A-646
[1-6]

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

Research on Phylogenetic Relationship of Lotus Populations Collected in Thua Thien Hue Province, Vietnam based on the Chloroplast Genome by DNA Barcode

Dang Thanh Long[1], Hoang Thi Kim Hong[2], Le Ly Thuy Tram[3], Nguyen Thi Quynh Trang[4] 10.18805/IJARe.A-646

ABTRACT

Background: The DNA barcoding is currently an effective and widely used tool that enables rapid and accurate identification of plant species.

Methods: DNA barcoding of 9 chloroplast genes (*rbcL, matK, trnH-psbA, accD-psaI, ndhA, psbE-petL, Rpl32-trnL, trnW-psaJ, trnS*^{GCU}*-trnG*^{GCC}) were used to provide the theoretical basis for species identification, genetic diversity analysis of lotus population collected in Thua Thien Hue province, Vietnam. Universal primers were used and sequence products were analyzed using the MEGA X program.

Result: The results showed that high levels of haplotype diversity (Hd), ranging from 0.618-0.869 and low levels of nucleotide diversity (Pi), ranging from 0.180 x 10^{-3}-3.280 x 10^{-3} base on a total of nine gene regions of chloroplast genome. The neutrality tests show an excess of rare nucleotide position variations in individuals' white lotus and derived haplotypes recent expansion. While the evolution of the individuals in the pink lotus may have to decrease. The phylogenetic analyses indicated that combined sequences were not insufficient to make a difference to the DNA barcoding in the individual's lotus of the *N. nucifera* species this is in the study. The standardized and accurate barcode information of lotus is provided for researchers. It lays the foundation for the conservation, evaluation, innovative utilization and protection of Nelumbonaceae germplasm resources.

Key words: Chloroplast, Genetic diversity, Lotus, *N. nucifera*.

INTRODUCTION

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* (*N. nucifera*) is known by numerous common names like sacred lotus, bean of India and simply lotus. *N. nucifera* is an aquatic herb with white or red-colored flowers (Maqbool *et al.* 2019). Since ancient times, the lotus flower has become familiar and close to Vietnamese's life as well as other countries such as India, China and Japan. Besides, 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 that are related to the ecosystem. Therefore, it is necessary to conserve and manage this species (Long *et al.* 2020).

Recently, there has been increasing interest in molecular techniques that have been widely used to analyze phylogenetic relationships among various breeds/populations (Avise *et al.* 1987a). 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. An ideal DNA barcode should 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 on the studied group taxa (Long *et al.* 2020). *N. nucifera* is an aquatic flowering plant. This perennial usually lives in lakes and ponds.

This study was undertaken to test the utility of DNA barcoding to provide the theoretical basis for species

[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. (2021). Research on Phylogenetic Relationship of Lotus Populations Collected in Thua Thien Hue Province, Vietnam based on the Chloroplast Genome by DNA Barcode. Indian Journal of Agricultural Research. DOI: 10.18805/IJARe.A-646.

Submitted: 23-04-2021 **Accepted:** 23-09-2021 **Online:** 08-10-2021

identification, genetic diversity analysis of lotus population collected in Thua Thien Hue province, Vietnam. This study shall help make an informed decision on conservation, utilization and exploitation of the lotus genetic resources of the Thua Thien Hue province.

MATERIALS AND METHODS

Plant materials

The lotus leaf samples (thirty-three samples: white and pink lotus) were collected from 33 different locations in Thua Thien Hue province (Fig 1). These 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, May, 2018.

Methods

DNA extraction, PCR amplification and sequencing analysis

Total DNA was extracted from fresh leaf tissue (lotus leaf samples were stored at 4°C in the dark for about 1 to 2 days to remove part of the starch existing in leaf tissue) by using a CTAB method as described by Sharma *et al.* (2008).

The chloroplast genome DNA barcodes were amplified in a 25 µL reaction volume with universal primer sets presented in Table 1, 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 Applied Biosystems-Life Technologies (Thermo Fisher Scientific Inc.. United States). PCR cycling conditions used in this study: 95°C/5 minutes; 30 cycles x (95°C/60 seconds; A°C/30 seconds; 72°C/60 seconds); 72°C/ 10 minutes (Annealing temperature is shown in Table 1).

