Plant Cell Biotechnology and Molecular Biology 20(23&24):1160–1171; 2019 ISSN: 0972-2025

# GENETIC DIVERSITY ANALYSIS OF LOTUS SPECIES (*Nelumbo nucifera* GAERTN.) IN THUA THIEN HUE BASED ON ITS4-5 GENETIC REGION

**DANG THANH LONG[#], HOANG THI KIM HONG[*], LE LY THUY TRAM, NGUYEN THI QUYNH TRANG, NGUYEN VAN HIEP AND NGUYEN PHAN THUY TIEN**

Institute of Biotechnology, Hue University, Provincial Road No. 10, Phu Vang, TTHue, Vietnam [DTL, NVH].

University of Sciences, Hue University, 77 Nguyen Hue, Hue, Vietnam [HTKH, NPTT].

University of Science and Technology, Da Nang University, 54 Nguyen Luong Bang, Da Nang, Vietnam [LLTT].

University of Education, Hue University, 32 Le Loi, Hue, Vietnam [NTQT].

[*For Correspondence: E-mail: hkhong@hueuni.edu.vn]

[#]Email: dtlong@hueuni.edu.vn

***Received: 10 October 2019***
***Accepted: 14 December 2019***
***Published: 20 December 2019***

***Original Research Article***

**ABSTRACT**

**In this study, we isolated and analysed the sequence of *ITS4-5* genetic region of thirty three lotus samples which were collected in Thua Thien Hue province. The attained *ITS4-5* genetic region had a length of ranged from 729 to 744 bp, which had high similarity with species *Nelumbo nucifera* (accession number: FJ599761.1), query cover was at 98%. Eight parameters were used to measure the DNA polymorphism of thirty three homologous DNA sequences in this population. The results indicated that, the number of separate polymorphic sites (S), total number of mutant sites (Eta), number of haplotype (h), haplotype diversity (Hd), average number of nucleotide differences (k), nucleotide diversity (π), theta (per site) from Eta (Ø) were 5 (S), 5 (Eta), 2(h), 0,458(Hd), 2,292 (k), 0,314 x $10^{-3}$ (π) and 0,169 x$10^{-3}$ (Ø), respectively (*P < 0,05)*. Our analysis indicated that Rm of ITS4-5 genetic region in lotus populations did not appear (Rm = 0). Two methods (Tajima's D test, Fu and Li's D* and F* test) were used to execute neutrality test. The results showed that, the evolution of lotus population was balancing selection, sudden population contraction, in other words, rare alleles present at low frequency and there are very few individuals that show significant differences from other individuals in the population. Phylogenetic tree was built based on three methods, namely Neightbour -Joining, Maximum Likelihood và Maximum Parsimony (bootstrap = 1000) showed that, thirty three collected lotus samples closely linked and they were divided into two groups. Group I included 22 samples of pink lotus varieties and group II included 11 samples of white lotus varieties. Through this result, we found that the ability to distinguish between white and pink lotus varieties of *N. nucifera* was significant when using the *ITS4-5* genetic region.**

**Keywords:** *Nelumbo nucifera*; ITS4-5; genetic diversity analysis; Thua Thien Hue; Lotus.

## INTRODUCTION

Lotus (*Nelumbo nucifera* Gaertn.) is a perennial aquatic basal eudicot belonging to a small family Nelumbonaceace, which contains only one genus with two species. It is an important horticultural plant, with its uses ranging from ornamental, nutritional to medicinal values, and has been

widely used, especially in Southeast Asia. Recently, the lotus obtained a lot of attention from the scientific community. An increasing number of research papers focusing on it have been published, which have shed light on the mysteries of this species [1]. 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 [2]. Molecular techniques and genetic technologies are now widely used in plant biodiversity researches, including species identification [3], in which DNA barcoding technique helps to classify species among the largest groups [4], solves the problems of different species classification [5], discovers new species at the same time [6], and builds phylogenetic trees allowing to research on community ecology [7,8].

DNA barcoding is a process of using common DNA sequences associated with the names of species for individuals which are sampled for research (http: //www.barcodeo fl ife.org/) [9]. Numerous studies have conducted the use of experimental bar code DNA regions in plants based on large data, spreading over many terrestrial plant species, or at least angiosperms [9,10,11,12,13]. The purpose of these studies is to assess the applicability of common bar code DNA regions in distinguishing species. However, the research groups have all argued that success in distinguishing species based on bar code DNA regions will create some problems (i) a number of plant groups have complex biological characteristics and (ii) closely related species in the same genus (or in recent evolutionary groups) [14,15,16].

In recent classified botanical studies, the ITS genetic region is the most commonly decoded locus. The region is highly effective in the classification of a variety of plants and fungi (except ferns), and this is a locus used for short DNA sequencing [17]. At the species level, the ITS genetic region has a high diversity (about 13.6% between closely related species) and has been demonstrated in almost all studies. The ITS genetic region has also been shown to have low levels of variation within the species [6]. Today, in the presence of more than 100.000 ITS sequences (as of December 2016) published on Genbank, this is a valuable resource, opening great prospects for species identification researches. The amount of sequences continues to be added daily. Because of these reasons, we isolated and read the ITS4-5 genetic sequence for 33 selected and collected lotus samples in Thua Thien Hue. The results of the study will be the basis for using the procedure to isolate the ITS4-5 genetic region and other genetic regions for *Nelumbo* genus species. This will be the data source for us to conduct a comprehensive analysis of molecular phylogeny for the genus *Nelumbo.*

## MATERIALS AND METHODS

### Materials

There are 33 lotus leaf samples, belonging to two varieties of white and pink lotus which were selected and collected on 33 different locations in Thua Thien Hue province (Table 1). After collecting, they are washed with distilled water and then refrigerated in the dark for further experiments.

