Bangladesh J. Plant Taxon. **30**(2): 213-232, 2023 (December) DOI: https://doi.org/10.3329/bjpt.v30i2.70498

MOLECULAR PHYLOGENETICS AND DATING OF ARECACEAE IN BANGLADESH INFERRED FROM *MATK* AND *RBCL* GENES

SHEIKH SUNZID AHMED, M. OLIUR RAHMAN[1], M. AJMAL ALI[2], FAHAD AL-HEMAID[2] AND JOONGKU LEE[3]

Department of Botany, Faculty of Biological Sciences, University of Dhaka, Dhaka 1000, Bangladesh

Keywords: Molecular dating; Molecular phylogenetics; Neighbor joining; Maximum likelihood; Bayesian inference; Arecaceae; *matK*; *rbcL*.

Abstract

A molecular phylogenetic investigation was undertaken for 30 species belonging to 15 genera of the palm family Arecaceae in Bangladesh to infer evolutionary relationships and molecular dating utilizing plastid-based *matK* and *rbcL* genes through multifaceted-algorithm driven approaches with Neighbor-Joining, Maximum-Likelihood, and Bayesian Inference methods. The study revealed that *matK* has better species discrimination efficiency than *rbcL* gene due to its highly variable nature. Transition/transversion bias test corroborated this finding as *matK* showed higher bias (2.632) than *rbcL* (2.235). Nucleotide substitution patterns were visualized via HYPERMUT program, which unveiled higher variability in *matK* and lower variability in *rbcL* alignment. Phylogenetic trees constructed with *matK* revealed monophyletic nature of origin for all the three subfamilies, *viz*. Arecoideae, Coryphoideae and Calamoideae, while *rbcL* trees exhibited polyphyly for Coryphoideae and monophyly for Arecoideae and Calamoideae. All the nine tribes belonging to three subfamilies demonstrated monophyletic nature in *matK* trees. Bootstrap support and Bayesian posterior probability were found to be higher in *matK* topologies than that of *rbcL*. The molecular clock test unraveled an equal evolutionary rate for *matK* and unequal rate for *rbcL* sequences. Molecular dating approach unveiled Calamoideae to be the most ancient subfamily (65.75 MYA) among the three subfamilies that originated during the Late Cretaceous period in the Mesozoic era, whereas Coryphoideae and Arecoideae were found to have originated in the Cenozoic era.

Introduction

The Arecaceae or Palmae family, commonly referred to as Palms, represents an iconic group of flowering plants, encompassing approximately 2600 species belonging to 181 genera, and is distributed across tropical and subtropical regions (Christenhusz and Byng, 2016). This monocot family boasts a well-documented fossil history tracing back to the Turonian period, approximately 89 to 93.5 million years ago (Harley, 2006). Nevertheless, molecular dating analyses suggest that the lineage predates this period by a considerable margin (Bremer, 2000; Bremer *et al.*, 2004). Palms play vital roles in numerous ecosystems, exerting significant ecological influence. Furthermore, they hold immense economic importance, featuring prominently in international trade (e.g., date palm, palm oil, coconut, rattan etc.) and sustaining the livelihoods of some of the world's most impoverished communities, both at the subsistence level and beyond. In Bangladesh,

[1]Corresponding author. Email: oliur.bot@du.ac.bd
[2]Department of Botany and Microbiology, College of Sciences, King Saud University, Riyadh 11451, Saudi Arabia
[3]Department of Environment and Forest Resources, Chungnam National University, Daehak-ro, Yuseong-gu, Daejeon, Republic of Korea

Arecaceae is represented by 40 species belonging to 20 genera. The family is characterized by the presence of large, compound leaves that are often long, narrow, either palmate or pinnate and arranged spirally at the top of the trunk, giving the palm its iconic appearance. Flowers are usually small, bisexual or unisexual, actinomorphic, sessile or very shortly pedicellate. Palms produce a variety of fruit types, including drupes (e.g., coconuts, dates), berries, and capsules (Siddiqui *et al.*, 2007).

Chloroplast genome or plastome data is pivotal for resolving phylogenetic relationships among plants, providing a rich source of genetic information that is highly conserved and offers insights into the evolutionary history and relatedness of plant species (Palmer *et al.*, 1988). The *matK* gene holds substantial significance in molecular phylogenetics due to its unique combination of conserved and variable regions. These features make it an indispensable genetic marker for studying the evolutionary relationships among plant species. Its variable regions, in particular, offer the necessary genetic diversity to distinguish closely related taxa, making *matK* particularly valuable for resolving phylogenetic relationships at lower taxonomic levels, such as intergeneric and interspecific points. Moreover, *matK* is often used in conjunction with other genetic markers to achieve a more comprehensive understanding of plant evolution, resulting in robust and accurate phylogenetic reconstructions. The contribution of this gene to finer-scale resolution in phylogenetic studies makes it an essential tool in the biologist's toolkit (Dong *et al.*, 2012; Watto *et al.*, 2016). The *rbcL* gene is of paramount importance in molecular phylogenetics due to its conserved nature and essential role in photosynthesis. This gene encodes a critical enzyme involved in carbon fixation and is highly conserved across plant taxa. Its slow evolutionary rate in coding regions, combined with its widespread presence in the chloroplast genome, makes it an ideal candidate for investigating evolutionary relationships and resolving the deep branches of the plant tree of life. The conserved nature of *rbcL* also enhances its utility in cross-species comparisons, allowing for robust phylogenetic analyses even at higher taxonomic levels (Soltis *et al.*, 2001).

Molecular dating analysis holds a central role in the field of phylogenetics, providing a potent instrument for gauging the temporal aspects of evolutionary events by leveraging genetic data (Roger and Hug, 2006). This approach helps to reconstruct the temporal dimension of phylogenetic trees, shedding light on diverged and evolved species. By examining the rate of genetic changes in specific molecular markers, such as *matK*, *rbcL*, ITS and so on, it is possible to calibrate an evolutionary "clock" and estimate the ages of common ancestors and branching points in the tree of life. Molecular dating information is vital for understanding the evolutionary history of organisms, including when and how they adapted to changing environments, migrated to new regions, or underwent significant speciation events. In addition, it helps in investigating the impact of geological and climatic events on diversification and biogeography as well as providing explanations for inquiries regarding the timing of significant evolutionary shifts, like the emergence of essential characteristics or the establishment of particular habitats. In sum, molecular dating analysis serves as a crucial bridge between genetic data and evolutionary time, enhancing our understanding of the intricate tapestry of plant life on earth and its historical development (Marshall *et al.*, 2016; Muellner-Riehl *et al.*, 2016).