PCR products were tested by electrophoresis on 1% agarose gel in TAE 1X buffer with Ethidium bromide dye and read electrophoresis images by direct UV reading system (UV-transilluminator, Model: Dyna Light). Samples showing a clear single band were sent to Maccrogen Company, Korea and sequenced in both directions with the same primers used for PCR by the method dideoxy terminator method on the ABI PRISM® 3100 Avant Genetic Analyzer (Applied Biosystems). Sequences were uploaded to the GenBank (Accession numbers: *rbcL* (MN011708 to MN068956); *matK* (MN011719 to MN068978) and *trnH-psbA* (MN011730 to MN086252); *accD-psaI* (MN086253 to MN086285); *psbE-petL (*MT901764 to MT901796); *Rpl32-trnL (*MT901731 to MT901763); *trnW-psaJ (*MT905225 to MT905257); *trnS*^{GCU}*-trnG*^{GCC} (MT905258 to MT905290) and *ndhA* ((MZ611976 to MZ612008), respectively.

Data analysis

Raw sequences for each region were assembled and edited using BioEdit v7.2.5. Edited sequences were then aligned by ClustalW in MEGA X and the non-overlapping sequence regions at the 52 - and 32 -ends were trimmed (Kumar *et al.* 2018). The evolutionary history was inferred by using the Maximum Likelihood method and Tamura-Nei model (Tamura and Nei, 1993). The tree with the highest log likelihood (-12544.64) is shown. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. Evolutionary analyses were conducted in MEGA X (Kumar *et al.* 2018). The barcode sequences were queried against the GeneBank database (NCBI) using the Nucleotide BLAST algorithm.

The seven parameters including the number of separate polymorphic sites (S), the total number of mutant sites (Eta), number of haplotypes (h), haplotype diversity (Hd), the average number of nucleotide differences (k), nucleotide diversity (Pi) and Minimum number of recombination events (Rm) are considered as a polymorphic measurement in the population. Neutrality is tested based on five methods namely (Tajima's D test (Tajima, 1989), Fs, Fu's statistic (Fu, 1997); S, Strobeck's statistic (Stroreck, 1987); D* and

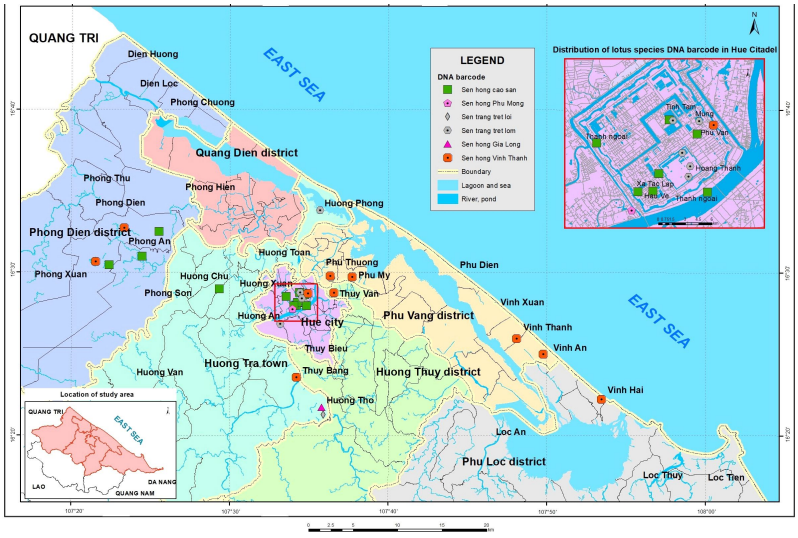

Fig 1: Location of collecting lotus samples.

F*, Fu and Li's statistics (Fu and Li, 1993) were used to DNAsp 6.0 software (Rozas *et al*. 2017).

