BIODIVERSITAS
Volume 25, Number 8, August 2024
Pages: 3299-3308

ISSN: 1412-033X
E-ISSN: 2085-4722
DOI: 10.13057/biodiv/d250801

Application of DNA barcode for the genetic analysis and identifying a *May nuoc mo* species in Quang Nam Province, Central Vietnam

HUYNH KIM HIEU[1], NGUYEN VAN MINH[2], NGUYEN VAN LOI[2], HO THANH HA[2,♥]

[1]Institute of Research and Development, University of Agriculture and Forestry, Hue University. 102 Phung Hung, Hue City, Thua Thien Hue Province, Vietnam

[2]Faculty of Forestry, University of Agriculture and Forestry, Hue University. 102 Phung Hung, Hue City, Thua Thien Hue Province, Vietnam.
Tel.: +84-234-3539531, ♥email: hothanhha@hueuni.edu.vn, hkhieu.dhnl22@hueuni.edu.vn

Manuscript received: 4 April 2024. Revision accepted: 1 August 2024.

Abstract. *Hieu HK, Minh NV, Loi NV, Ha HT. 2024. Application of DNA barcode for the genetic analysis and identifying a* May nuoc mo *species in Quang Nam Province, Central Vietnam. Biodiversitas 25: 3299-3308.* Precise identification in *Calamus*, the genus in the subfamily Calamoideae (Arecaceae), is challenging due to taxonomic complexities. This study aimed to discriminate and identify the 23 *May nuoc mo* samples collected from Quang Nam, Vietnam using four barcoding regions (*rbcL*, *matK*, *psbA-trnH*, and ITS nrDNA) and combined with morphological characters. DNA sequence data from those regions were analyzed using distance, tree, and similarity-based statistical methods. All region sequences presented no variable sites, except for *rbcL* gene region with a low rate of nucleotide differences of 0.270%. The results of the 23 *May nuoc mo* samples were highly similar from 99.73 to 100% to *Calamus* sp. N_XT142 (ON248649.1 and MK692404.1) and no gene regions were proposed as barcodes for species identification in this study. All samples shared certain morphological characteristics, including leaflets arranged regularly but with gaps; petioles with yellow spine groups; and leaf sheaths with hairs, densely arranged with brown, triangular, and flattened spines, and interspersed among many short, and black spines. Based on these results, we recommend using some other gene regions (*rpoC*, and *rpoB*)) in species identification to shed more light on these *May nuoc mo* samples with subsequent studies.

Keywords: *Calamus*, DNA barcoding, *May nuoc mo*, morphological characteristics, Quang Nam

INTRODUCTION

May nuoc mo (local name according to Peter and Henderson 2014; Dung et al. 2021) is one of the rattan species belonging to the subfamily Calamoideae (Arecaceae), a large subfamily of palms. It is an essential component offering both conventional local commodities and ecosystem-related benefits. This *May nuoc mo* species needs to be managed and conserved. In Vietnam, the *May nuoc mo* is distributed in Central Vietnam, including Quang Nam Province. According to many studies based on morphological characteristics, it has been shown that the *May nuoc mo* species belongs to the genus *Calamus* with different species names depending on each author report, including *C. eugenei* W.J.Baker (Baker 2015) and *C. applanata* (A.J.Hend. & N.Q.Dung) A.J.Hend. (Henderson 2020).

The study of biodiversity using different molecular techniques to identify species has drawn more attention lately. Efficient delineation and identification of a species is essential for preserving biodiversity and is a crucial factor in enhancing species conservation and management (Trias-Blasi and Vorontsova 2015). DNA barcoding techniques, in particular, offer the best capacity to classify species among groups (Ojeda et al. 2014), resolve issues with classifying distinct species (Meher et al. 2016; Antil et al. 2023), while also uncovering new species (Lopez-Vaamonde et al. 2021), and construct phylogenetic trees that facilitate community ecology research (Kress et al. 2015).

DNA barcoding may be a particularly valuable tool for confirming the identification of palm species, especially for specimens at immature stages of development, where diagnostic floral characteristics are rarely present in many botanical garden collections (Le et al. 2020). In plant classification studies at the species level, have shown that within the same species, the ITS nrDNA region has also been shown to have low variability within species (Alam et al. 2020; Ahmadi et al. 2022). With differing degrees of success in species identification depending on the taxonomic group under study, several plastid genome regions, such as the *trnH-psbA*, *rbcL*, *matK*, *rpoC1*, and *rpoB* intergenic spacer area and ITS nrDNA have been proposed, and tested for DNA barcoding of terrestrial plants (Long et al. 2021a; Antil et al. 2023).

Taxonomic complexities, like environmental plasticity and homoplasy, make precise identification challenging in *Calamus*, the genus of spiny climbing palms of the subfamily Calamoideae (Arecaceae) (Kurian et al. 2020). Although the rate of species discrimination within the same genus in palm based on DNA barcodes between different studies is different, several studies have also been carried out to solve this problem and determine the relationship between species within a group (Jeanson et al. 2011; Yang et al. 2012; Naeem et al. 2014). According to research by Yang et al. (2012), using the *matK*, *rbcL*, and *trnHpsbA* gene regions on 15 *Calamus* species in China, The result shows that the *trnHpsbA* region is proposed as a DNA barcode for this species. Phylogenetic studies of palm trees

based on several gene regions of the chloroplast genome have also been studied, including *matK*, *rbcL*, *rps16*, and *trnL-trnF*, and the results demonstrate variation in low genetic variation (Asmussen and Chase 2001; Asmussen et al. 2006; Le et al. 2020).

Given the discrepancy among previous studies, the present study aimed to test the usefulness of DNA barcoding in species identification of the *May nuoc mo* samples collected from different sub-areas in two closely related districts of Nam Giang and Dong Giang, Quang Nam province, Vietnam. We used the region ITS nrDNA, *trnH-psbA*, *rbcL*, and *matK* for the 23 *May nuoc mo* samples. This will serve as the data source for our in-depth examination of the molecular phylogeny of rattan species belonging to any previously published genera based on morphological characteristics.

MATERIALS AND METHODS

Plant materials

A total of three samples from the *May nuoc mo* species, comprising stems, leaf sheaths, ocreas, knees, rachis, cirri, inflorescences, and fruits were collected during the field survey in Dong Giang and Nam Giang districts, Quang Nam province, Vietnam. The samples were labelled, dried, and stored in the plant herbarium of Forestry Faculty, University of Agriculture and Forestry, Hue University, Vietnam (Table 1; Figure 1).