A few endeavors have been made to unravel the molecular phylogeny of Arecaceae, occasionally delving into its subfamilies and tribes by investigating chloroplast DNA (cpDNA) and nuclear ribosomal DNA (nrDNA) sequences. Hahn (2002) evaluated Arecaceae based on *atpB*, *rbcL* and 18S nrDNA sequences without molecular dating assessments. Asmussen *et al.* (2006) conducted phylogenetic analysis focusing on *matK*, *rbcL*, *rps16* intron and *trnL-trnF* intergenic spacer wherein molecular dating is missing. Baker *et al.* (2011) performed a study on the Arecoideae subfamily only, while Comer *et al.* (2016) analyzed the subfamily Arecoideae

targeting nuclear genes. Nevertheless, as of yet, no concerted efforts have been undertaken to elucidate the phylogenetic and evolutionary relationships among Arecaceae members with a specific focus on the *matK* and *rbcL* genes alongside molecular clock dating. Furthermore, there has been no study towards uncovering the molecular phylogenetics and molecular dating pertinent to the Arecaceae taxa occurring in Bangladesh. Therefore, in the present investigation, we aimed to reconstruct a robust phylogeny of Arecaceae using a multi-algorithmic approach with Neighbor-Joining (NJ), Maximum-Likelihood (ML) and Bayesian Inference (BI) analyses inferred from *matK* and *rbcL* genes to shed light on the molecular evolutionary relationships of taxa. In addition, molecular dating initiative was undertaken to highlight temporal aspects of evolutionary events that impacted the diversification of Arecaceae throughout the geological time scale.

Materials and Methods

Taxon selection and retrieval of sequences

The NCBI (National Center for Biotechnology Information) Nucleotide database was explored to select and retrieve gene sequences of the member taxa of Arecaceae reported from Bangladesh (Siddiqui *et al.*, 2007). Based on availability, a total of 30 taxa belonging to 15 genera were chosen and both *matK* and *rbcL* gene sequences were downloaded in FASTA format. In addition, two species of Marantaceae, viz. *Maranta arundinacea* L. and *Marantochloa leucantha* (K. Schum.) Milne-Redh. were chosen as outgroups, and their sequences were retrieved in FASTA as well. Marantaceae was selected as the outgroup due to its close taxonomic affinity with Arecaceae, as both families belong to the clade Commelinidae, and the availability of *matK* and *rbcL* sequences for these two species. The FASTA files were accumulated together using Notepad of Windows 10 to create two separate multifasta files for *matK* and *rbcL* sequences.

Sequence alignment

All the *matK* and *rbcL* sequences were subjected to Multiple Sequence Alignment (MSA) following sequence retrieval. Multifasta *matK* and *rbcL* files were uploaded to the MAFFT server (Katoh *et al.*, 2019) for MSA. For iterative refinement, the E-INS-i method was selected which utilized CLUSTAL Omega to perform MSA. Afterwards, BLOSUM62 was fixed as the scoring matrix for amino acid sequences. All other settings were kept default before running MAFFT. The aligned sequences were retrieved in FASTA format for subsequent analyses in MEGA 11 (Tamura *et al.*, 2021). Transition-transversion bias was calculated for *matK* and *rbcL* genes using the Models module of MEGA 11. Nucleotide substitution patterns were investigated and visualized further using the HYPERMUT server (Rose and Korber, 2000).

Phylogenetic analyses

The Phylogeny module of MEGA 11 was employed to construct phylogenetic trees for *matK* and *rbcL* sequences using both distance-based and character-based approaches to corroborate the findings. Both the Neighbor-Joining (NJ) tree and the Maximum-Likelihood (ML) tree were generated with 1000 bootstrap replicates with the Kimura-2 parameter model as the substitution model. Both transition and transversion types of substitutions were included and uniform rates were selected as the substitution rates among sites. Partial deletion was preferred for the gaps or missing data treatment with a site coverage cutoff value of 95%. For Maximum-Likelihood analysis, tree inference options were additional where Nearest-Neighbor-Interchange (NNI) mode was implemented as ML heuristic method. Initial tree for ML was selected automatically and the branch swap filter was set to ‘none’.

Bayesian evolutionary analyses were carried out further using four-software packages, *i.e.*, BEAST 1.10.4, BEAUTi 1.10.4, TreeAnnotator 1.10.4 and FigTree 1.4.4 (Naro-Maciel *et al.*, 2008; Drummond *et al*., 2012). To carry out analyses, both *matK* and *rbcL* alignments were imported first in BEAUTi to generate parameter files for BEAST. Hasegawa–Kishono–Yano (HKY) was selected as the nucleotide substitution model in BEAUTi and Yule Speciation Process (Reid and Carstens, 2012) was selected as the prior tree. All settings in BEAUTi were maintained as default, and after generating parameter files, the BEAGLE library was installed to run BEAST properly. For running BEAST, the XML file was given as input, keeping double preferred precision and default rescaling scheme. The output was analyzed using TreeAnnotator to produce a FigTree editable file and later visualized using FigTree.

Molecular dating analyses

The Clocks module of MEGA 11 was employed for molecular dating analysis. Prior to that, molecular clock hypothesis was tested first for both *matK* and *rbcL* alignments. Based on the null hypothesis, the *matK* alignment was selected for subsequent molecular dating analyses using the RelTime-ML module. Firstly, the *matK* alignment file was loaded followed by Maximum-Likelihood Tree file in NEWICK format. Afterwards, the outgroup taxa of Marantaceae were specified. The TimeTree server was explored to add constraints in calibration nodes and based on the availability of taxa four nodes were selected keeping uniform distribution type of calibration (Hedges *et al*., 2006). Kimura-2 parameter model was selected as a nucleotide substitution model with uniform substitution rates. Partial deletion was followed by gaps or missing data treatment with a site coverage cutoff value of 95%. The time tree was visualized using the default Tree Explorer of MEGA 11.