### Methods

#### Total DNA extraction method

*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 65ºC 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 (-20ºC); 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 37ºC for 1h, and after stored at -20ºC. The process of DNA extraction is performed by basic CTAB method as described by Doyle [18].

**Table 1. Sample list collected lotus used in the study**

| Number | Sign | Collected sample location |
|---|---|---|
| 1 | ST01 | Concave white lotus in Tinh Tam lake<br>(Thuan Loc ward – Hue) |
| 2 | ST02 | Convex white lotus in Tinh Tam lake<br>(Thuan Loc ward – Hue) |
| 3 | ST03 | Concave white lotus, in Institute of biotechnology lake<br>(Phu Thuong - Phu Vang - Thua Thien Hue) |
| 4 | ST04 | Convex white lotus in Gia Long tomb lake<br>(Huong Tho - Huong Tra - Thua Thien Hue) |
| 5 | ST05 | Concave white lotus in Minh Mang tomb lake<br>(Huong Tho - Huong Tra - Thua Thien Hue) |
| 6 | ST06 | Concave white lotus in Thai dich lake<br>(Citadel, Thuan Thanh – Hue) |
| 7 | ST07 | Concave white lotus in Hoa Binh lake (Noi Kim Thuy)<br>(Citadel, Thuan Thanh – Hue) |
| 8 | ST08 | Concave white lotus in Ngoc Dich lake<br>(Citadel, Thuan Thanh – Hue) |
| 9 | ST09 | Concave white lotus in Huong Phong<br>Huong Tra – Thua Thien Hue |
| 10 | ST10 | Concave white lotus in Dien lake<br>Thuy Bieu - Thua Thien Hue |
| 11 | ST11 | Concave white lotus in Mung lake<br>Thuan Loc – Hue |
| 12 | SH01 | Phu Mong pink lotus<br>(Kim Long – Hue) |
| 13 | SH02 | Pink lotus in Gia Long tomb lake<br>(Huong Tho – Huong Tra) |
| 14 | SH03 | Phu Mong pink lotus in Tinh Tam lake<br>(Thuan Thanh - Hue) |
| 15 | SH04 | High yield pink lotus in Cua Huu lake<br>(Thuan Thanh - Hue) |
| 16 | SH05 | Phu My red lotus (Hoang Van Tuan)<br>Phu My – Phu Vang – Thua Thien Hue |
| 17 | SH06 | Thuy Van red pink lotus<br>Thuy Van – Huong Thuy – Thua Thien Hue |
| 18 | SH07 | High yield pink lotus in Huong Xuan ward<br>Huong Xuan – Huong Tra – Thua Thien hue |
| 19 | SH08 | Vinh Hai pink lotus<br>Vinh Hai – Phu Loc – Thua Thien Hue |
| 20 | SH09 | High yield pink lotus Phong An<br>Phong An – Phong Dien – Thua Thien Hue |
| 21 | SH10 | Vinh Thanh red lotus<br>Vinh Thanh - Phu Vang – Thua Thien Hue |
| 22 | SH11 | Vinh An red lotus<br>Vinh An – Phu Vang – Thua Thien hue |
| 23 | SH12 | Phu Thuong red lotus<br>Phu thuong – Phu Vang – Thua Thien Hue |

| Number | Sign | Collected sample location |
|---|---|---|
| 24 | SH13 | Pink lotus in Minh Mang tomb<br>Huong Tho – Huong Tra - Thua Thien Hue |
| 25 | SH14 | High yield pink lotus in Hoang Thanh lake<br>(Thuan Hoa ward – Hue) |
| 26 | SH15 | High yield lotus in Xa Tac lake<br>(Thuan Hoa ward – Hue) |
| 27 | SH16 | High yield lotus in Tan Mieu lake<br>(Thuan Hoa ward – Hue) |
| 28 | SH17 | High yield pink lotus in Huu Bao lake<br>(Tay Loc ward – Hue) |
| 29 | SH18 | High yield lotus in Nhon Hau lake<br>(Thuan Loc ward – Hue) |
| 30 | SH19 | Xuan Phu pink lotus<br>(Xuan Phu ward – Hue) |
| 31 | SH20 | Phu Mong pink lotus Phong An<br>Phong An – Phong Dien – Thua Thien Hue |
| 32 | SH21 | Phu Mong pink lotus Phong Son<br>Phong Son – Phong Dien – Thua Thien Hue |
| 33 | SH22 | Phu Mong pink lous Phong Xuan<br>Phong Xuan – Phong Dien – Thua Thien Hue |

**PCR method**

Performing PCR reaction to amplify the ITS4-5 genetic region, originating from genome with a pair of ITS4 primers: TCC TCC GCT TAT TGA TAT GC and ITS5: GGA AGG AGA AGT CGT AAC AAG G [19]. The PCR reaction is performed on the MJ - MiniTM Persanol Thermal Cycle, Bio - Rad thermal cycler with a reaction component of: 25 µl PCR master mix 2 × (2.4 mM dNTP each, 0.3 units Taq DNA polymerase), 10 pmol of ITS4 primer, 10 pmol of ITS5 primer, 1 µl of total DNA (50 ng/µl) and sterile distilled water to a final volume of 50 µl. The ITS4-5 gene region is amplified with the following thermal cycle: 95°C/5 minutes; 30 cycles x (95°C/60 seconds; 55°C/50 seconds; 72°C/60 seconds); 72°C/10 minutes. PCR products are tested by electrophoresis on 1% agarose gel in TAE 1X buffer with Ethydium bromide dye and read electrophoresis images by direct UV reading system (UV-transillumnator, Model: DyNa Light).