The 23 leaf samples of the *May nuoc mo* were chosen and collected from 23 different sites across various compartments of Dong Giang and Nam Giang districts, Quang Nam province. The leaf samples were rinsed and cleaned with distilled water and then stored in a dark refrigerator at 4°C to be used for further research (Table 1; Figure 1).

Procedures

Morphological description

The morphological description was conducted based on three herbarium collections and photographs of living plants (code: HMNGV01; HMNGV12 and HMNGV22). Identification of the *May nuoc mo* species was taken by morphological comparison from Ho (1999), Peter and Henderson (2014). The terminology used for the morphological description of the *May nuoc mo* species follows Thin (2007).

Genomic DNA extraction

DNA from the leaf samples of the *May nuoc mo* were extracted and purified using the Cetyl Trimethyl Ammonium Bromide (CTAB) method described by Vaze et al. (2010). The method relies on CTAB forming complexes with sample DNA. Subsequently, the phenol: chloroform method was used to purify the total DNA. The quality of the total DNA was tested based on electrophoresis with a 1% agarose gel in 1X TAE buffer and staining with GelRed dye (Biotium, USA). It was then observed under UV light using a direct UV reading system (UV-transilluminator, Model: DyNa Light). The total DNA was stored at -20°C for further use.

Table 1. List of samples used in this study

Code sample	Sampling location	Altitude (m asl.)	Coordinate (Latitude; Longitude)	Accession number from GenBank (*rbcL; matK; trnH-psbA; ITS*)
HMNGV01	Sub-region 223, Nam Giang	335	15°75'; 107°68'	PP104953; PP104978; PP109235; PP117900
HMNGV02	Sub-region 223, Nam Giang	227	15°77'; 107°69'	PP104954; PP104979; PP109236; PP117901
HMNGV03	Sub-region 223, Nam Giang	66	15°78'; 107°69'	PP104955; PP104980; PP109237; PP117902
HMNGV04	Sub-region 294, Nam Giang	445	15°68'; 107°66'	PP104956; PP104981; PP109238; PP117903
HMNGV05	Sub-region 294, Nam Giang	444	15°68'; 107°66'	PP104957; PP104982; PP109239; PP117904
HMNGV06	Sub-region 294, Nam Giang	453	15°68'; 107°66'	PP104958; PP104983; PP109240; PP117905
HMNGV07	Sub-region 286, Nam Giang	225	15°69'; 107°60'	PP104959; PP104984; PP109241; PP117906
HMNGV08	Sub-region 286, Nam Giang	218	15°69'; 107°60'	PP104960; PP104985; PP109242; PP117907
HMNGV09	Sub-region 286, Nam Giang	226	15°69'; 107°60'	PP104961; PP104986; PP109243; PP117908
HMNGV10	Sub-region 299, Nam Giang	238	15°64'; 107°62'	PP104962; PP104987; PP109244; PP117909
HMNGV11	Sub-region 299, Nam Giang	240	15°64'; 107°62'	PP104963; PP104988; PP109245; PP117910
HMNGV12	Sub-region 299, Nam Giang	237	15°64'; 107°62'	PP104964; PP104989; PP109246; PP117911
HMNGV13	Sub-region 154, Dong Giang	401	15°83'; 107°64'	PP104965; PP104990; PP109247; PP117912
HMNGV14	Sub-region 154, Dong Giang	401	15°83'; 107°64'	PP104966; PP104991; PP109248; PP117913
HMNGV15	Sub-region 154, Dong Giang	442	15°83'; 107°64'	PP104967; PP104992; PP109249; PP117914
HMNGV16	Sub-region 159, Dong Giang	170	15°83'; 107°74'	PP104968; PP104993; PP109250; PP117915
HMNGV17	Sub-region 159, Dong Giang	280	15°83'; 107°73'	PP104969; PP104994; PP109251; PP117916
HMNGV18	Sub-region 166, Dong Giang	375	15°80'; 107°72'	PP104970; PP104995; PP109252; PP117917
HMNGV19	Sub-region 166, Dong Giang	283	15°80'; 107°70'	PP104971; PP104996; PP109253; PP117918
HMNGV20	Sub-region 166, Dong Giang	228	15°79'; 107°71'	PP104972; PP104997; PP109254; PP117919
HMNGV21	Sub-region 158, Dong Giang	419	15°83'; 107°70'	PP104973; PP104998; PP109255; PP117920
HMNGV22	Sub-region 158, Dong Giang	420	15°82'; 107°70'	PP104974; PP104999; PP109256; PP117921
HMNGV23	Sub-region 158, Dong Giang	634	15°83'; 107°69'	PP104975; PP105000; PP109257; PP117922