Results and Discussion

Taxon sampling

The NCBI nucleotide database unveiled available *matK* and *rbcL* sequences for 30 species of Arecaceae reported from Bangladesh, belonging to 15 genera and three subfamilies. Among these three subfamilies, Coryphoideae contained the highest number of taxa (15) followed by Arecoideae (9) and Calamoideae (6). *Maranta arundinacea* L. and *Marantochloa leucantha* (K. Schum.) Milne-Redh. of the family Marantaceae were selected as outgroups. Among the investigated genera, *Calamus* L. appears to be the largest genus comprising six species. The list of the studied species along with their accession numbers and subfamilies are appended in Table 1.

Sequence alignment and phylogenetic analyses

Multiple sequence alignments revealed average nucleotide frequencies for A, T, G and C bases on both *matK* and *rbcL* alignments. In the *matK* alignments, the average frequency was recorded as 29.8%, 37.2%, 15.5% and 17.4% for A, T, G and C bases, respectively, whereas, in the *rbcL* alignment, this frequency was recorded as 27.9%, 28.9%, 22.4% and 20.8%, respectively. The number of variable and conserved sites was recorded to be 182 and 681 for *matK* alignment, while in *rbcL* alignment, these were recorded as 53 and 653, respectively. Therefore, considering variability *matK* alignment was more justified than *rbcL* alignment. Transition/transversion bias was found to be higher in *matK* (2.632) than *rbcL* (2.235) alignment (Table 2).

Transitional and transversional substitution rates were recorded as 74.56% and 25.44%, respectively for *matK* alignment. In the *rbcL* alignment, the transitional substitution rate was 69.47% and the transversional rate was 30.53%. Hence, the two genes differ in transitional and transversional substitutions, and this variation was clarified further with the physical

representation of substitutions sites in Figures 1 and 2. Figure 1 illustrates a higher variability of the *matK* gene, while Figure 2 distinctly shows lower variability in *rbcL*.

Table 1. Taxon used in the present investigation to infer phylogenetic relationships using *matK* and *rbcL* barcodes.

No.	Taxon	Subfamily	*matK* Accession	*rbcL* Accession
Ingroup				
1	*Areca catechu* L.	Arecoideae	KX783635.1	MK753924.1
2	*A. triandra* Roxb. *ex* Buch.-Ham.	Arecoideae	MK705059.1	MK753941.1
3	*Borassus flabellifer* L.	Coryphoideae	MK705088.1	MK753416.1
4	*Calamus erectus* Roxb.	Calamoideae	JQ041985.1	MK754002.1
5	*C. gracilis* Roxb.	Calamoideae	JQ041982.1	JQ042033.1
6	*C. guruba* Buch.-Ham. *ex* Mart.	Calamoideae	JQ042013.1	JQ042064.1
7	*C. longisetus* Griff.	Calamoideae	JX185542.1	JQ906811.1
8	*C. tenuis* Roxb.	Calamoideae	JX390640.1	JX185534.1
9	*C. viminalis* Willd.	Calamoideae	MK705230.1	JX502779.1
10	*Caryota mitis* Lour.	Coryphoideae	KJ708862.1	JF344847.1
11	*C. urens* L.	Coryphoideae	MK705128.1	JF344863.1
12	*Cocos nucifera* L.	Arecoideae	KX783653.1	MK753840.1
13	*Corypha umbraculifera* L.	Coryphoideae	MK705154.1	MK753393.1
14	*Dypsis lutescens* (H. Wendl.) Beentje & J. Dransf.	Arecoideae	KX783673.1	MK753724.1
15	*D. madagascariensis* (Becc.) Beentje & J. Dransf.	Arecoideae	MK705003.1	MK753718.1
16	*Elaeis guineensis* Jacq.	Arecoideae	MG648356.1	OM837689.1
17	*Licuala grandis* (T. Moore) H. Wendl.	Coryphoideae	OL354144.1	OL536973.1
18	*Livistona chinensis* (Jacq.) R. Br. *ex* Mart.	Coryphoideae	KX783705.1	GU135214.1
19	*L. speciosa* Kurz	Coryphoideae	MK704536.1	MK753429.1
20	*Phoenix acaulis* Roxb.	Coryphoideae	MK704670.1	MK753959.1
21	*P. paludosa* Roxb.	Coryphoideae	MK704675.1	MK753964.1
22	*P. rupicola* T. Anderson	Coryphoideae	MK704669.1	MK753973.1
23	*P. sylvestris* (L.) Roxb.	Coryphoideae	MK704660.1	MK753976.1
24	*Ptychosperma macarthurii* (H. Wendl. *ex* H. J. Veitch) H. Wendl. *ex* Hook. f.	Arecoideae	MK704980.1	MK753661.1
25	*Rhapis excelsa* (Thunb.) A. Henry	Coryphoideae	KX783766.1	MK753583.1
26	*R. humilis* Blume	Coryphoideae	MK704592.1	MK753575.1
27	*Roystonea oleracea* (Jacq.) O. F. Cook	Arecoideae	MK704872.1	MK753867.1
28	*R. regia* (Kunth) O.F. Cook	Arecoideae	KX783772.1	MK753868.1
29	*Wallichia caryotoides* Roxb.	Coryphoideae	MK705141.1	MK753528.1
30	*W. oblongifolia* Griff.	Coryphoideae	MK705143.1	MK753495.1
Outgroup				
1	*Maranta arundinacea* L.	Marantaceae (family)	JQ588311.1	JQ592612.1
2	*Marantochloa leucantha* (K. Schum.) Milne-Redh.	Marantaceae (family)	OL690061.1	OL536989.1

Table 2. Analysis of substitution matrix using transition/transversion rates. Each entry indicated the probability of substitution from one row (base) to another row (column).

