), 0.080 x 10^{-3} (*rbcL*) and 0.330 x 10^{-3} (*trnH-psbA*), the number of effective populations for the rate of mutations per nucleotide position per generation (Ø) accounts for 20.010 x 10^{-3} (*matK*), 0.330 x 10^{-3} (*rbcL*) and 0.700 x 10^{-3} (*trnH-psbA*), the minimum number of recombinants (Rm=0) to occur does not exist. All indicators were processed with statistical significance $p < 0.05$.

Two methods (Tajima's D test, Fu and Li's D*) were used to execute neutrality test. The results showed that, the evolution of lotus population was may be affected by the recent bottleneck (or being solved) or we may have evidence

Table 4: Neutral test results based on three *rbcL, matK* and *trnH-psbA* genetic regions of lotus population

DNA polymorphism	Tajima's D test	Fu and Li's D* test
rbcL	-1.140*	-1.713*
matK	0.366*	1.916**
trnH-psbA	-0.792*	0.584*
rbcL + matK + trnH-psbA	0.325*	1.820**

**Not significant: p > 0.10*

***Statistical significance: p < 0.02.*

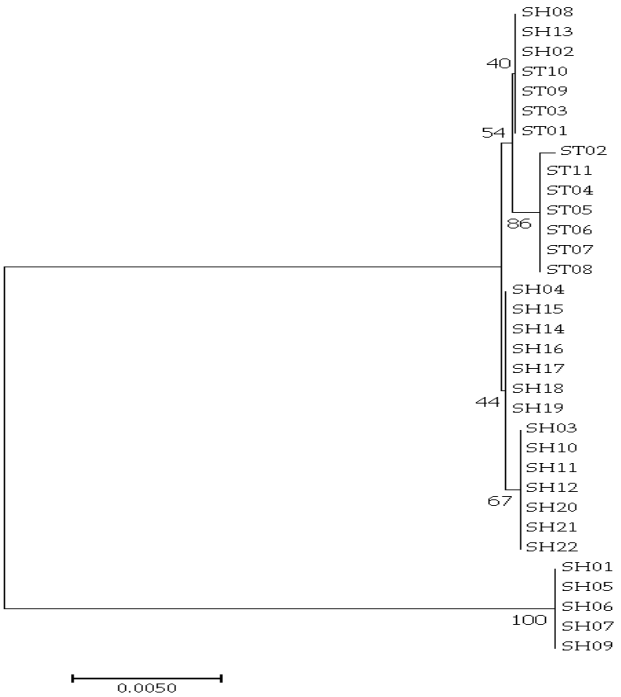

Fig 3: Evolutionary relationships of taxa based on the three markers of the barcodes region *rbcL, matK* and *trnH-psbA*.

of excessive selection at this location. In addition, the value of Fu and Li's D* of the combination of *rbcL* + *matK* + *trnH-psbA* series (*Statistical significance*: *p <0.02*) indicates that the research population has very few individuals. There is a big difference compared to other individuals in the population.

Phylogenetic tree was built based on Minimum Evolution (bootstrap = 1000) showed that, thirty three collected lotus samples closely linked and they were divided into two groups. Group I included 5 samples of pink lotus varieties and group II included 11 samples of white lotus and 17 samples of pink lotus varieties.

In turn, many studies have indicated that *matK* is a key marker discriminating specific groups (Newmaster *et al.* 2009; DeMattia *et al.* 2011), although many authors questioned the usefulness of this gene as a barcode due to poor amplification and sequencing efficiency and problems related to primers' universality (Yan *et al.* 2011; Theodoridis *et al.* 2012). The research presented in the study indicates that despite PCR and sequencing efficiency, unfortunately, this region can not be considered as an effective white and pink lotus varieties of *N. nucifera* specie barcode. Analyses involving this sequence showed only 8.013% polymorphism in the studied taxa.

However, in terms of molecular variability, *rbcL* was the most conservative sequence among the three analyzed regions, as indicated by the lowest number of polymorphic sites and the obtained haplotypes (Fig. 2). This was also confirmed by other authors (Bieniek *et al.* 2015; Zimmermann *et al.* 2013; Bolson *et al.* 2015; Gamache and Sun, 2015). This region also demonstrated reasonably good effectiveness at lower taxonomic levels in *Hordeum* (Bieniek *et al.* 2015; Gamache and Sun, 2015). Bieniek *et al.* (2015) identified *Hordeum bulbosum* or *H. bogdani* using the *rbcL* region.