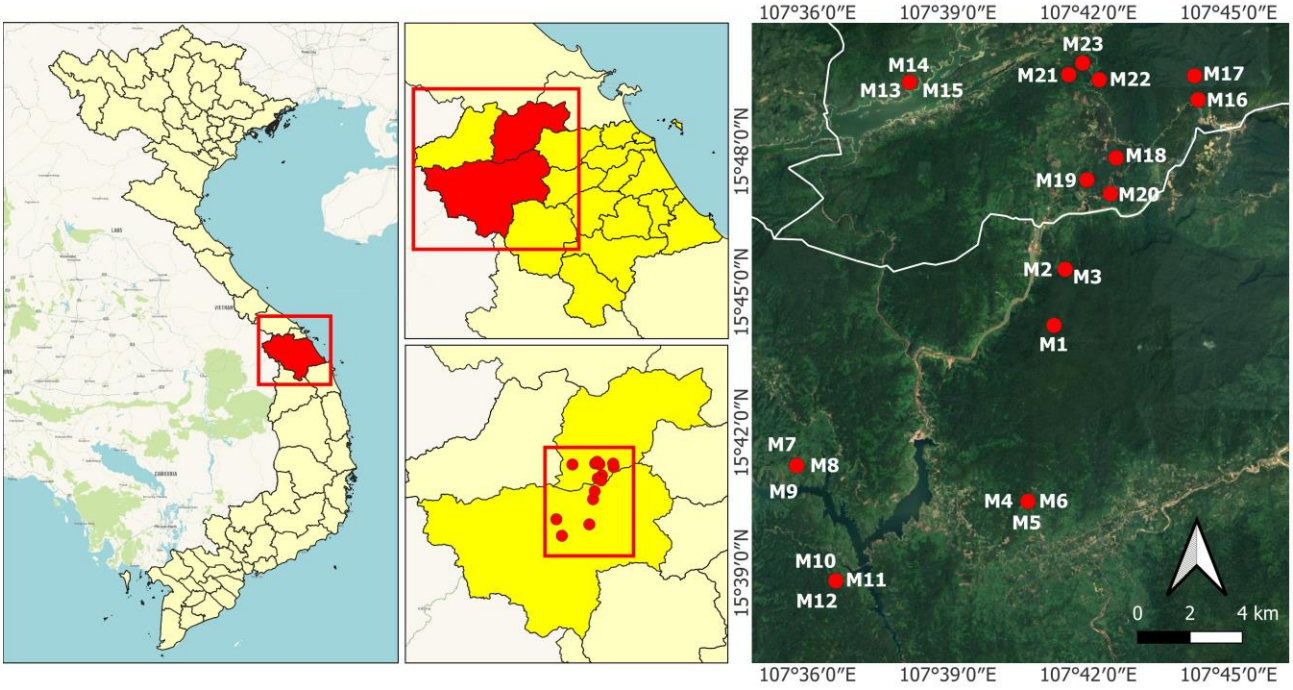

Figure 1. Sample collection locations in Quang Nam Province, Central Vietnam. Note: M: HMNGV on Table 1

Table 2. Nucleotide sequences of primers and thermocycling for PCR reaction

Regions	Primer	Sequence 5′→3′	References
rbcL	1F	ATGTCACCACAAACAGAGAC	Long et al. (2021b)
	743R	TCACATGTACCTGCAGTAGC	
matK	385F	CGATCAATTCATTCAATATTTTC	Dong et al. (2012)
	1320R	ACTTCGACTTTCGTGTGCTAGA	
trnH-psbA	46F	ACTGCCTTGATCCACTTGGC	
	25R	TGAAGCTCCATCTACAAATGG	
ITS	ITSF	AATTGCAGAATCCCGTGAACC	MK692394.1
	ITSR	TACAATTCGAGCGGCGCAGCCG	

DNA amplification

Amplification of the chloroplast genome's ribulose-1,5-bisphosphate carboxylase (*rbcL*), Matturase K (*matK*), and *trnH-psbA* gene regions, as well as the ITS nrDNA (Internal Transcribed Spacer) region of the nuclear genome were conducted using primer pairs designed by Dong et al. (2012), Long et al. (2021b), and based on Genbank code MK692394.1 (Table 1). PCR reactions were performed on a SimpliAmp™ Thermal Cycler (ThermoFisher, USA) in a volume of 50 µL containing 25 µL GoTaq® Green Master Mix, 2X (2.4 mM dNTP each, 0.3 Taq units DNA polymerase, Promega), 10 pmol forward primer (IDT, USA), 10 pmol reverse primer (IDT, USA), 100 ng total DNA template and sterile deionized water. The thermal cycle used for the PCR reaction to amplify gene regions consists of 95°C for 5 min (predenaturation), then 30 cycles of 95°C for 45 seconds (denaturation) followed by 55°C for 45 seconds (annealing) and 72°C for 1 min (extension), and one cycle at 72°C for 10 min (final extension). PCR reaction products were electrophoresed on a 1% agarose gel in 1X TAE buffer with Gelred dye (Biotium, USA) and analyzed using a UV-transilluminator (Model: DyNa Light) and a 100-1000 bp DNA standard scale (Bioline, UK) for size determination.

DNA sequencing

PCR amplification products of the target gene regions were purified using the Isolate II PCR and Gel kit (Bioline, UK). These products were then subjected to direct gene sequencing reactions with specific primer pairs using the fluorescent labeling method (dideoxy terminator) on the ABI 3100 system (Applied Biosystems, USA) at the 1st Base company-Malaysia (https://base-asia.com/).

Data analysis

DNA sequence data were analyzed and modified using BioEdit 7.0.5 software. The DNA sequences are aligned by the Clustal program and The species name was ascertained by comparing them to the NCBI (www.ncbi.nlm.nih.gov) by the BLAST tool. Based on MEGA X software (The Molecular Evolution Genetics Analysis), the Maximum Likelihood (ML) method was used for inferring evolutionary trees between research samples. The analysis included a bootstrap value of 2000 replicates to consolidate the branch positions in the phylogenetic tree (Kumar et al. 2016). The Tamura-Nei model (1993) was used for the phylogenetic tree construction. The Maximum Composite Likelihood (MCL) approach was used to estimate the initial tree, which was then automatically constructed using the Neighbor-Join and BioNJ algorithms based on the pairwise distance matrix. The topology with the highest stability value was selected. The number of substitutions at each nucleotide site is used to calculate branch length, and the tree was drawn to scale.

Measures of DNA polymorphism in the population analysis are based on six parameters, i.e. the number of polymorphic sites (S), the total number of mutations (Eta), the number of haplotypes (h), haplotype diversity (Hd), the average number of different nucleotides (k), and nucleotide diversity (π) (Rozas et al. 2017). Five methods are used to test for neutrality: Tajima's D test (Tajima 1989), Fu and Li's D* and F* test (Fu and Li 1993), Fu's statistic (Fu 1997), and Strobeck's statistic utilizing DNAsp 6.0 software (Rozas et al. 2017).

RESULTS AND DISCUSSION

Morphological description

Based on three specimens collected in the wild HMNGV01, HMNGV12, and HMNGV20, morphological characteristics of the *May nuoc mo* species were described as follows: Stems are clustered, climbing, with mature stems up to 20 m long, nodes ranging from 18 to 31 cm in length and up to 2.9 cm in diameter. Leaf sheaths are yellowish-green or brown with reddish-brown hairs, densely arranged with brown, triangular, flattened, spines 1.5 to 2.4 cm in length, interspersed amongst many, short, black spines. Petioles are 12 to 27 cm in length, rachis is 1.2-1.8 m long with 35-45 lanceolate leaflets on each side, regularly arranged to 2.5-2.7 cm or sometimes irregularly arranged in groups from 6,5 to10 cm; leaflets are 31-48 x 2 -2.5 cm with 3 main veins; bristly on margins and lateral veins on upper surfaces and main veins on lower surfaces. Ocreas are obscure, knees on leaf sheaths usually present; cirrus is 1.2-1.4 m long. Normally, the leaf of a young seedling is brown-red and has regularly arranged leaflets compared to the mature stem, with petioles having yellow spines groups of 3-5 spines, and the distance between clusters of spines in the leaf axis is from 9 to 12.5 cm. Inflorescences are surrounded by bracts that are 30 to 85 cm long, arching; bracts falling from the elongating infructescence, only the basal bract persistent. Fruits are globose, 10 x 11 mm in diameter, sharp-pointed, and when ripe, the color of the fruit bark changes from green to yellowish-brown. The fruit bark is plump with black seeds, seed bark is fragrant and sweet (Figure 2).