DNA bases	A	T	C	G
***matK* alignment**				
A	-	4.63	2.22	**11.45**
T	3.86	-	**13.34**	2.02
C	3.86	**27.87**	-	2.02
G	**21.9**	4.63	2.22	-
***rbcL* alignment**				
A	-	4.38	3.25	**13.37**
T	4.23	-	**16.84**	3.41
C	4.23	**22.68**	-	3.41
G	**16.58**	4.38	3.25	-

In the present investigation, both *matK* and *rbcL* aligned sequences underwent through Neighbor-Joining (NJ) and Maximum-Likelihood (ML) analyses for a better understanding of the tree topology. The NJ tree of *matK* taxa revealed a clear segregation pattern among the member taxa of the three subfamilies of Arecaceae (Fig. 3). The subfamily Arecoideae was represented by three tribes such as Areceae, Cocoseae and Roystoneeae. All the members of Areceae, viz. *Areca catechu*, *A. triandra*, *Ptychosperma macarthurii*, *Dypsis lutescens* and *D. madagascariensis* clustered together with a bootstrap support value of 76. Two members of the tribe Cocoseae. such as *Cocos nucifera* and *Elaeis guineensis* grouped together with a bootstrap support value of 63 and another two members of the tribe Roystoneeae, such as *Roystonea regia* and *R. oleracea* clustered together showing a bootstrap value of 96. All the members of Arecoideae exhibited monophyletic nature of origin. Subfamily Coryphoideae was represented by five tribes, e.g. Livistoneae, Phoeniceae, Borasseae, Corypheae and Caryoteae, where they formed two subclusters. The first major subcluster incorporated Livistoneae and Phoeniceae whereas, the second major subcluster included the remaining three tribes. The Livistoneae tribe formed a single cluster with a bootstrap value of 97 incorporating its five member taxa, viz. *Livistona chinensis*, *L.* speciosa, *Rhapis excela*, *R. humilis* and *Licuala grandis.* Four species of Phoeniceae including *Phoenix acaulis*, *P. paludosa*, *P. rupicola* and *P. sylvestris* grouped together separately with bootstrap support of 97. Both the Borasseae and Corypheae tribes were represented by a single species, namely *Borassus flabellifer* (Borasseae) and *Corypha umbraculifera* (Corypheae). These two tribes grouped together and got separated from Caryoteae. Caryoteae clustered together with four taxa belonging to two genera, viz. *Caryota* and *Wallichia* with good bootstrap support values. Like Arecoideae, Coryphoideae also exhibited monophyletic nature of origin. Subfamily Calamoideae demonstrated monophyletic nature of origin and was represented by a single tribe, Calameae which circumscribed six species of the genus *Calamus* with a very good bootstrap support value. The outgroup taxa *Maranta arundinacea* and *Marantochloa leucantha* clustered distinctively, clearly showing the point of divergence for Arecaceae, and formed the root of the tree with perfect bootstrap support (100).

Maximum-Likelihood (ML) tree of *matK* reflected the relationships of three subfamilies, presenting a very close affinity with the NJ-*matK* tree topology (Fig. 4). Bootstrap support for the ML-*matK* tree was much better than that of the NJ-*matK* tree. All the terminal and internal nodes showed bootstrap scores above 50, with the majority of them showing scores exceeding 80. Three tribes of Arecoideae such as, Areceae, Cocoseae and Roystoneeae clustered together and separated

from Coryphoideae and Calamoideae. Coryphoideae members formed two subclusters incorporating its five tribes. Livistoneae and Phoeniceae grouped together in the first subcluster and the remaining three tribes, Borasseae, Corypheae and Caryoteae clustered together in the second subcluster. Members of Calamoideae represented very good bootstrap support (100) and clustered together showing their monophyletic nature of origin.