**Sequencing and analyzing genetic relationships**

The PCR products of the ITS4-5 genetic region are purified with Isolate II PCR and Gel (Bioline) kits. Then, they are sequenced directly by the dideoxy terminator method on the ABI PRISM® 3100 Avant Genetic Analyzer (Applied Biosystems) at Maccrogen Company, Korea.

The nucleotide sequences are arranged based on the Clustals program [20] and corrected by using BioEdit 7.0.5 software [21].

Phylogenetic tree showing genetic relationship will be built by MEGA 7.0 software (The Molecular Evolution Genetics Analysis), based on 3 methods of Maximum Likehood (ML); Neighbor-Joining and Maximum Parsimony. To reinforce the position of branches in the phylogenetic tree, the bootstrap value was used 1.000 times [22].

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 ($\pi$), 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 [23]. Neutrality is tested based on two methods, Tajima's D test [24] and Fu and Li's D * and F * test [25] using DNASP 6.0 software.

**RESULTS AND DISCUSSION**

**Total DNA Extraction Results**

The DNA extraction results in Plate 1 showed that the quality of DNA was clean, unbroken and gave

a single tape about 23 kb. This DNA material was used for further experiments (Plate 1).

**PCR Result**

The results in Plate 2 indicated that all PCR products of the ITS4-5 genetic region in the studied lotus samples showed a single band with 100% amplification rate. All samples gave high DNA concentration and clear. The obtained size was approximately 750 bp in accordance with the initial expected size (Plate 2).

The PCR products of ITS4-5 genome were sequenced on ABI PRISM® 3100 Avant Genetic Analyzer (Applied Biosystems) by dideoxy terminator method. The results of the ITS4-5 genetic region were 729 bp (ST01, ST09 and ST10) and 744 bp for the remaining lotus samples. The BLAST result on NCBI was used to verify and compare with the sequences of the *N. nucifera* (accession number: FJ599761.1) showed that the nucleotide sequences obtained were highly similar to those of the *N. nucifera* (accession number: FJ599761.1), ranging from 96,89 to 99,32% and coverage of 98%. The percentage of occurrence of each type of nucleotide in the ITS4-5 genetic region showed that Guanidin (G) accounts for the highest proportion, ranging from 24,5 to 24,7%, followed by Adenin (A) accounting for 24,3 to 24,5% and the lowest was Timin (Uracin) accounting for 20,4 to 20,7% (Table 2). The percentage (G + C) contained in the genome was the highest at 55,2% and between different lotus samples, the differences ranged from 54,8% to 55,2% and reached an average of 55,1% (Table 2).

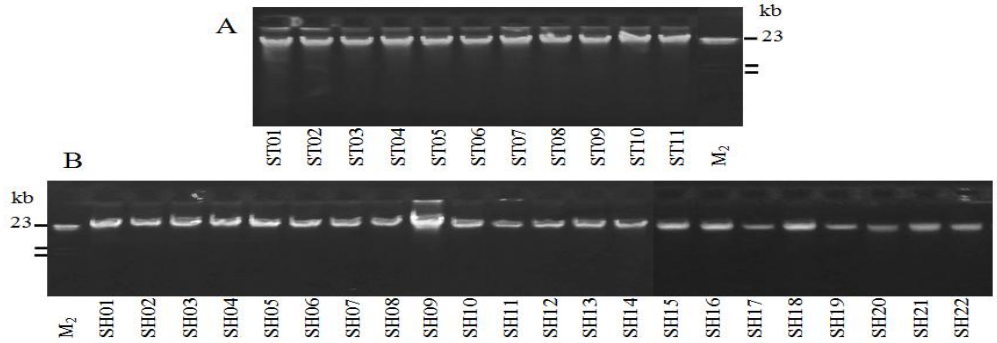


**Plate 1. Total DNA electrophoresis results of 33 lotus samples on 1% agarose gel. $M_2$: DNA mass scale (Lambda DNA / HindIII Marker (0,125 kb to 23 kb), Biotols); Plate 1A. Total DNA of 11 white lotus leaf samples; Plate 1B. Total DNA of 22 pink lotus leaf samples**

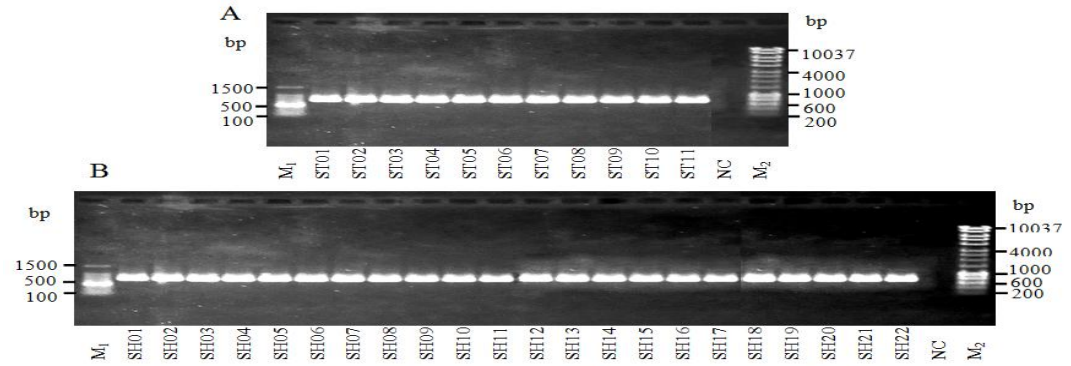


**Plate 2. Electrophoresis of PCR product. $M_1$: weight of DNA standard scale (100-1500 bp, BioBase); $M_2$: weight of DNA standard scale (200-10037 bp, Bioline), NC: negative control does not contain DNA template, Plate 2A. PCR product amplifies ITS4-5 gene region of 11 white lotus samples; Plate 2B. PCR product amplifies the ITS4-5 gene region of 22 pink lotus samples**