PCR product

The results from revealed that all PCR products of gene regions in the examined the *May nuoc mo* samples showed a single DNA band with an amplification rate of 100%. DNA concentrations were all high and clear in the PCR results of the study samples. The DNA bands obtained size showed to correspond to the initial anticipated size (Table 3; Figure 3).

Sequence characteristics

The sequencing results showed that the partial sequence sizes of gene regions after correction are 740 bp of *rbcL*, 924 bp of *matK*, 678 bp of *trnH-psbA*, and 467 bp of ITS nrDNA region (Table 3). BLAST results on NCBI were used to verify and compare the nucleotide sequences of the gene regions in the chloroplast genome of the studied *May nuoc mo* samples showed high similarity to *Calamus* sp. N_XT142 (ON248649.1) (Yao et al. 2023), with similarity ranging from 99.73 to 100% and gene region coverage reaching 100% (Table 3). Meanwhile, for the *ITS* region, it showed that the 23 *May nuoc mo* samples used in the study have high similarities with *Calamus* sp., but the nucleotide sequence of the gene region has many nucleotide positions that are different from the sequence published on GenBank, corresponding to the highest similarity of 88.64% (MK692404.1). All obtained nucleotide sequences of gene regions were deposited on GenBank with reference (Table 1).

Figure 2. Morphological features of the *May nuoc mo*: A. Clustered stem; B. Leaf with regularly arranged leaflets but with gaps; C. Leaflet; D. Thorny petioles; E. Stem; F. Inflorescence; G. Infructescence; 1. Cirrus; 2. Linear leaflets; 3. Leaflets bristly on the veins; 4. Petioles with yellow thorny groups; 5. Ocreas obscure; 6. Knees on leaf sheaths usually present; 7. Leaf sheath spines conical-based; 8. Petioles; 9. Bracts persistent; 10. Inflorescences; 11. Spherical fruit

Table 3. Molecular characteristics sample studied

Regions	PCR success	Sequencing success	Total aligned length bp	Number of monomorphic sites	Variable sites %	Intraspecific distance mean
rbcL	100	100	740	738	0.270276	0-0.001355 (0.000236)
atK	100	100	924	924	0	0
trnH-psbA	100	100	678	678	0	0
ITS	100	100	467	467	0	0

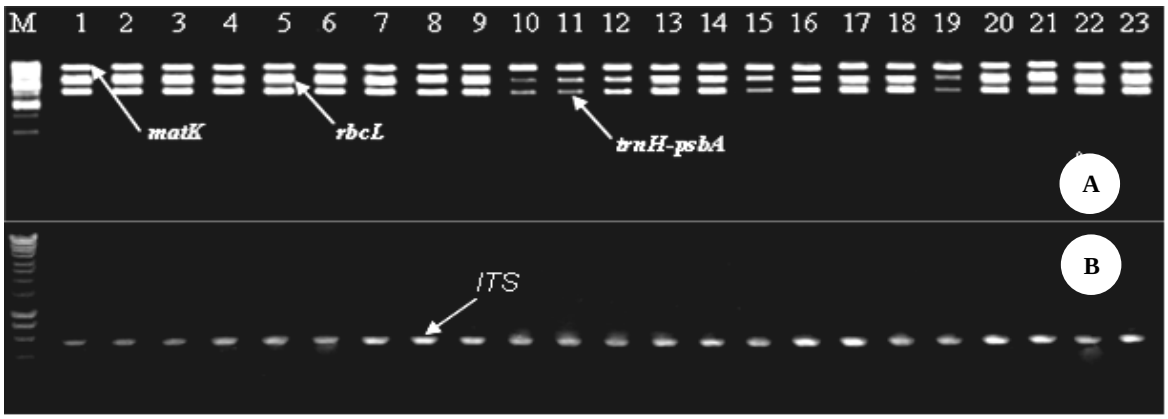

Figure 3. Electrophoresis of PCR products: M. Weight of DNA standard (100-1000 bp, Bioline; 1kb, Bioline); 1-23. Samples analyzed; A. Gene regions of the chloroplast genome; B. ITS nrDNA region

The percentage occurrence of each type of nucleotide in different gene regions varies. The *rbcL* gene region showed that all four types of nucleotides have the same occurrence rate in the gene region (25%). The *matK* gene region showed A = 29.65%, T/U = 37.12%, C = 17.97%, and G = 15.26%. The *trnH-psbA* gene region obtained the percentage of nucleotides as A = 38.49%, T/ U = 30.53%, C = 16.67%, and G = 14.31% and the *ITS* region obtained the corresponding percentage of nucleotide types A = 20.13%, T/U = 15.85%, C = 34.05%, and G = 29.98%. The nucleotide sequences of the two gene regions *matK, trnH-psbA*, and *ITS* region did not differ among the 23 *May nuoc mo* samples studied (Table 3). Meanwhile, the gene region contained two polymorphic nucleotide positions, accounting for 0.270276% of the total length of the gene region. The genetic distance between the 23 studied *May nuoc mo* samples ranged from 0-0.001355 (mean = 0.000236) (Table 3; Figure 4). These results showed that the three gene regions of the chloroplast and one gene region of the nuclear genome of the *May nuoc mo* plant used in the study have high genetic stability, and there are no differences between the *May nuoc mo* samples collected in the two districts of Dong Giang and Nam Giang, Quang Nam province.