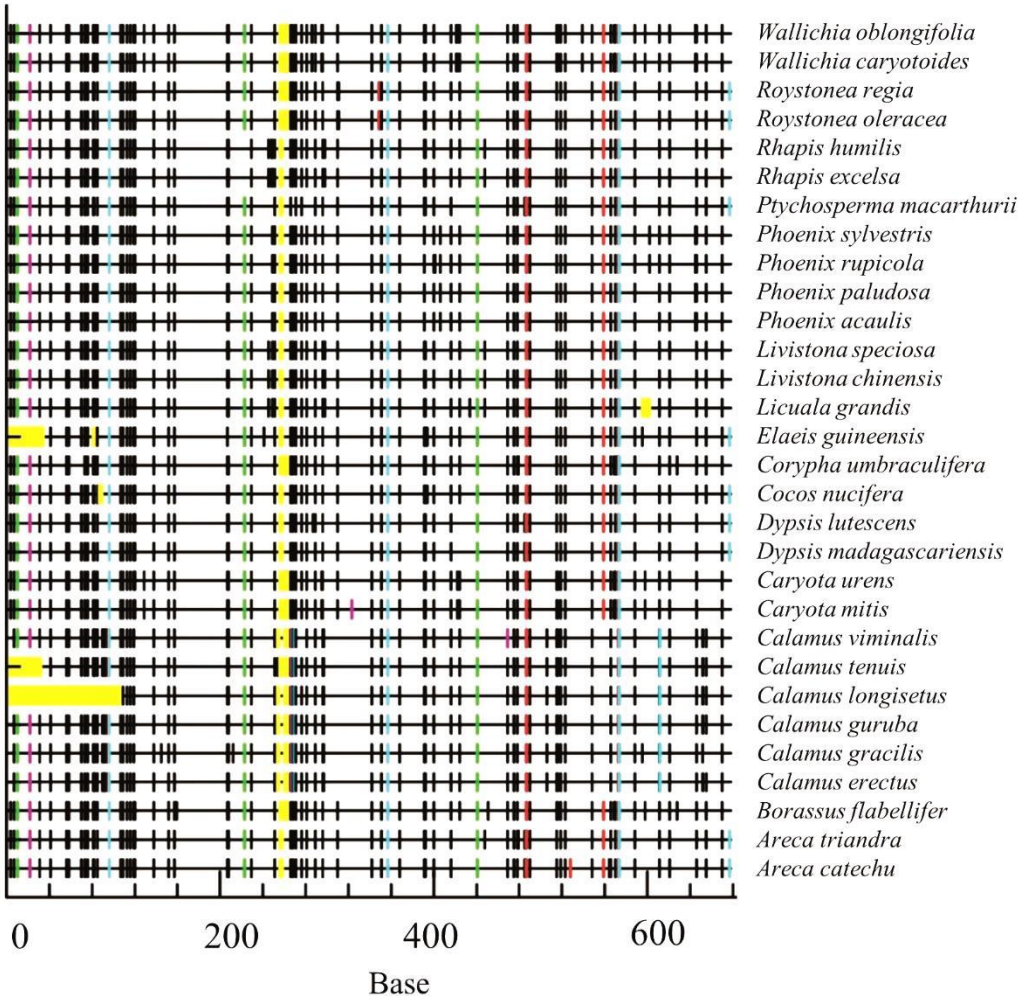

Fig. 1. Distribution of substitution sites across the *matK* region obtained from 30 species of Arecaceae using *Marantochloa leucantha* as reference (Red=GG to AG, Cyan=GA to AA, Green=GC to AC, Magenta=GT to AT, Black= not G to A transition, Yellow=gap).

The NJ tree of *rbcL* unveiled the polyphyletic nature of origin for the subfamily Coryphoideae, and monophyly for Arecoideae and Calamoideae (Fig. 5). In the subfamily Arecoideae, tribe Areceae and tribe Roystoneeae demonstrated monophyletic nature; however, the tribe Cocoseae exhibited polyphyletic nature of origin. Two taxa of Cocoseae (e.g. *Cocos nucifera* and *Elaeis guineensis*) did not share any common ancestor with each other. *Cocos nucifera*

grouped with the *Roystonea* clade (Roystoneeae) and *Elaeis guineensis* clustered with *Areca* clade (Areceae). In the Coryphoideae subfamily, four tribes, i.e. Livistoneae, Caryoteae, Borasseae and Corypheae claded together but the tribe Phoeniceae was claded outside the subcluster. Tribe Livistoneae was polyphyletic as *Licuala grandis* did not share common ancestry with *Rhapis* and *Livistona* clade. Tribe Caryoteae, Borasseae and Corypheae were found to have monophyletic origin. In the subfamily Calamoideae, *Calamus tenuis*, *C. guruba* and *C. erectus* demonstrated closer similarity than the other three species within Calamoideae. Bootstrap support values were moderately well at the terminal nodes than the internal nodes, however, in overall consideration Boostrap support was found to be weaker than *matK* trees.

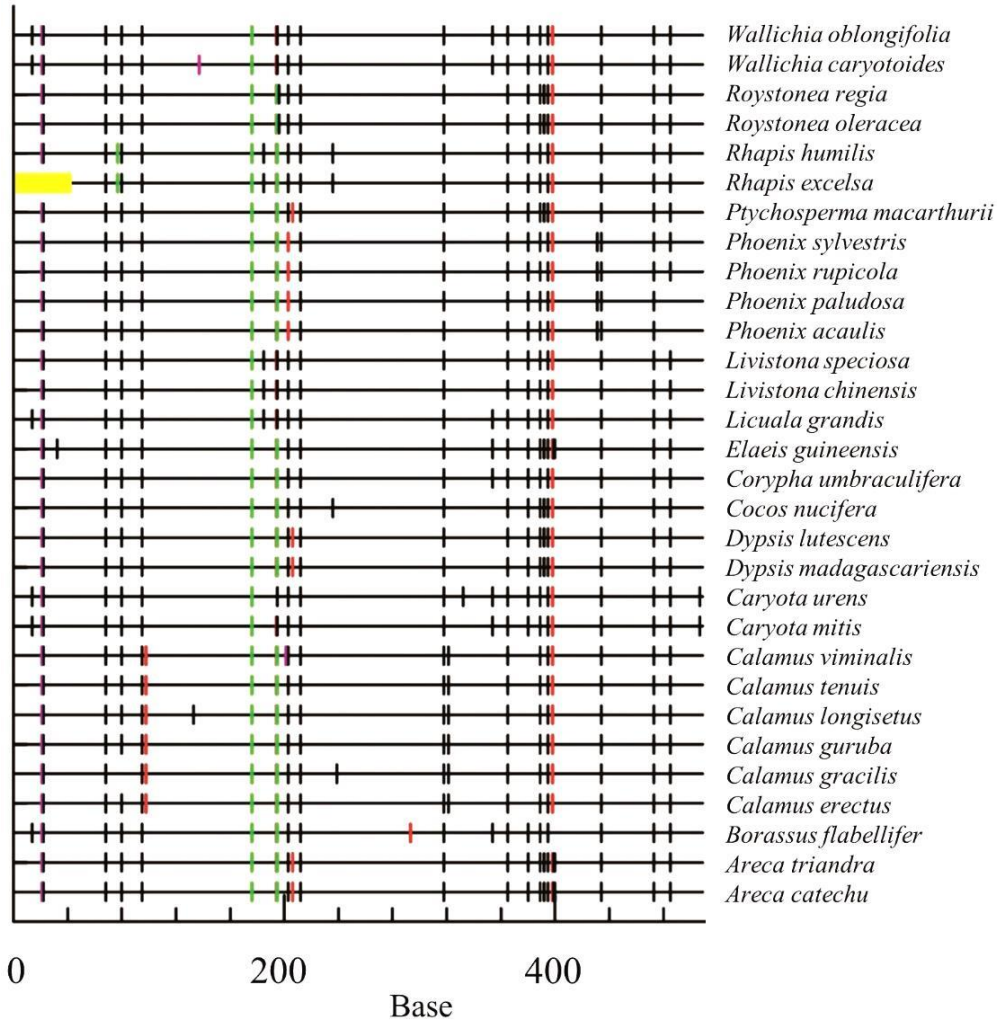

Fig. 2. Distribution of substitution sites across the *rbcL* region obtained from 30 species of Arecaceae using *Marantochloa leucantha* as reference (Red=GG to AG, Cyan=GA to AA, Green=GC to AC, Magenta=GT to AT, Black=not G to A transition, Yellow=gap).