**Table 2. Nucleotide components of ITS4-5 genetic region of *N. nucifera***

| Sample | T(U) | C | A | G | Total | G+C (%) |
|---|---|---|---|---|---|---|
| ST01 | 20,7 | 30,2 | 24,4 | 24,7 | 729,0 | 54,9 |
| ST02 | 20,7 | 30,4 | 24,5 | 24,5 | 744,0 | 54,8 |
| ST03 | 20,7 | 30,4 | 24,5 | 24,5 | 744,0 | 54,8 |
| ST04 | 20,7 | 30,4 | 24,5 | 24,5 | 744,0 | 54,8 |
| ST05 | 20,7 | 30,4 | 24,5 | 24,5 | 744,0 | 54,8 |
| ST06 | 20,7 | 30,4 | 24,5 | 24,5 | 744,0 | 54,8 |
| ST07 | 20,7 | 30,4 | 24,5 | 24,5 | 744,0 | 54,8 |
| ST08 | 20,7 | 30,4 | 24,5 | 24,5 | 744,0 | 54,8 |
| ST09 | 20,7 | 30,2 | 24,4 | 24,7 | 729,0 | 54,9 |
| ST10 | 20,7 | 30,2 | 24,4 | 24,7 | 729,0 | 54,9 |
| ST11 | 20,7 | 30,4 | 24,5 | 24,5 | 744,0 | 54,8 |
| SH01 | 20,4 | 30,6 | 24,3 | 24,6 | 744,0 | 55,2 |
| SH02 | 20,4 | 30,6 | 24,3 | 24,6 | 744,0 | 55,2 |
| SH03 | 20,4 | 30,6 | 24,3 | 24,6 | 744,0 | 55,2 |
| SH04 | 20,4 | 30,6 | 24,3 | 24,6 | 744,0 | 55,2 |
| SH05 | 20,4 | 30,6 | 24,3 | 24,6 | 744,0 | 55,2 |
| SH06 | 20,4 | 30,6 | 24,3 | 24,6 | 744,0 | 55,2 |
| SH07 | 20,4 | 30,6 | 24,3 | 24,6 | 744,0 | 55,2 |
| SH08 | 20,4 | 30,6 | 24,3 | 24,6 | 744,0 | 55,2 |
| SH09 | 20,4 | 30,6 | 24,3 | 24,6 | 744,0 | 55,2 |
| SH10 | 20,4 | 30,6 | 24,3 | 24,6 | 744,0 | 55,2 |
| SH11 | 20,4 | 30,6 | 24,3 | 24,6 | 744,0 | 55,2 |
| SH12 | 20,4 | 30,6 | 24,3 | 24,6 | 744,0 | 55,2 |
| SH13 | 20,4 | 30,6 | 24,3 | 24,6 | 744,0 | 55,2 |
| SH14 | 20,4 | 30,6 | 24,3 | 24,6 | 744,0 | 55,2 |
| SH15 | 20,4 | 30,6 | 24,3 | 24,6 | 744,0 | 55,2 |
| SH16 | 20,4 | 30,6 | 24,3 | 24,6 | 744,0 | 55,2 |
| SH17 | 20,4 | 30,6 | 24,3 | 24,6 | 744,0 | 55,2 |
| SH18 | 20,4 | 30,6 | 24,3 | 24,6 | 744,0 | 55,2 |
| SH19 | 20,4 | 30,6 | 24,3 | 24,6 | 744,0 | 55,2 |
| SH20 | 20,4 | 30,6 | 24,3 | 24,6 | 744,0 | 55,2 |
| SH21 | 20,4 | 30,6 | 24,3 | 24,6 | 744,0 | 55,2 |
| SH22 | 20,4 | 30,6 | 24,3 | 24,6 | 744,0 | 55,2 |
| Avg. | 20,5 | 30,5 | 24,4 | 24,6 | 742,6 | 55,1 |

Eight parameters including number of polymorphic sites (S), total number of mutant sites (Eta), number of haplotypes (h), haplotype diversity (Hd), average number of nucleotide differences (k), nucleotide diversity ($\pi$), minimum number of recombinant processes (Rm) and number of effective populations for mutation rate at each nucleotide position per generation (Ø) were used to evaluate the diversity of 33 studied lotus samples belonging to two varieties of white lotus and pink lotus. The results of the sequencing analysis were calibrated, Align and analysed by using MEGA 7.0 software, obtained a conserved region of 724/744 nucleotide positions and a modified region of 20/744 nucleotide positions (Fig. 1). Five separate polymorphic nucleotide sites were found in the ITS4-5 gene region and one defective region (from 133 to 147 is the ACGTCCAGCATTCCA sequence) was presented in ST02, ST03, ST04, ST05, ST06, ST07, ST08, ST11 and SH01 to SH22, but not in ST01, ST09 and ST10 models. This defective gene sequence could be considered as a sign to build identification and distinguish between the lotus samples ST01, ST09 and ST10 compared to the remaining lotus samples.