The six parameters included the mean number of different nucleotides (k), total number of mutations (Eta), number of haplotypes (h), haplotype diversity (Hd), and number of polymorphic sites (S), and nucleotide diversity (π), which were used to evaluate the genetic diversity of the 23 investigated *May nuoc mo* samples of the genus *Calamus* in this study. Following editing and alignment, the sequence analysis outcomes showed that the 23 *May nuoc mo* samples that were studied had two different nucleotide positions in the *rbcL* gene region of 18 and 271 (Figure 4). These positions correspond to the average number of nucleotide differences in the gene region's total length (k) and nucleotide diversity coefficient (π), which are 0.174 and 0.240 $x10^{-3}$, respectively. With a matching haplotype diversity coefficient of 0.170, this divergence allowed for dividing the 23 *May nuoc mo* samples into three distinct haplotypes. The gene region sequences acquired from the 23 samples under study did not reveal nucleotide changes in other gene regions, such as *matK, trnH-psbA*, and ITS nrDNA region. In other words, the present *May nuoc mo* samples of *Calamus* exhibit a high degree of genetic diversity conservation for these gene regions. The ITS nrDNA region (0.640), followed by the *rbcL* gene region (0.426), and the *matK* gene region (0.332) have the largest G+C contents, while the *trnH-psbA* gene region has the lowest ratio (0.310). All indicators were analyzed with a statistical significance level of $p<0.05$ (Table 4; Table 5).

The number of polymorphic nucleotide sites relative to neutral predictions occurred at a low frequency among samples, according to the statistical analysis of D, D*, F*, and Fs for the *rbcL* gene region of the 23 investigated *May nuoc mo* samples of *Calamus*. Several rare nucleotide variations have recently been found in the sequence of the *rbcL* gene region, causing an increasing number of haplotypes. The number of populations has tended to increase due to the division of the number of the *May nuoc mo* samples under the influence of selection. The statistical analyses of D, D *, F *, and Fs (Table 5) yielded negative values (Doğan and Doğan 2017), which illustrate these observations.

Table 4. DNA diversity based on genomic regions of the *May nuoc mo* populations

Regions	G+C content	S	Eta	h	Hd	k	π (x10⁻³)
ITS nrDNA	0.640	0	0	1	0	0	0
rbcL	0.426	2	2	3	0.170	0.174	0.240
matK	0.332	0	0	1	0	0	0
trnH-psbA	0.310	0	0	1	0	0	0
rbcL+matK+trnH-psbA+ ITS nrDNA	0.427	2	2	3	0.170	0.174	0.070

Note: S: Number of polymorphic sites; Eta: Total number of mutations; h: Number of haplotypes; Hd: Haplotype diversity; k: Average number of different nucleotides; π: Nucleotide diversity

Table 5. Neutrality test results based on gene regions

Regions	Tajima's D test	Fu and Li's D* test	Fu and Li's F* test *	Fu's Fs statistic	Strobeck's S statistic
rbcL	-1,51496*	-2,13487**	-2,26023**	-2,027	0,981
matK	0	0	0	0	0
trnH-psbA	0	0	0	0	0
ITS	0	0	0	0	0

Note: *Not significant, p>0.10; **Not significant, 0.10>p>0.05

Figure 4. Comparison alignment of nucleotide sequences of the *rbcL* gene region

Phylogenetic relationships

Phylogenetic analysis was conducted on 23 nucleotide sequences of the *May nuoc mo* samples from *Calamus* in Dong Giang and Nam Giang districts, Quang Nam. The sequences were cleared of any blank spaces and missing information, resulting in a final sequence of 2342 nucleotide locations (total of *rbcL+matK+trnH-psbA* gene regions) and 467 nucleotide positions (for ITS nrDNA region). The total gene sequence of the *May nuoc mo* species of *Calamus* was obtained from GenBank and used as a reference (Figure 5 shows the codes).

The result demonstrates that the 23 investigated *May nuoc mo* samples in Dong Giang and Nam Giang districts, Quang Nam showed a strong genetic relationship and were on the same evolutionary branch (Figure 5). Two *May nuoc mo* samples (HMNGV06 and HMNGV11) have the fastest rates of evolution among them. Based on the sequences of the gene regions published on Genbank with reference number ON248649.1 (chloroplast region gene) and MK692404.1 (nuclear genomic region), there is a significant degree of similarity between all of these *May nuoc mo* samples and the genus *Calamus*. However, the species has not been definitively recognized. Based on these findings, we discovered that when looking at the *rbcL*, *matK*, *trnH-psbA*, and ITS nrDNA regions, there was substantial genetic stability among the examined rattan samples. Nucleotide sequences of gene regions obtained from samples are not sufficient to determine the species level of the studied *May nuoc mo* samples of *Calamus* (Figure 5).