The ML tree of *rbcL* taxa unraveled a tree topology very similar to that of the NJ-*rbcL* tree (Fig. 6). In comparison to bootstrap support, the ML-*rbcL* tree was found to have weaker support than the NJ-*rbcL* tree. The root was supported with an almost perfect bootstrap value (99), though in the internal nodes the values decreased significantly. A significant downfall was observed at the terminal nodes with a relatively small number having higher bootstrap support of over 80. Coryphoideae was found to be polyphyletic, while Arecoideae and Calamoideae were observed to be monophyletic. *Phoenix* clade was the underlying reason for the polyphyletic nature of Coryphoideae. Due to *Licuala grandis*, tribe Livistoneae became polyphyletic while tribe Caryoteae, Borasseae and Corypheae remained monophyletic. In Arecoideae, the tribes Areceae and Roystoneeae were monophyletic, and the tribe Cocoseae was found to be polyphyletic. Subfamily Calamoideae exhibited good bootstrap support at the internal nodes as well as monophyletic nature of origin with its six member taxa of *Calamus.*

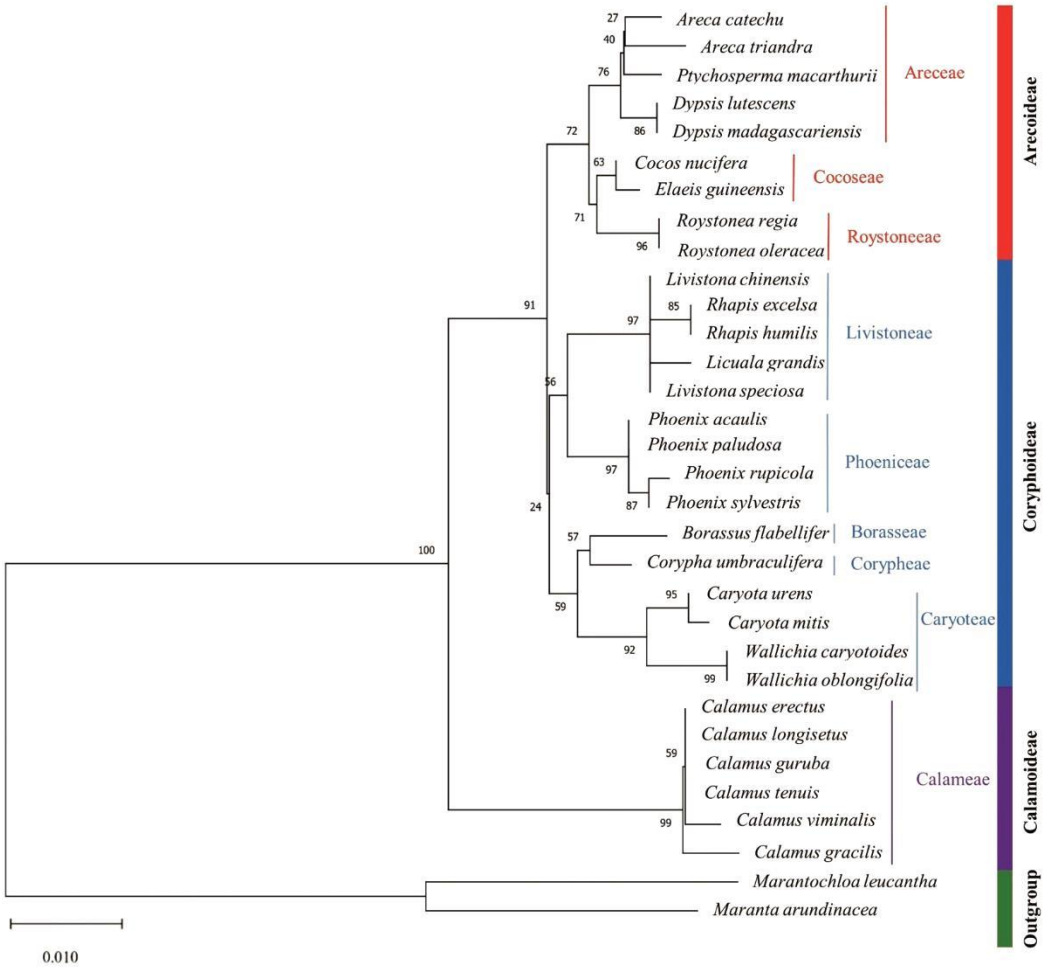

Fig. 3. Neighbor Joining tree showing inter-relationships among three subfamilies of Arecaceae using 1000 bootstrap replicates based on *matK* gene.