RESULTS AND DISCUSSION

Sequence characteristics of the barcode

The nine DNA barcodes of the lotus chloroplast genome showed high success rates for PCR amplification and sequencing using specific single primer pair. Sequence analysis results showed that the total length of nine chloroplast DNA regions of the two lotus study populations ranged from 8892-8921 bp (white lotus) and 8861-8916 bp (pink lotus) (Table 2). BLAST results on the National Center for Biotechnology Information (NCBI) showed that the obtained nucleotide sequence is very similar to that of the lotus species *N. nucifera* (Accecssion number: KF009944.1).

Sequence characteristics of a total of nine regions are present in Table 2 shows the nine sequences the within mean group distance in each lotus population was 0.180±0.900 × 10^{-3} (white Lotus) and 3.390±0.450 × 10^{-3} (pink Lotus). G+C content contained in the nine genetic regions is between different lotus samples was low, ranging from 0.334-0.335 (Table 2). The appearing percentage of each nucleotide type showed that Adenine (A) and Timin (Uracin) accounted for the highest proportion and there was no high difference between studied lotus samples and ranged from 32.863 to 33.180% (A) and 33.546 to 33.676% (T), reaching an average of 33.085 and 33.595, respectively. Meanwhile, the lowest proportions were Cysteine (C) and Guanidin (G), ranging from 17.207% to 17.454% (C) and from 15.971% to 16.138% (G), reaching an average of 17.266% (C) and 16.054% (G) of the lotus samples studied.

Besides, analysis of a total of nine genomic regions of two lotus populations using DNAsp 6.0 software showed that there are five separate polymorphic positions (S), create five mutant positions (Eta) for the white lotus population; eighty-four separate polymorphic positions (S) create eighty-five mutant positions (Eta) for the pink lotus population and eighty-five separate polymorphic positions (S) create eighty-six mutant positions (Eta) for the total of both lotus populations. A total of nine genomic regions with these differences were divided into three types of haplotypes (h) for the white lotus population; seven types of haplotypes (h) for the pink lotus population and nine types of haplotypes (h) for the total of both lotus populations. The haplotype diversity coefficient (Hd) accounts for 0.618±0.104 (white lotus); 0.818±0.050 (pink lotus) and 0.869±0.026 (total of both lotus populations), the average number of nucleotide

Table 1: Primers used for amplification and sequencing.

Regions	Primer names	Primer sequences 5¹'!3¹	Annealing temperature (°C)	References
rbcL	1F	ATGTCACCACAAACAGAGAC	50	KF009944.1 and
	743R	TCACATGTACCTGCAGTAGC		Dong *et al.* (2012)
matK	385F	CGATCAATTCATTCAATATTTTC	50	
	1320R	ACTTCGACTTTCGTGTGCTAGA		
trnH-psbA	46F	ACTGCCTTGATCCACTTGGC	50	
	25R	TGAAGCTCCATCTACAAATGG		
accD-psaI	accD-F	GGTAAAAGAGTAGTTGAACAAAC	55	
	psaI-R	GGAAATACTAGGCCCACTAAAGGCACA		
ndhA	ndhA-F	CAACTATATCAACTGTACTTGAAC	53	
	ndhA-R	CGAGCTGCTGCTCAATCGAT		
psbE-petL	psbE-F	ATCTACTAAATTCATCGAGTTGTTCC	53	
	petL-R	TATCTTGCTCAGACCAATAAATAGA		
Rpl32-trnL	rpl32-F	GCGTATTCGTAAAAATATTTGGAA	51	
	trnL-R	TTCCTAAGAGCAGCGTGTCTACC		
trnW - psaJ	trnW-F	TACCGAACTGAACTAAGAGCGC	55	
	psaJ-R	CGATTGATCTCTATCAAAAGACCTGC		
$trnS^{GCU}$-$trnG^{GCC}$	trnS1-F	AACGGATTAGCAATCCGACGCTTTA	51	
	trnG1-R	CTTTTACCACTAAACTATACCCGC		

Table 2: The results of DNA polymorphism of lotus populations.