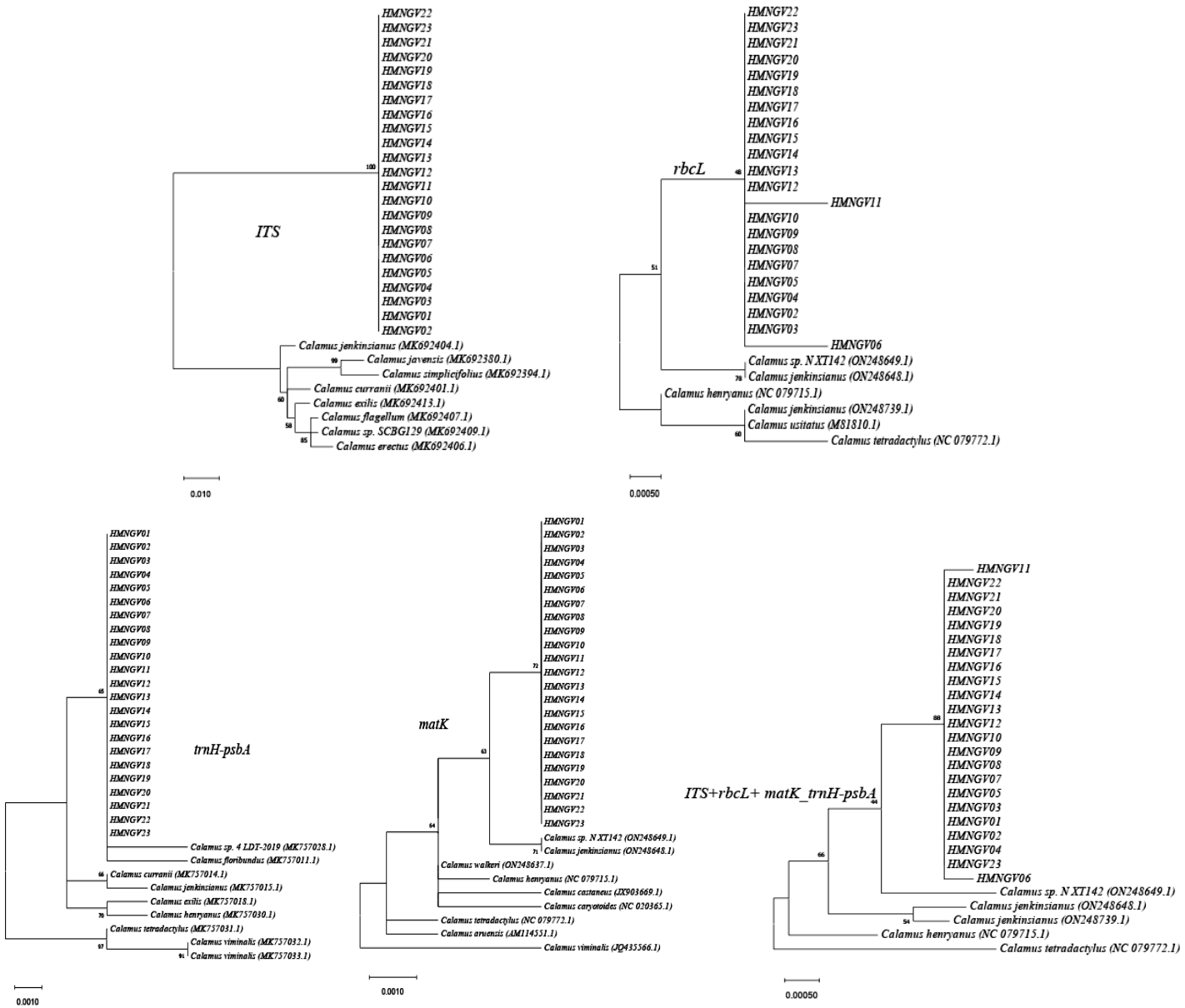

Figure 5. Phylogenetic tree of the 23 *May nuoc mo* samples based on ITS, *rbcL, matK, trnH-psbA,* and ITS+*rbcL*+*matK*+*trnH-psbA* regions

Discussion

Based on morphological characters, the *May nuoc mo* species can be distinguished from all species of the genus *Calamus* found in Vietnam by its leaf sheaths with hairs, densely arranged with brown, triangular, and flattened spines, and interspersed among many short, and black spines. This corresponds with the previous description of the species by Peter and Henderson (2014). Further, leaflets arranged regularly but with gaps, and petioles with yellow spine groups are also certain morphological characteristics to identify this species. Peter and Henderson (2014) suggested that newly opened leaves are often reddish-brown, which will usefully contribute to identifying this rattan species.

The following characteristics of the perfect DNA barcode, according to the CBOL Plant Working Group (2009) are significant genetic variation to allow successful PCR amplification using universal primers, high PCR amplification efficiency, and the ability to analyze DNA sequences. The conservative nature of DNA sequence analysis also makes it possible to distinguish between individuals belonging to the same species. This was made possible by the successful PCR amplification of all applicable genomic regions analyzed in this study, which made sequencing straightforward and of high quality. With a length of less than 1000 bp (about 930 bp), the PCR-amplified gene sections we acquired in our tests allow for highly efficient direct sequencing using specific primer pairs. Comparable outcomes were observed in other classes of terrestrial plants. The *trnH-psbA* region can be considered a barcode due to the PCR amplification of this gene region and the high quality of the sequencing data obtained (Bieniek et al. 2015; Su et al. 2016).

The present study involved the isolation and sequencing of three gene regions chloroplast and region ITS of nrDNA from the 23 *May nuoc mo* samples collected in Dong Giang and Nam Giang districts, Quang Nam province. The regions included the non-coding gene region *trnH-psbA*, which is situated between the *trnH* and *psbA* genes, as well as the two gene regions encoding *rbcL* and *matK* of the chloroplast genome and one region ITS of nrDNA. After correcting the respective gene regions, the sizes obtained were 740 bp for *rbcL*, 924 bp for *matK*, 678 bp for *trnH-psbA*, and 467 bp for region *ITS*. The results of the gene regions of the chloroplast genome show that all of these *May nuoc mo* samples have a high degree of similarity of 99.73% to 100% *Calamus* sp. (ON248649.1) and are located close to *C. jenkinsianus* with petioles of black thorny groups (ON248648.1) for the *rbcL* and *matK* gene regions and *C. exilis* (MK757018.1) for the *trnH-psbA* gene region. Meanwhile, the *ITS* region showed that the 23 *May nuoc mo* samples used in the study have high similarities with *Calamus* sp. Still, the obtained nucleotide sequences have large differences compared to the species published on GenBank, corresponding to the highest similarity of 88.64% (MK692404.1). The phylogenetic tree shows that all the examined rattan samples are on the same branch and have the same rate of evolution, except for two *May nuoc mo* samples (HMNGV06 and HMMGV11), which have a greater evolutionary rate than the other samples.

The use of three plastid genes (*matK, rbcL, trnH-psbA)* and ITS nrDNA did not provide strong enough distinctions between the *May nuoc mo* samples in the research population or determine the scientific names. This suggests that the nuclear and chloroplast genome of the rattan species studied has high genetic stability and does not undergo frequent genetic changes. The negative values obtained for the D, D *, F *, and Fs statistics indicate that uncommon nucleotide changes occur and nucleotide polymorphisms are less frequent than predicted suggesting that the populations have undergone selection (Tajima 1989; Fu and Li 1993; Fu 1997).

Nonetheless, several studies demonstrate that the *matK* gene region is a significant marker that aids in differentiating between members of the same or different species (Ho et al. 2021; Long et al. 2021a). However, due to issues with primer universality across species, inefficient amplification and sequencing, and other issues, numerous previous studies presented the value of this genomic region as a barcode (Theodoridis et al. 2012; Long et al. 2021b). Recently, in plant classification studies at the species level, the ITS nrDNA region is the most commonly decoded locus. The ITS nrDNA region is highly effective in taxonomic studies of many plants and fungi (except ferns), and this is a locus used for sequencing with short DNA (Ansari et al. 2018). For this gene region in research samples, the PCR amplification efficiency in our study was 100%. However, the results of the nucleotide sequence analysis were consistent across the *May nuoc mo* samples under investigation. This outcome was similar for gene regions of *trnH-psbA*.

The *rbcL* gene region, which has the nucleotide sequence, is thought to be the most conservative of the three gene regions examined in terms of molecular variation, as seen by the low number of polymorphic nucleotide locations. Numerous scientists from around the world have published and verified this observation, such as Bieniek et al. (2015), Bolson et al. (2015), Gamache and Sun (2015), and Le et al. (2020). Additional research indicates that the classification efficiency of this gene region can be applied to the genus *Hordeum* at the subspecies level (Bieniek et al. 2015; Gamache and Sun 2015) used the *rbcL* gene region to identify the scientific names of two species *H. bulosum* and *H. bogdani*.