Our research shows that the *matK* gene sequences are also highly similar in the analyzed taxa (75 polymorphic sites have been identified) and allow only the identification of *Nelumbo* ancestrale. Bieniek *et al.* (2015) obtained different results, demonstrating high species identification capacity, but also for the genus, using the *matK* gene alone in the genera Elymus, Loptiopyrum, Pseudoroegneria and Thinopyrum. These results are in contradiction with the study of Zimmerman *et al.* (2013) in relation to the genus Panicum. This might result from a larger number of species selected for analysis – 9 (Zimmermann *et al.* 2013).

CONCLUSION

The present study is the first to analyze selected white and pink lotus varieties of *N. nucifera* specie, in which the usefulness of the combinations of the plastid *rbcL* and *matK* coding regions and intergenic *trnH-psbA* region for DNA barcoding was assessed. The results confirm that the use of non-coding *trnH-psbA* and *rbcL* is insufficient for DNA barcoding in white and pink lotus varieties of *N. nucifera* specie. Our results also indicate the need to use a different region, e.g., the ITS region and different regions from the plastid genome, *trnH-psbA* intergenic spacer, *rbcL, matK*, *rpoC1* and *rpoB*, have been proposed and tested for DNA barcoding of land plants with different level of species identification success depending of the studied group taxa (Kress and Erickson, 2007; Singh *et al.* 2012), in order to correctly identify white and pink lotus varieties of *N. nucifera* specie.

ACKNOWLEDGEMENT

This study supported by the Domestic Master/ PhD Scholarship Programme of Vingroup Innovation Foundation and by the People's Committee -Department of Science and Technology Province Thua Thien Hue, Vietnam for the project (code No: TTH.2017-KC.02).

Competing interests

Authors have declared that no competing interests exist.

REFERENCES

Alam, A., Chadha, N.K., Kumar A.P., Chakraborty, S.K., Joshi, K.D., Sawant, P.B., Das, S.C.S., Kumar, J., Kumar, T. (2020). DNA Barcoding and Biometric Investigation on the Invasive Oreochromis niloticus (Linnaeus, 1758) from the River Yamuna of Uttar Pradesh. Indian Journal of Animal Research. 54: 856-863.

Bieniek, W., Mizianty, M., Szklarczyk, M. (2015). Sequence variation at the three chloroplast loci (*matK, rbcL, trnH-psbA*) in the Triticeae tribe (Poaceae): comments on the relationships and utility in DNA barcoding of selected species. Plant Systematics and Evolution. 301: 1275-1286.

Bolson, M., de Camargo, S.E., Brotto, M.A., Silva-Pereira V. (2015). *ITS* and *trnH-psbA* as Efficient DNA Barcodes to Identify Threatened Commercial Woody Angiosperms from Southern Brazilian Atlantic Rainforests. PLoS One. 10(12): e0143049.

CBOL Plant Working Group, (2009). A DNA barcode for land plants. Proceedings of the National Academy of Sciences of the United States of America. 106: 12794-12797.

Chase, M.W., Cowan, R.S., Hollingsworth, P.M., van den Berg, C., Madriñán, S., Petersen, G., Wilkinson, M. (2007). A proposal for a standardized protocol to barcode all land plants. Taxon. 56: 295-299.

DeMattia, F., Bruni, I., Galimberti, A., Cattaneo, F., Casiraghi, M., Labra, M. (2011). A comparative study of different DNA barcoding markers for the identification of some members of Lamiaceae. Food Research International. 44: 693-702.

Sharma, K., Mishra, A.K., Misra, R.S. (2008). A simple and efficient method for extraction of genomic DNA from tropical tuber crops. African Journal of Biotechnology. 7(8): 1018-1022.

Fu, Y.X., Li, W.H. (1993). Statistical tests for neutrality of mutations. Genetics. 133: 693-709.

Gamache, J., Sun, G. (2015). Phylogenetic analysis of the genus Pseudoroegneria and the Triticeae tribe using the *rbcL* gene. Biochemical Systematics and Ecology. 62: 73-81.