In addition, five mutation positions 404; 408; 576; 625 and 626 when performing analysis with DNASP 6.0 software showed that there were 2 synonymous mutation positions 408 and 576 (G (T) and C (A)) and 3 mutation sites replaced 404; 625 and 626 (C (T); G (A) and A (G) (Fig. 1). Five separate polymorphic positions (S) created five mutant positions (Eta) shown in 33 studied lotus samples classified into two types of haplotype (h) with haplotype diversity coefficient accounting for 0,458 (Hd), the average number of nucleotide differences is 2,292 (k), the nucleotide diversity coefficient accounts for 0,314 x $10^{-3}$ ($\pi$),

the number of effective populations for the rate of mutations per nucleotide position per generation accounts for 0,169 $x10^{-3}$ (Ø), the minimum number of recombinants (Rm) to occur does not exist. All indicators were processed with statistical significance $p < 0,05$ (Table 3).

Number of separate polymorphic sites (S), total mutations (Eta), number of haplotypes (h), haplotype diversity (Hd), average number of distinct nucleotides (k), nucleotide diversity (π) and the number of viable populations for the rate of mutation per nucleotide position per generation (Ø) and the minimum number of recombinants (Rm) to occur.

Two methods namely (Tajima's D test, Fu and Li's D * and F * test) were used to test neutrality. The results in Table 4 with D value and Fu and Li's F * both yield positive values with statistical significance $p < 0,05$, this showed that the evolution of the studied lotus population was balancing selection, sudden contraction or in other words, rare alleles appeared in populations with low frequency. In addition, the value of Fu and Li's D * = 1,13477 (statistically insignificant with $p > 0,10$) indicated that the studied population had very few individuals showing large differences in comparision with other individuals in the population (Table 4).

After analyzing the genetic relationship of 33 nucleotide sequences of 33 lotus samples belonging to two varieties of white lotus and pink lotus, all gaps and missing data in the sequence have been removed. There have been a total of 717 nucleotide positions shown in the last sequence. The ITS4-5 sequence of *N.nucifera* lotus species was taken from Genebank for reference (accession number: FJ599761.1). Genetic tree was built based on three Neighbour-Joining methods; Maximum Likelihood and Maximum Parsimony with bootstrap values (1000 repetitions) on MEGA 7.0 software. For the Neighbour-Joining method, the optimal tree of genetic development in figure 2 [26] had the total branch length = 0,01115760. The percentage of trees that were replicated among the taxon linked together in the bootstrap test (1000 replicates) was displayed next to the branches [11]. The trees were plotted in proportion, with branch lengths in the same unit and evolutionary distance used to infer genetically generated trees. The evolutionary distance was calculated using the p-distance method [2]. For the Maximum Likelihood method, phylogenetic trees were built based on the Tamura-Nei model [27]. Displayed tree was the most reasonable (-1047,42). The percentage of trees that were repetitively processed between the classification units linked next to the branches (Fig. 3). The tree was initially built automatically using Neighbour-Join and BioNJ algorithms based on the distance of each pair and was estimated using the maximum aggregate optimization method (MCL), then selected the structure. Topology with the highest stable value. Plants were drawn in proportion, with branch lengths measured by the number of substitutions per nucleotide position. Genetically generated trees are based on the Maximum Parsimony method with a uniform index of (1.000000), a maintenance index of (1.000000) and a combined index of 1.000000 for all nucleotide chains to provide analytical information. The percentage of trees that are repetitively processed between the taxonomic units were linked next to the branches. Trees are plotted proportionally, with branch lengths calculated using the moving average method and calculated as a number of changes in the whole series (Fig. 4).

**Table 3. DNA diversity based on ITS4-5 genetic region of lotus population**

| Genetic region | S | Eta | H | | Hd | K | π ($x10^{-3}$) | Ø ($x10^{-3}$) | Rm |
|---|---|---|---|---|---|---|---|---|---|
| | | | $H_1$ | $H_2$ | | | | | |
| *ITS 4-5* | 5 | 5 | ST01; ST02; ST03; ST04; ST05; ST06; ST07; ST08; ST09; ST10 và ST11 | SH01; SH02; SH03; SH04; SH05; SH06; SH07; SH08; SH09; SH10; SH11; SH12; SH13; SH14; SH15; SH16; SH17; SH18; SH19; SH20; SH21 và H22 | 0,458 | 2,292 | 0,314 | 0,169 | 0 |