Population	n	Total aligned length (bp)	Within mean group distance x10^{-3}	G+C content	S	Eta	h	Hd ± SD	k x 10^{-2}	Pi ± SD x 10^{-3}	Rm
Lotus white	11	8892-8921	0.180 ± 0.900	0.334	5	5	3	0.618 ± 0.104	156.364	0.180 ± 0.050	0
Lotus pink	22	8861-8916	3.390 ± 0.450	0.335	84	85	7	0.818 ±0.050	2908.225	3.280 ± 0.830	1
Total	33	-	-	0.335	85	86	9	0.869 ±0.026	2210.985	2.550 ± 0.720	1

Note- n: Number of samples; S: Number of variable sites; Eta: Total number of mutations; h: Number of Haplotypes; Hd: Haplotype (gene) diversity; Pi: Nucleotide diversity (per site); k: Average number of nucleotide differences; Rm: Minimum number of recombination events; SD: Standard deviation.

Table 3: Neutrality tests results of lotus populations.

Population	Fu's	Tajima's		Fu and Li's				Strobeck's S
	Fs	*D*	*P*	*D**	*P*	*F**	*P*	
Lotus white	1.545	-0.32197	Not significant, $p > 0.10$	-0.76384	Not significant, $p > 0.10$	-0.73647	Not significant, $p > 0.10$	0.446
Lotus pink	14.465	0.99305	Not significant, $p > 0.10$	1.13376	Not significant, $p > 0.10$	1.27662	Not significant, $p > 0.10$	0.000
Total	13.184	0.16270	Not significant, $p > 0.10$	1.22398	Not significant, $p > 0.10$	1.02310	Not significant, $p > 0.10$	0.000

Note - D, Tajima's statistic; Fs, Fu's statistic; D* and F*, Fu and Li's.

differences (k) was 156.364 $\times 10^{-2}$ (white lotus); 2908.225 $\times 10^{-2}$ (pink lotus) and 2210.985 $\times 10^{-2}$ (total of both lotus population). The nucleotide diversity coefficient (Pi) accounts for 0.180±0.050 $\times 10^{-3}$ (white lotus), 3.280 ± 0.830 $\times 10^{-3}$ (pink lotus) and 2.550±0.720 $\times 10^{-3}$ (total of both lotus population), respectively. The minimum number of recombination events (Rm =1) for the pink lotus population but not expressed in the white lotus population. All the indicators were processed with statistical significance $p<0.05$ (Table 2).

The results test neutrality shown in Table 3 with D, Li's D* and F* values are negative for the white lotus population (D= -0.32197, D*= -0.76384 and F*= -0.73647, Not significant, $p>0.10$), shows an excess of rare nucleotide site variants and recently emerging haplotypes compared to what would be expected under a neutral model and either population expansion or background selection has occurred concerning for to the evolution of the lotus population under study (Fu and Li, 1993). While the D value of the pink lotus population and the combination of the total of both lotus populations yielded a positive value was with not significant $p > 0.10$, this indicates the evolution of the pink lotus population is studied and both total lotus populations may have been suffered a recent bottleneck or we may have evidence for overdominant selection at this locus. Besides, the values of Fu and Li's D* and F* of the pink lotus population and the combination of the white and pink lotus populations (*Not significant:* $p > 0.10$), indicated that the study population has very few individuals show large differences in when compared to other individuals in the population (Table 3). Also, the results of the Fu's Fs test, based on the distribution of haplotypes, showed positive values for the lotus populations as evidence for a deficiency of alleles, as would be expected from a recent population bottleneck or we may have evidence for overdominant selection at this population.

Following Fu's Fs test, the hypothesis of natural evolution was significantly rejected for all regions of the lotus population under study. Strobeck's S, the probability of obtaining equal or fewer haplotypes based on gene frequency and mutation rate is low in the white lotus population (Strobeck's S statistic = 0.446) and does not occur in the pink lotus population and total of both lotus populations (Table 3). These results are not consistent with deviation from neutrality due to either selection or population expansion.

Table 4: Fst value and gene ow among of lotus populations.