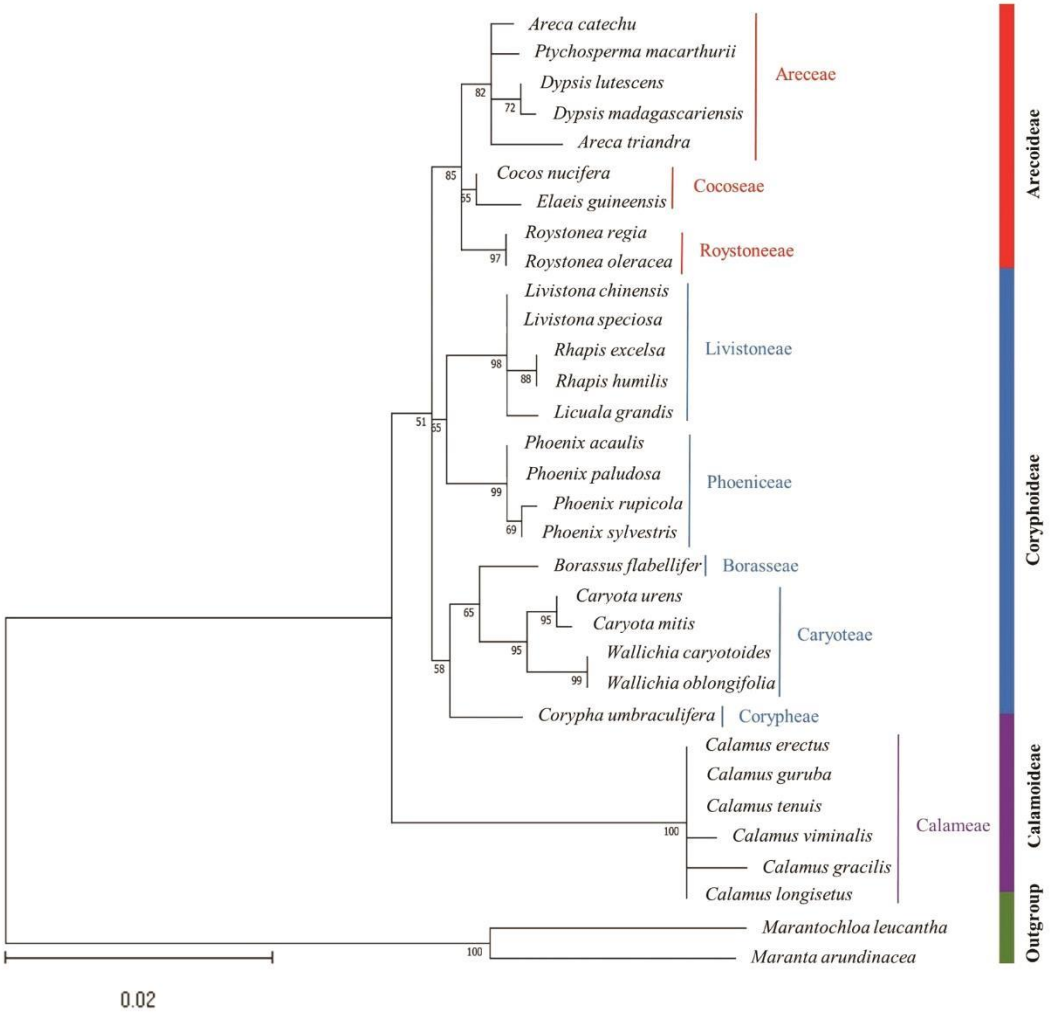

Fig. 4. Maximum Likelihood tree showing inter-relationships among three subfamilies of Arecaceae using 1000 bootstrap replicates based on *matK* gene.

The Bayesian evolutionary tree was analyzed further to corroborate our phylogenetic study, revealing significant results that correlated with both NJ and ML approaches for both *matK* and *rbcL* genes. The Bayesian tree demonstrated strong posterior probability support for the *matK* tree (Fig. 7). The *matK* tree exhibited 100% posterior probability support to signify monophyletic nature of origin for Arecoideae, Coryphoideae and Calamoideae. This finding is congruent with the NJ-*matK* and ML-*matK* trees, providing additional support for the constructed phylogeny of Arecaceae. Within the subfamily Arecoideae, all three tribes were monophyletic and did not converge with the members of Coryphoideae or Calamoideae. Of the two major subclusters of Coryphoideae, the tribes Corypheae, Borasseae and Caryoteae formed the first cluster, while the second one consisted of the species of the tribes Livistoneae and Phoeniceae. Most of the terminal nodes and many internal nodes showed nearly 100% posterior probability.

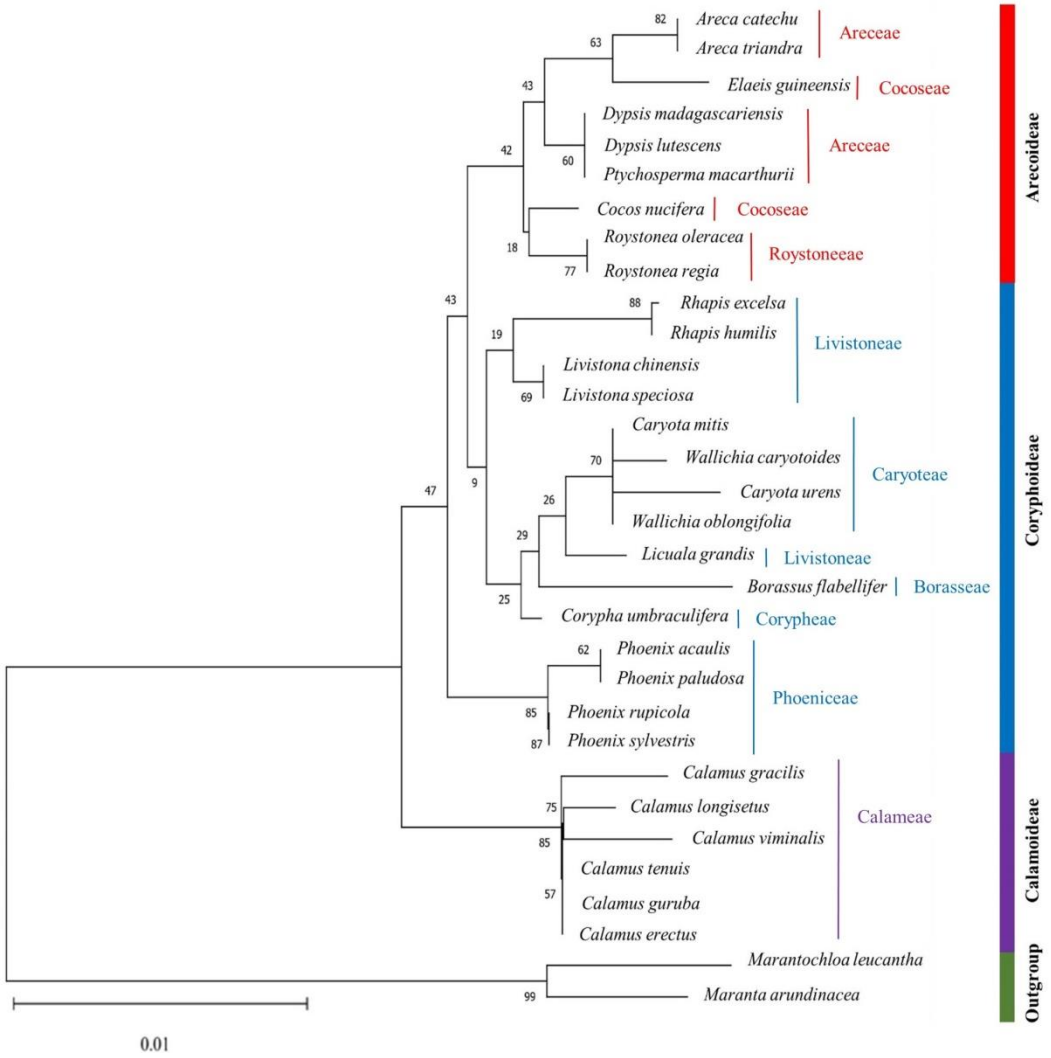

Fig. 5. Neighbor Joining tree showing inter-relationships among three subfamilies of Arecaceae using 1000 bootstrap replicates based on of *rbcL* gene.