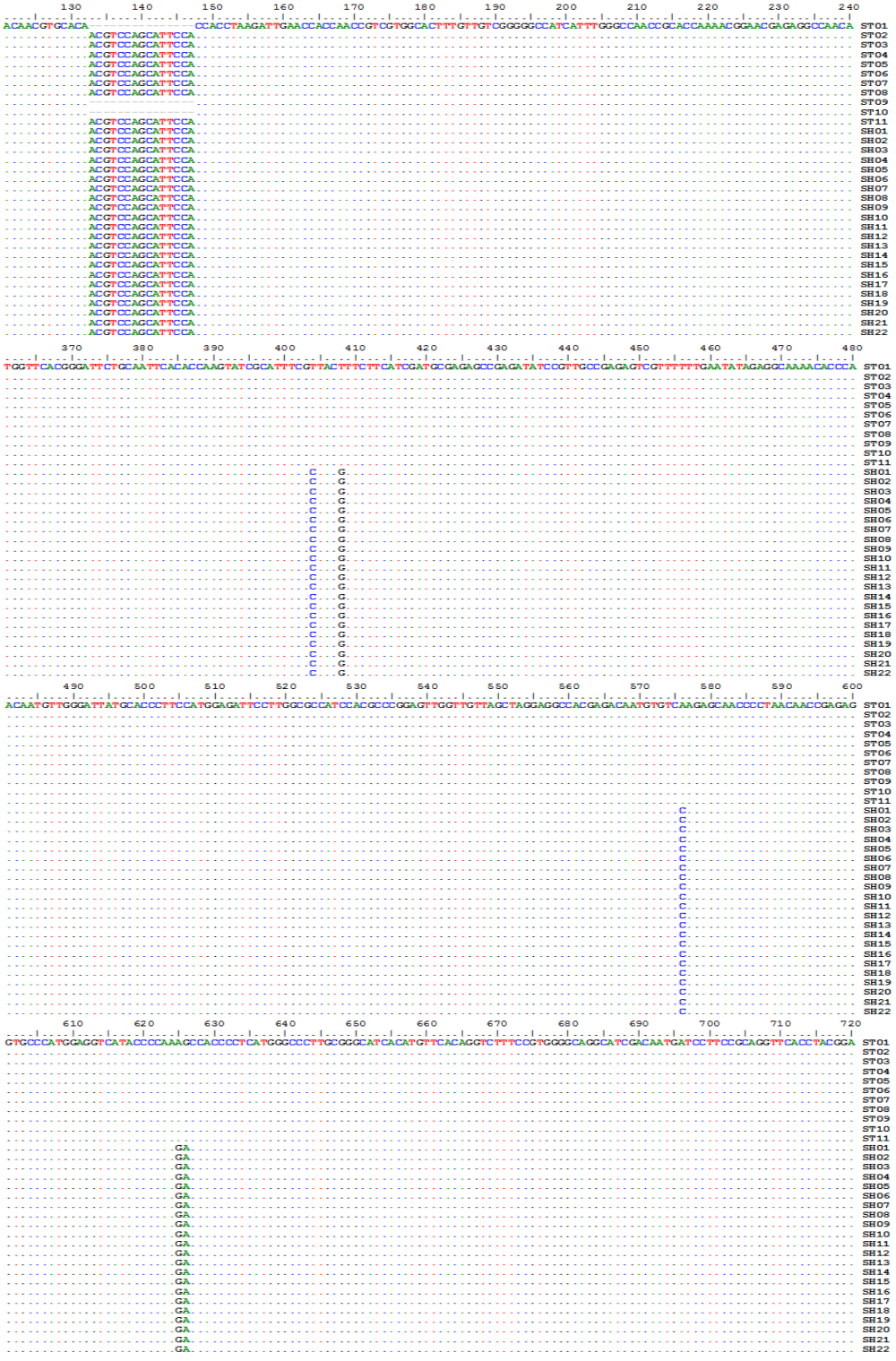

**Fig. 1. Results of analysis and comparison of nucleotide sequences of ITS4-5 gene region**

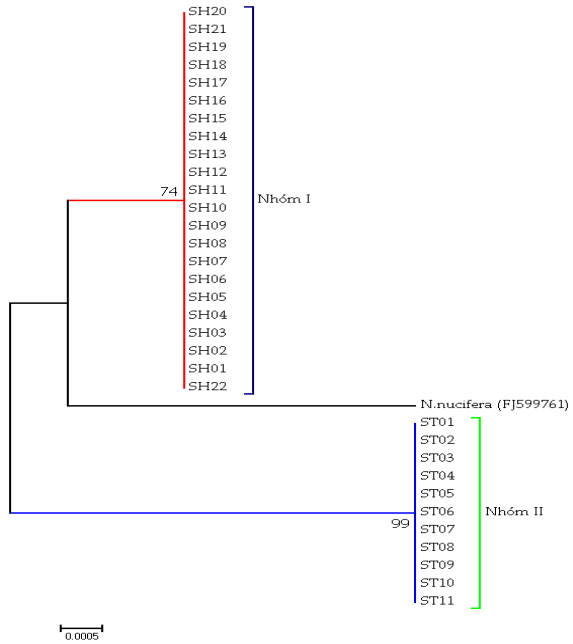


**Fig. 2. Phylogenetic tree of 33 lotus varieties collected based on the ITS4-5 gene region of the nucleus by Neighbour-Joining method [26] with a repeat of bootstrap 1000 times [11], the evolution distance was calculated by p-distance [28] based on MEGA7.0 software [22]**

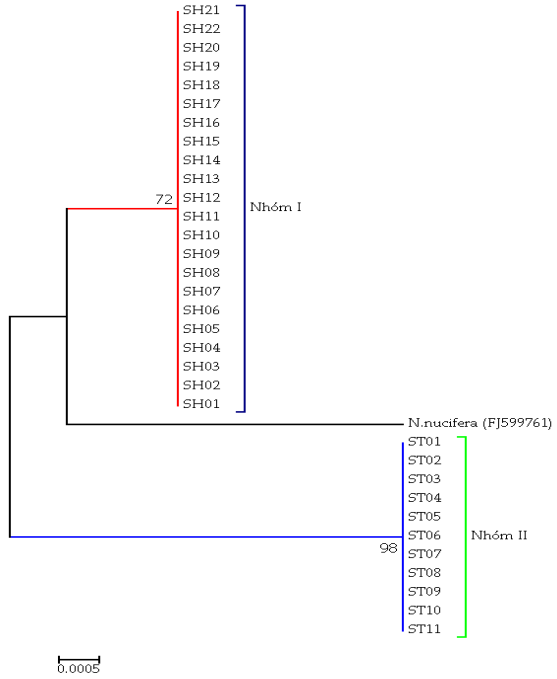


**Fig. 3. Phylogenetic tree of 33 lotus varieties collected based on the ITS4-5 genome of the nucleus by the Maximum Likelihood method based on the Tamura-Nei model [27] with a repeat of bootstrap 1000 times [11] on the section soft MEGA7 [22]**