Population	Lotus white	Lotus pink
Lotus white	-	0.00032
Lotus pink	0.00233*	-

Data allow the diagonal is Fst; Nm values are data above the diagonal. * indicates the significance level of Fst value at p<0.05.

Researchers often use Fst to assess gene flow, a higher Fst value indicates a lower level of gene flow (Nm) and higher genetic differentiation among populations (Hedrick, 2005). Fst reflects the level of inbreeding within populations (Wright, 1984) or the extent to which populations are differentiated (Hartl and Clark, 2007). The presence of genetic structure is an outcome of limited gene flow and a high level of genetic drift within each reproductively isolated group. Fst values below 0.05 indicate negligible genetic differentiation, whereas values greater than 0.25 indicate high genetic differentiation within the analyzed population (Weir, 1996). Fst values of the lotus populations in Thua Thien Hue, Viet Nam were significant but weak (Fst = 0.00233 ; $p < 0.05$, Table 4). This may be because primitive and highly conservative of the chloroplast genome in this species. This genetic distance implies that the kinship between unrelated individuals of the same ancestry relative to the lotus study population is equivalent to the kinship between individuals in a randomly mating lotus population. The presence of genetic structure is an outcome of limited gene ow and a high level of genetic drift within each reproductively isolated group. This study indicated that the two population genetic structure of lotus collection from different places in Thua Thien Hue, Vietnam are stable in both spatial and temporal scales. The random dispersal and their year-round spawning behavior might inhibit the population differentiation.

Phylogenetic analysis

The Phylogenetic tree was built based on Maximum Likelihood method (bootstrap = 1000). The result is shown the first cluster groups a population of pink lotus (SH02, SH05, SH06, SH07 and SH09); The second cluster includes both the pink and white lotus samples and they are divided into 3 different sub-clusters. Sub-clusters one includes 7 pink lotus samples (SH04, SH10, SH11, SH12, SH20, SH21 and SH22), Sub-clusters two includes 11 white

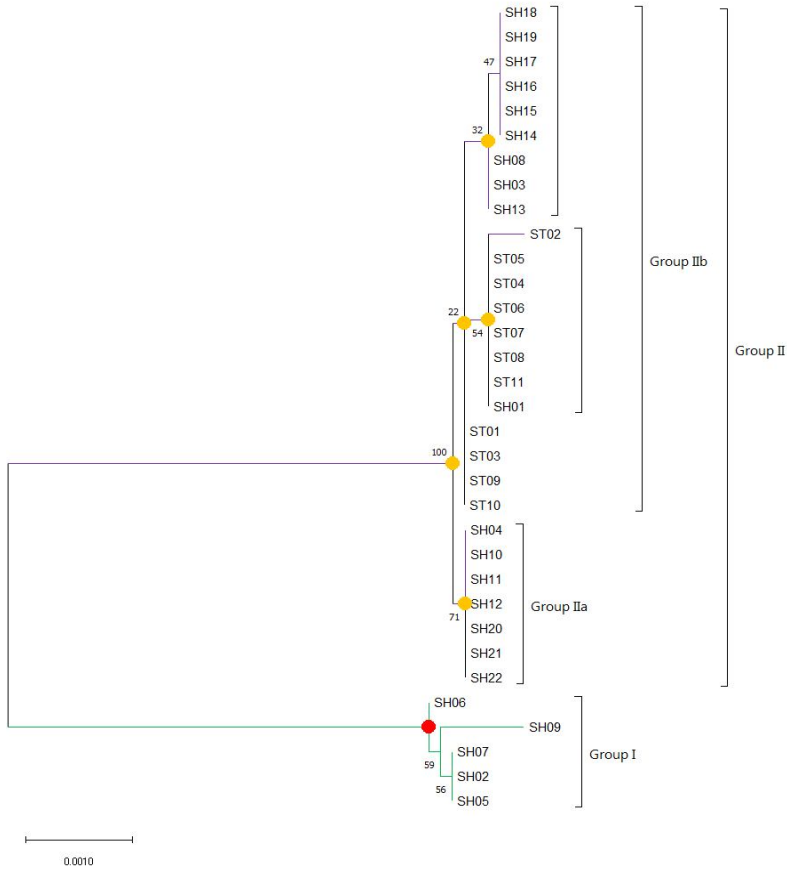

Fig 2: Evolutionary relationships of taxa based on the nine markers of the barcodes region by Maximum Likelihood method.