Hebert, P.D.N, Cywinska, A., Ball, S.L., de Waard, J.R. (2003). Biological identification through DNA barcodes. Proceedings of the Royal Society of London. 270: 313-322.

Kress, W.J., Erickson, D.L. (2007) .A Two-Locus Global DNA Barcode for land plants: The Coding *rbcL* Gene Complements the Non-Coding *trnH-psbA* Spacer Region. Plos One. 2e508.

Kress, W.J., Erickson, D.L., Jones, F.A., Swenson, N.G., Perez, R., Sanjur, O., Bermingham, E. (2009). Plant DNA barcodes and a community phylogeny of a tropical forest dynamics plot in Panama. Proceedings of the National Academy of Sciences USA. 106: 18621-18626.

Kumar, S., Stecher,G., Tamura, K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. Molecular Biology and Evolution. 33: 1870-1874.

Long, D.T., Hong, H.T.K., Tram, L.L.T., Trang, N.T.Q., Tien, N.P.T. and Hanh, N.T.N. (2020). Study of the Procedure for Bath Ultrasound-Assisted Extraction of Total Flavonoid from Lotus Seeds and Testing Some Biological Activities. Indian Journal of Agricultural Research. DOI: 10.18805/IJARe.A-525.

Maqbool, S., Ullah, N., Zaman, A., Akbar, A., Saeed, S., Nawaz, H., Samad, N., Ullah, R., Bari, A. and Ali, S.S. (2019). Phytochemical Screening, In-vitro and In-vivo Anti-diabetic Activity of *Nelumbo nucifera* Leaves Against Alloxan-Induced Diabetic Rabbits. Indian Journal of Animal Research. 54(4): 1-6.

Newmaster, S.G., Ragupathy, S. (2009). Testing plant barcoding in a sister species complex of pantropical Acacia (Mimosoideae, Fabaceae). Molecular Ecology Resources. 9: 172-180.

Rozas, J., Ferrer-Mata, A., Sánchez-DelBarrio, J.C., Guirao-Rico, S., Librado, P., Ramos-Onsins, S.E., Sánchez-Gracia, A. (2017). DnaSP 6: DNA Sequence Polymorphism Analysis of Large Data Sets. Molecular Biology and Evolution. 34(12): 3299-3302.

Saitou, N., Nei, M. (1987). The neighbor-joining method: A new method for reconstructing phylogenetic trees. Molecular Biology and Evolution. 4: 406-425.

Seberg, O., Petersen, G. (2009). How many loci does it take to DNA barcode a crocus. PLoS ONE. 4 e4598.

Singh, H.K., Parveen, I., Raghuvanshi, S., Babbar, S.B.(2012). The loci recommended as universal barcodes for plants on the basis of floristic studies may not work with congeneric species as exemplified by DNA barcoding of Dendrobium species. BMC Research Notes. 5: 42.

Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics. 123: 585-595.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., Kumar, S. (2013). MEGA6: Molecular evolutionary genetics analysis version 6.0. Molecular Biology and Evolution. 30: 2725-2729.

Theodoridis, S., Stefanaki, A., Tezcan, M., Aki, C., Kokkin, S., Vlachonasios, K.E. (2012). DNA barcoding in native plants of the Labiatae (Lamiaceae) family from Chios Island (Greece) and the adjacent Çesme-Karaburun Peninsula (Turkey). Molecular Ecology Resources. 12: 620-633.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., Higgins, D.G. (1997). The CLUSTAL_X windows interface:flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Research. 24: 4876-4882.

Tripathi, A.M., Tyagi, A., Kumar. A,, Singh, A., Singh, S., Chaudhary, L.B., Roy, S. (2013). The Internal Transcribed Spacer (ITS) region and *trnH-psbA* are suitable candidate loci for DNA barcoding of tropical tree species of India. PLoS One. 8: e57934.

Yan, H.F., Hao, G., Hu, C.M., Ge, X.J. (2011). DNA Barcoding in closely related species: A case study of Primula L. sect. Proliferae Pax (Primulaceae) in China. Journal of Systematics and Evolution. 49: 225-236.

Zimmermann, T., Bocksberger, G., Brüggemann, G., Berberich, W. (2013). Phylogenetic relationship and molecular taxonomy of African grasses of the genus Panicum inferred from four chloroplast DNA-barcodes and nuclear gene sequences. Journal of Plant Research. 126(3): 363-371.