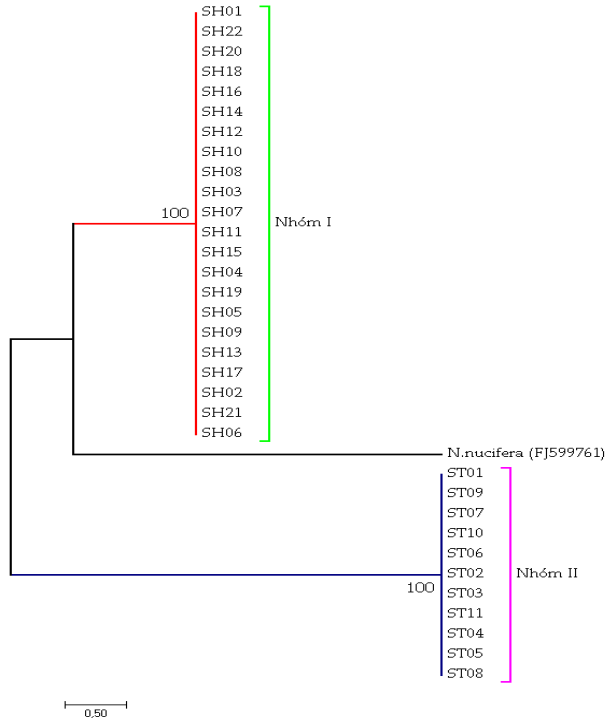


**Fig. 4. Phylogenetic tree of 33 lotus varieties collected based on the ITS4-5 gene region of the nucleus by Maximum Parsimony method with a repeat of bootstrap 1000 times [11], the evolution distance was calculated by p -distance [28] based on MEGA7 software [22]**

**Table 4. Neutral test results based on ITS4-5 genetic region of lotus population**

| Genetic region | Tajima's D test | Fu and Li's D* test | Fu and Li's F* test * |
|---|---|---|---|
| *ITS 4-5* | 2,29096<br>*Statistical significance with p<0,05* | 1,13477<br>*Statistical significance is not significant with p> 0,10* | 1,72353<br>*Statistical significance with p<0,05* |

The results in Figs. 2, 3 and 4 showed that the studied lotus samples closely relate to each other and are divided into two characteristic groups. Group I consists of 22 samples belonging to the pink lotus variety (including high-yielding lotus varieties imported from other localities and the pink lotus traditionally planted in Thua Thien Hue). Group II includes 11 specimens of white lotus variety with two different types of convex and concave lotus seed pods. All varieties of lotus belong to the species *N. nucifera*. Through these results, we found that the ability to distinguish between white and pink lotus varieties of *N. nucifera* is high when using ITS4-5 genetic region (Figs. 2, 3 and 4).

## CONCLUSION

In this study, we isolated and sequenced the ITS4-5 genetic region of 33 lotus samples collected in Thua Thien Hue, obtained its size ranging from 729 to 744 bp which was highly similar with *N. nucifera* lotus (accession number: FJ599761.1), query cover was at 98. The conservation zone contains 724/744 nucleotide positions, the transformed region contains 20/744 nucleotide positions. Guanidin (G) accounted for the highest proportion, ranging from 24,5 to 24,7%, followed by Adenin (A) accounting for 24,3 to 24,5% and the lowest was Timin (Uracin) accounting for 20,4 to 20,7%. The percentage (G + C)

fluctuated from 54,8% to 55,2% and averaged at 55,1%.

Five separate polymorphic positions (S) created five mutation positions (Eta) (404; 408; 576; 625 and 626) including two synonymous mutation positions 408 and 576 (G (T) and C (A)) and 3 replaced mutation positions 404; 625 and 626 (C (T); G (A) and A (G)). Thirty three lotus samples were classified into 2 types of haplotype (h) with haplotype diversity factor accounting for 0,458 (Hd), the average number of nucleotides was 2,292 (k), the nucleotide diversity factor was 0,314 x $10^{-3}$ ($\pi$ ), the number of effective populations for mutation rate at each nucleotide position per accounted for 0,169 x$10^{-3}$ (Ø), the minimum number of recombinant processes (Rm) was not available (all both indices were treated with statistical significance $p < 0,05$). Neutrality was tested based on two methods (Tajima's D test, Fu and Li's D * and F * test) showing that the evolution of lotus population was balancing selection, sudden population contraction, in other words, rare alleles were present at low frequency and there are very few individuals that showed significant differences from other individuals in the population. Phylogenetic tree was built based on three methods, namely Neightbour -Joining, Maximum Likelihood và Maximum Parsimony (bootstrap = 1000) showed that, thirty three collected lotus samples closely linked and they were divided into two groups. Group I included 22 samples of pink lotus varieties and group II included 11 samples of white lotus varieties. Through these results, we found that the ability to distinguish between white and pink lotus varieties of *N. nucifera* was significant when using the *ITS4-5* genetic region.

## ACKNOWLEDGEMENTS

This work was supported by the Domestic Master/PhD Scholarship Programme of Vingroup Innovation Foundation.