Our study demonstrates that the combination of the three gene regions *trnH-psbA, matK, rbcL*, and *ITS* region, as well as their nucleotide sequences only permit the identification of the *May nuoc mo* samples collected in Dong Giang and Nam Giang districts, Quang Nam province that a highly similar to *Calamus* sp. While, according to many studies based on morphological characteristics, it has been shown that the *May nuoc mo* species belongs to the genus *Daemonorops* (now *Calamus*) with different species names depending on each author report, including *C. applanata* (Dung et al. 2021) and *C. poilanei* (Peter and Henderson 2014; Lợi et al. 2018). Using these gene regions, our research results are comparable to Long et al. (2021a, 2021b) for the genus *Panicum*. When utilizing the *matK* gene region to identify species in the genera *Elymus*, *Loptiopyrum*, *Pseudoroegneria*, and *Thinopyrum*, these results are in contradiction to Bieniek et al. (2015). For the two gene areas, *rbcL* and *matK*, Hunt et al. (2014)

demonstrated great effectiveness in identifying species of the genus *Panicum.*

In conclusion, based on similarities in the barcode sequence data of three gene regions *rbcL*, *matK*, and the non-coding spacer region *trnH-psbA* located between the two genes *trnH* and *psbA*, ITS nrDNA and morphological characters, this study reported the 23 *May nuoc mo* samples as belonging to the *Calamus* sp. N_XT142 (ON248649.1 and MK692404.1). Based on some morphological characteristics, all samples showed similarities, including leaflets arranged regularly but with gaps; petioles with yellow spine groups; and leaf sheaths with hairs, densely arranged with brown, triangular, and flattened spines, and interspersed among many short, and black spines. The gene regions therefore emerge as a strong candidate barcode for the genus *Calamus* of this rattan species. Therefore, our findings highlight the need to use many distinct gene areas in diverse genomes, such as the *rpoC1* and *rpoB*egion to accurately identify species of the *May nuoc mo* samples obtained in Dong Giang and Nam Giang districts, Quang Nam province.

ACKNOWLEDGEMENTS

This article is a part of the collaborative activities within WWF (World Wide Fund for Nature) and IKEA Partnership in the Greater Mekong region. In addition to financial support, this article has received guidance and technical support from WWF and IKEA colleagues. This research also benefited from the partial funding from Hue University under the Core Research Program, Grant No. NCTB.DHH.2024.10. We are grateful to the administration of the Nam Giang and Dong Giang watershed protection forest management boards for providing us with an opportunity to carry out the field surveys. We would like to express our sincere thanks to Dr. Dang Thanh Long (the Institute of Biotechnology, Hue University, Vietnam) for his technical assistance in the application of DNA barcoding.

REFERENCES

Ahmadi H, Solouki M, Fazeli-Nasab B, Heidari F, Sayyed RZ. 2022. Internal Transcribed Spacer (ITS) regions: A powerful tool for analysis of the diversity of wheat genotypes. Indian J Exp Biol 60: 137-143.

Alam A, Chadha NK, Kumar A-P, Chakraborty SK, Joshi KD, Sawant PB, Das SCS, 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 J Anim Res 54 (7): 856-863. DOI: 10.18805/ijar.B-3833.

Ansari S, Solouki M, Fakheri B, Fazeli-Nasab B, Mahdinezhad N. 2018. Assessment of molecular diversity of Internal transcribed spacer region in some lines and landrace of persian clover (*Trifolium resupinatum* L). Potravinarstvo Slovak J Food Sci 12 (1): 657-666. DOI: 10.5219/960.

Antil S, Abraham JS, Sripoorna S, Maurya S, Dagar J, Makhija S, Bhagat P, Gupta R, Sood U, Lal R, Toteja R. 2023. DNA barcoding, an effective tool for species identification: A review. Mol Biol Rep 50 (1): 761-775. DOI: 10.1007/s11033-022-08015-7.

Asmussen CB, Chase MW. 2001. Coding and noncoding plastid DNA in palm systematics. Am J Bot 88 (6): 1103-1117. DOI: 10.2307/2657094.

Asmussen CB, Dransfield J, Deickmann V, Barfod AS, Pintaud J-C, Baker WJ. 2006. A new subfamily classification of the palm family (Arecaceae): Evidence from plastid DNA phylogeny. Bot J Linn Soc 151 (1): 15-38. DOI: 10.1111/j.1095-8339.2006.00521.x.

Baker WJ. 2015. A revised delimitation of the rattan genus *Calamus* (Arecaceae). Phytotaxa 197: 139-152. DOI: 10.11646/phytotaxa.197.2.7.

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 Syst Evol 301: 1275-1286. DOI: 10.1007/s00606-014-1138-1.

Bolson M, de Camargo Smidt E, Brotto ML, 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. DOI: 10.1371/journal.pone.0143049.

CBOL Plant Working Group. 2009. A DNA barcode for land plants. Proc Natl Acad Sci USA 106: 12794-12797. DOI: 10.1073/pnas.0905845106.

Doğan I, Doğan N. 2017. Statistical tests for neutrality: Review. Turkiye Klinikleri J Biostat 9 (2): 167-174. DOI: 10.5336/biostatic.2016-53446.

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. DOI: 10.1371/journal.pone.0035071.

Dung NQ, Hai TN, Henderson A, Phuong NTB. 2021. Proposal on conservation and development of high value rattan species in Vietnam. J For Sci Technol 5: 67-77.

Fu YX, Li WH. 1993. Statistical tests of neutrality of mutations. Genetics 133 (3): 693-709. DOI: 10.1093/genetics/133.3.693.

Fu YX. 1997. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. Genetics 147 (2): 915-925. DOI: 10.1093/genetics/147.2.915.

Gamache J, Sun G. 2015. Phylogenetic analysis of the genus *Pseudoroegneria* and the Triticeae tribe using the *rbcL* gene. Biochem Syst Ecol 62: 73-81. DOI: 10.1016/j.bse.2015.07.038.

Henderson A. 2020. A revision of *Calamus* (Arecaceae, Calamoideae, Calameae, Calaminae). Phytotaxa 445 (1): 1-656. DOI: 10.11646/phytotaxa.445.1.1.

Ho P. 1999. An Illustrated Flora of Vietnam. Part 3. Young Publishing House, Ho Chi Minh City.