lotus (ST01-ST11) and 1 pink lotus (SH01) samples and Sub-clusters three includes the rest of the pink lotus samples (SH03, SH08, SH13-19). The tree topology is supported by a good bootstrap value. The differences between the two pink and white lotus populations were not found in the total nine regions of barcodes. Although, the two lotus populations have a different flower color, shared the same haplotype for the nine markers of the barcodes region, which are considered the most variable coding and non-coding regions of the plastid genome (Chase *et al.* 2007) (Fig 2).

CONCLUSION

Although the chloroplast genome contains many noncoding regions, relatively few have been exploited for interspecific phylogenetic and intraspecific phylogeographic studies. In our recent evaluation of the phylogenetic utility of non-coding chloroplast regions, we found the most widely used non-coding regions are among the least variable, but the more variable regions have rarely been employed (Long *et al*. 2020). To explore the potential variability of previously unexplored regions, we used nine gene regions in the study of thirty-three lotus samples that were collected in Thua Thien Hue province. The result 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 variety of *N. nucifera* species barcode. Analyses involving this sequence showed only 0.562% polymorphism in the studied taxa. Our results also indicate the need to use a different region, e.g., the *ITS* region, trnH-psbA, *rbcL, matK, rpoC1* and *rpoB*, to correctly identify white and pink lotus varieties of *N. nucifera* species. All of the new sequences have been deposited in GeneBank.

ACKNOWLEDGEMENT

Dang Thanh Long was funded by Vingroup Joint Stock Company and supported by the Domestic Master/ PhD Scholarship Programme of Vingroup Innovation Foundation (VINIF), Vingroup Big Data Institute (VINBIGDATA), code No: VINIF.2020.TS.119.

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.

Avise, J.C., Arnold, J., Ball, R.M., Bermingham, E., Lamb, T., Neigel, J.E., Reeb, C.A. and Saunders N.C. (1987a). Intraspecific phylogeography: The mitochondrial DNA bridge between population genetics and systematics. Annual Review of Ecology, Evolution and Systematics. 18: 489-522.

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(2): 295-299.

Dong, W., Liu, J., Yu, J., Wang, L., Zhou, S. (2012). Highly variable chloroplast markers for evaluating plant phylogeny at low taxonomic levels and for DNA barcoding. PLoS one. 7(4): e35071.

Fu, Y.X. (1997). Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. Genetics. 147: 915-925.

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

Hartl, D.L., Clark, A.G. (2007). Principles of Population Genetics, 4th edn. Sinauer Associates, Sunderland.

Hedrick, P.W. (2005). A standardized genetic differentiation measure. Evolution. 59(8): 1633-1638.

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., Li, M., Knyaz, C. and Tamura, K. (2018). MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. Molecular Biology and Evolution. 35: 1547-1549.

Long, D.T., Hong, H.T.K., Tram, L.L.T., Trang, N.T.Q. (2020). Evalution of genetic diversity by dna barcoding of local lotus populations from thua thien hue province. Indian Journal of Agricultural Research. 55(2):121-128.

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.

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 Evolutio. 34(12): 3299-3302.

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.

Stroreck, C. (1987). Average number of nucleotide difference in a sample from a single subpopulation: A test for population subdivision. Genetics. 117: 149-153.

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

Tamura, K. and Nei, M. (1993). Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Molecular Biology and Evolution. 10: 512-526.

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.

Weir, B.S. (1996). Genetic Data Analysis II: Methods for Discrete Population Genetic Data. Sinauer Associates, Inc., Sunderland.

Wright, S. (1984). Evolution and the Genetics of Populations, vol 4: Variability Within and Among Natural Populations. University of Chicago Press.