## COMPETING INTERESTS

Authors have declared that no competing interests exist.

## REFERENCES

1. Lin Z, Zhang C, Cao D, Damaris RN, Yang P. The latest studies on Lotus (*Nelumbo nucifera*)-an emerging horticultural model plant: Review. International Journal of Molecular Sciences. 2019;20:1-13. DOI: 10.3390/ijms20153680
2. Trias-Blasi A, Vorontsova M. Botany: Plant identification is key to conservation. Nature. 2015;521:161. DOI: 10.1038/521161c
3. Savolainen O, Karhu A. Assessment of biodiversity with molecular tools in forest trees. In Molecular Biology of Woody Plants, Eds S. M. Jain and S. C. Minocha (Dordrecht: Springer). 2000;395-406.
4. Ojeda DI, Santos-Guerra A, Oliva-Tejera F, Jaen-Molina R, Caujapé-Castells J, Marrero-Rodríguez A, Cronk Q. DNA barcodes successfully identified Macaronesian Lotus (Leguminosae) species within early diverged lineages of Cape Verde and Mainland Africa. AoB PLANTS. 2014;6:plu050. DOI: 10.1093/aobpla/plu050
5. Feau N, Vialle A, Allaire M, Maier W, Hamelin RC. DNA barcoding in the rust genus Chrysomyxa and its implications for the phylogeny of the genus. Mycologia. 2011;103:250-1266. DOI: 10.3852/10-426
6. Baldwin BG, Sanderson MJ, Porter JM, Wojciechowski MF, Campbell CS, Donoghue MJ. The ITS region of nuclear ribosomal DNA-A valuable source of evidence on Angiosperm phylogeny. Annals of the Missouri Botanical Garden. 1995;82:247-82.
7. Joly S, Davies TJ, Archambault A, Bruneau A, Derry A, Kembel SW, Peres-Neto P, Vamosi J, Wheeler TA. Ecology in the age of DNA barcoding: The resource, the promise and the challenges ahead. Molecular Ecology Resources. 2013;14: 221-232.
8. Kress WJ, García-Robledo C, Uriarte M, Erickson DL. DNA barcodes for ecology, evolution, and conservation. Trends in Ecology & Evolution. 2015;30:25-35. DOI: 10.1016/j.tree.2014.10.008

9. CBOL Plant Working Group. A DNA barcode for land plants. Proceedings of the National Academy of Sciences of the USA. 2009;106:12794-12797.
10. Fazekas AJ, Burgess KS, Kesanakurti PR, Graham SW, Newmaster SG, Husband BC, Percy DM, Hajibabaei M, Barrett SC. Multiple multilocus DNA barcodes from the plastid genome discriminate plant species equally well. PLoS ONE. 2008;3:e2802.
11. Felsenstein J. Confidence limits on phylogenies: An approach using the bootstrap. Evolution. 1985;39:783-791.
12. Kress WJ, Erickson DL. A two-locus global DNA barcode for land plants: The coding rbcL gene complements the non-coding trnH-psbA spacer region. PLoS ONE. 2007;2:e508.
13. Lahaye R, Van Der Bank M, Bogarin, Warner J, Pupulin F, Gigot G, Maurin O, Duthoit S, Barraclough TG, Savolainen V. DNA barcoding the floras of biodiversity hotspots. Proceedings of the National Academy of Sciences of the USA. 2008;105:2923-2928.
14. Miller D, Bohs ML, Spooner DM. DNA barcoding will frequently fail in complicated groups: An example in wild potatoes. Amerian Journal of Botany. 2009;96:1176-1189.
15. Pillon Y, Johansen J, Sakishima T, Chamala S, Barbazuk WB, Roalson EH, Price DK, Stacy EA. Potential use of low-copy nuclear genes in DNA barcoding: A comparison with plastid genes in two Hawaiian plant radiations. BMC Evolutionary; 2013.
16. Seberg O, Petersen G. How many loci does it take to DNA barcode a crocus. PLoS ONE. 2009;4:e4598.
17. Stoeckle M. Taxonomy, DNA and the bar code of life. BioScience. 2003;53:2-3.
18. Doyle JJ, Doyle JL. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem Bulletin. 1987;19:11-15.
19. White TJ, Bruns T, Lee S, Taylor J. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In PCR protocols a guide to methods and applications. Academic Press, San Diego. 1990;315-322.
20. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. The CLUSTAL_X windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Research. 1997;2(4):4876-4882.
21. Hall TA. BIOEDIT: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/ NT. Nucleic Acids Symposium Series. 1999;41:95-98.
22. Kumar S, Stecher G, Tamura K. MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. Molecular Biology and Evolution. 2016;33:1870-1874.
23. Rozas J, Ferrer-Mata A, Sánchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE, Sánchez-Gracia A. DNASP 6: DNA sequence polymorphism analysis of large datasets. Molecular Biology and Evolution. 2017;34:3299-3302.
DOI: 10.1093/mo.
24. Tajima F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics. 1989a;123:585-595.
25. Fu YX, Li WH. Statistical tests of neutrality of mutations. Genetics.1993;133:693-709.
26. Saitou N, Nei M. The neighbor-joining method: A new method for reconstructing phylogenetic trees. Molecular Biology and Evolution. 1987;4:406-425.
27. Tamura K, Nei M. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Molecular Biology and Evolution. 1993;10:512-526.
28. Nei M, Kumar S. Molecular evolution and phylogenetics. Oxford University Press, New York; 2000.