Ho VT, Tran TKP, Vu TTT, Widiarsih S. 2021. Comparison of *matK* and *rbcL* DNA barcodes for genetic classification of jewel orchid accessions in Vietnam. J Genet Eng Biotechnol 19 (1): 93. DOI: 10.1186/s43141-021-00188-1.

Hunt HV, Badakshi F, Romanova O, Howe CJ, Jones MK, Heslop-Harrison JSP. 2014. Reticulate evolution in *Panicum* (Poaceae): The origin of tetraploid broomcorn millet, *P. miliaceum*. J Exp Bot 65 (12): 3165-3175. DOI: 10.1093/jxb/eru161.

Jeanson ML, Labat JN, Little DP. 2011. DNA barcoding: A new tool for palm taxonomists? Ann Bot 108: 1445-1451. DOI: 10.1093/aob/mcr158.

Kress WJ, García-Robledo C, Uriarte M, Erickson DL. 2015. DNA barcodes for ecology, evolution, and conservation. Trends Ecol Evol 30 (1): 25-35. DOI: 10.1016/j.tree.2014.10.008.

Kumar S, Stecher G, Tamura K. 2016. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. Mol Biol Evol 33 (7): 1870-1874. DOI: 10.1093/molbev/msw054.

Kurian A, Dev SA, Sreekumar VB, Muralidharan EM. 2020. The low copy nuclear region, RPB2 as a novel DNA barcode region for species identification in the rattan genus *Calamus* (Arecaceae). Physiol Mol Biol Plants 26 (9): 1875-1887. DOI: 10.1007/s12298-020-00864-5.

Le D-T, Zhang Y-Q, Xu Y, Guo L-X, Ruan Z-P, Burgess KS, Ge X-J. 2020. The utility of DNA barcodes to confirm the identification of palm collections in botanical gardens. PLoS One 15 (7): e0235569. DOI: 10.1371/journal.pone.0235569.

Long DT, Hong HTK, Tram LLT, Trang NTQ. 2021a. Research on phylogenetic relationship of *Lotus* populations collected in Thua Thien Hue Province, Vietnam based on the chloroplast genome by DNA Barcode. Indian J Agric Res 56: 249-254. DOI: 10.18805/IJARe.A-646.

Long DT, Hong HTK, Tram LLT, Trang NTQ. 2021b. Evaluation of genetic diversity by DNA barcoding of local *Lotus* populations from Thua Thien Hue Province. Indian J Agric Res 55 (2): 121-128. DOI: 10.18805/IJARe.A-564.

Lopez-Vaamonde C, Kirichenko N, Cama A et al. 2021. Evaluating DNA barcoding for species identification and discovery in European Gracillariid moths. Front Ecol Evol 9: 626752. DOI: 10.3389/fevo.2021.626752.

Lợi NV, Hà HT, Thành DV, Hùng LT. 2018. Assessing suitability of two rattan species (*Daemonorops poilanei* J.Dransf. and *D. jenkinsiana* Mart.) in natural forests in Nam Dong district, Thua Thien Hue Province. Hue Univ J Sci Agric Rural Dev 127 (3B): 151-161. DOI: 10.26459/hueuni-jard.v127i3B.4863. [Vietnamese]

Meher PK, Sahu TK, Rao AR. 2016. Identification of species based on DNA barcode using k-mer feature vector and random forest classifier. Gene 592 (2): 316-324. DOI: 10.1016/j.gene.2016.07.010.

Naeem A, Khan AA, Cheema HMN, Khan IA, Buerkert A. 2014. DNA barcoding for species identification in the Palmae family. Genet Mol Res 13 (4): 10341-10348. DOI: 10.4238/2014.December.4.29.

Ojeda DI, Santos-Guerra A, Oliva-Tejera F, Jaen-Molina R, Caujapé-Castells J, Marrero-Rodríguez A, Cronk Q. 2014. DNA barcodes successfully identified Macaronesian *Lotus* (Leguminosae) species within early diverged lineages of Cape Verde and mainland Africa. AoB PLANTS 6: plu050. DOI: 10.1093/aobpla/plu050.

Peter CM, Henderson A. 2014. Systematics, Ecology and Management of Rattans in Cambodia, Laos and Vietnam. WWF-Greater Mekong and The New York Botanical Garden.

Rozas J, Ferrer-Mata A, Sánchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE, Sánchez-Gracia A. 2017. DNAsp 6: DNA sequence polymorphism analysis of large Datasets. Mol Biol Evol 34 (12): 3299-3302. DOI: 10.1093/molbev/msx248.

Su X, Liu YP, Chen Z, Chen KL. 2016. Evaluation of candidate barcoding markers in *Orinus* (Poaceae). Genet Mol Res 15 (2): 7714. DOI: 10.4238/gmr.15027714.

Tajima F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123 (3): 585-595. DOI: 10.1093/genetics/123.3.585.

Tamura K, Nei M. 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol Biol Evol 10 (3): 512-526. DOI: 10.1093/oxfordjournals.molbev.a040023.

Theodoridis S, Stefanaki A, Tezcan M, Aki C, Kokkini S, Vlachonasios KE. 2012. DNA barcoding in native plants of the Labiatae (Lamiaceae) family from Chios Island (Greece) and the adjacent Çesme-Karaburun Peninsula (Turkey). Mol Ecol Resour 12 (4): 620-633. DOI: 10.1111/j.1755-0998.2012.03129.x.

Thin NN. 2007. Methods of Plant Research. Hanoi National University Publishing House, Hanoi.

Trias-Blasi A, Vorontsova M. 2015. Botany: Plant identification is key to conservation. Nature 521 (7551): 161. DOI: 10.1038/521161c.

Vaze A, Nerkar G, Pagariya M, Devarumath RM, Prasad DT. 2010. Isolation and PCR amplification of genomic DNA from dry leaf samples of Sugarcane. Intl J Pharm BioSci 6 (2): 1-6.

Yang H-Q, Dong Y-R, Gu Z-J, Liang N, Yang J-B. 2012. A preliminary assessment of *matK*, *rbcL* and *trnHpsbA* as DNA barcodes for *Calamus* (Arecaceae) species in China with a note on *ITS*. Ann Bot Fenn 49 (5): 319-330. DOI: 10.5735/085.049.0603.

Yao G, Zhang Y-Q, Barrett C, Xue B, Bellot S, Baker WJ, Ge X-J. 2023. A plastid phylogenomic framework for the palm family (Arecaceae). BMC Biol 21 (1): 50. DOI: 10.1186/s12915-023-01544-y.