The subfamily Calamoideae unraveled moderately strong support for *Calamus tenuis*, *C. viminalis* and *C. guruba* compared to the other three species within Calamoideae. Marantaceae was supported with 100% confidence as outgroup. The Bayesian-*rbcL* tree unveiled similar tree topologies to the NJ-*rbcL* and ML-*rbcL* trees (Fig. 8). Arecoideae demonstrated polyphyletic nature of origin for the tribe Areceae. The tribe Cocoseae showed polyphyletic nature, while the tribe Roystoneeae exhibited monophyletic origin. Coryphoideae showed polyphyletic origin, and Calamoideae unraveled monophyletic origin. The outgroup Marantaceae was corroborated with perfect posterior probability support. Bayesian inference for *matK* and *rbcL* has been visualized as radiation diagram in Figure 9.

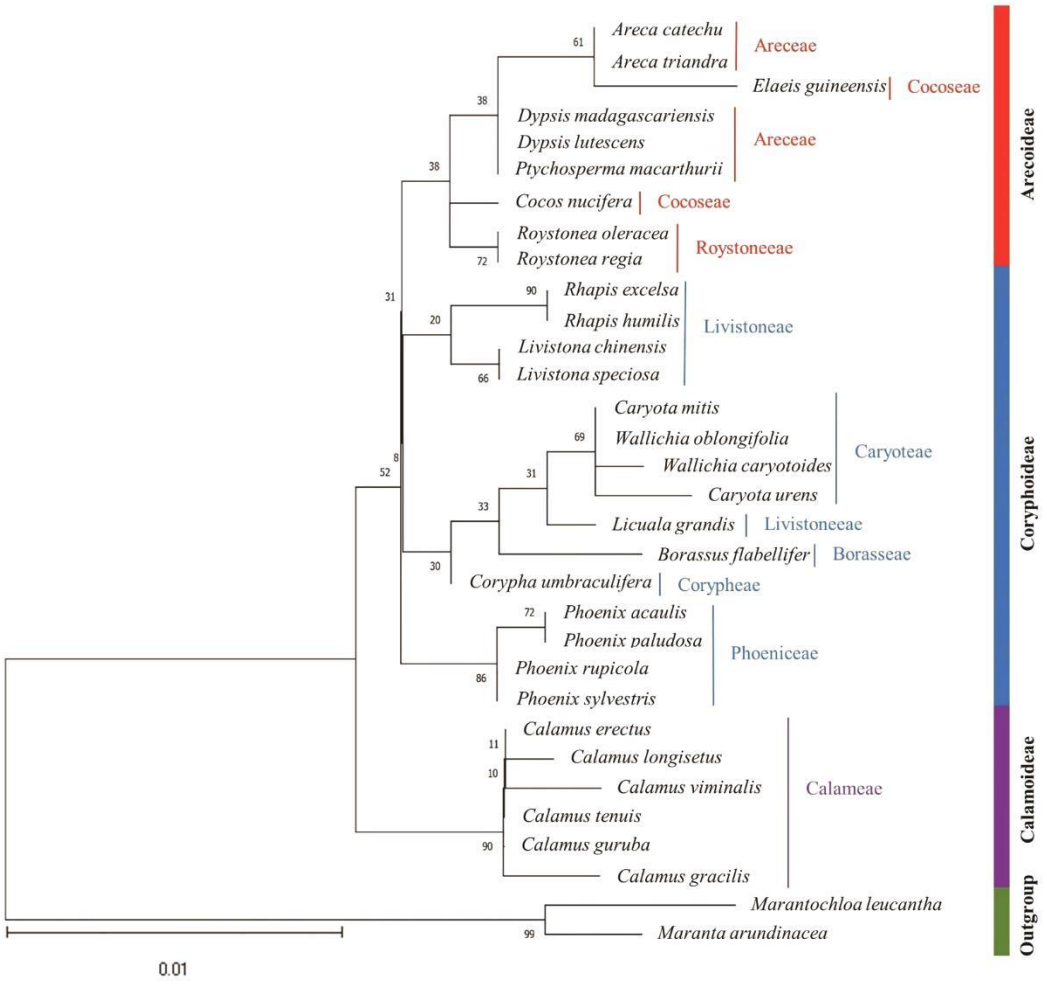

Fig. 6. Maximum Likelihood tree showing inter-relationships among three subfamilies of Arecaceae using 1000 bootstrap replicates based on of *rbcL* gene.

Molecular dating analyses

The molecular dating was performed based on the null hypothesis, where it was hypothesized that the rate of molecular evolution or the rate of nucleotide substitutions is constant across the branches of the phylogenetic trees. The test in MEGA 11 unveiled acceptance for *matK* sequences and rejection for *rbcL* sequences. The P value at 5% significant level was denoted as 0.1335, and a total of 630 positions were covered in the final dataset of *matK*. On the contrary, the P value at 5% significant level was marked as 0.4648 with a coverage of 508 positions in the final dataset of *rbcL* sequences. Consequently, we carried out molecular dating analysis for *matK* sequences. Prior to molecular dating analyses, the TimeTree server revealed a total of four calibration points using four pairs of taxa for efficient calculation of the time tree (Fig. 10). The pairs were: (a) *Borassus flabellifer* vs *Corypha umbraculifera*, (b) *Borassus flabellifer* vs *Cocos nucifera*, (c) *Areca catechu* vs *Roystonea regia* and (d) *Rhapis excelsa* vs *Phoenix sylvestris*. The calibration points were fixed based on the availability of data in the TimeTree server. The tree unraveled the first

point of divergence about 65.75 Million Years Ago (MYA) from the outgroup Marantaceae during the Late Cretaceous period in the Mesozoic era that resulted in the separate occurrence of Calamoideae (Fig. 11).

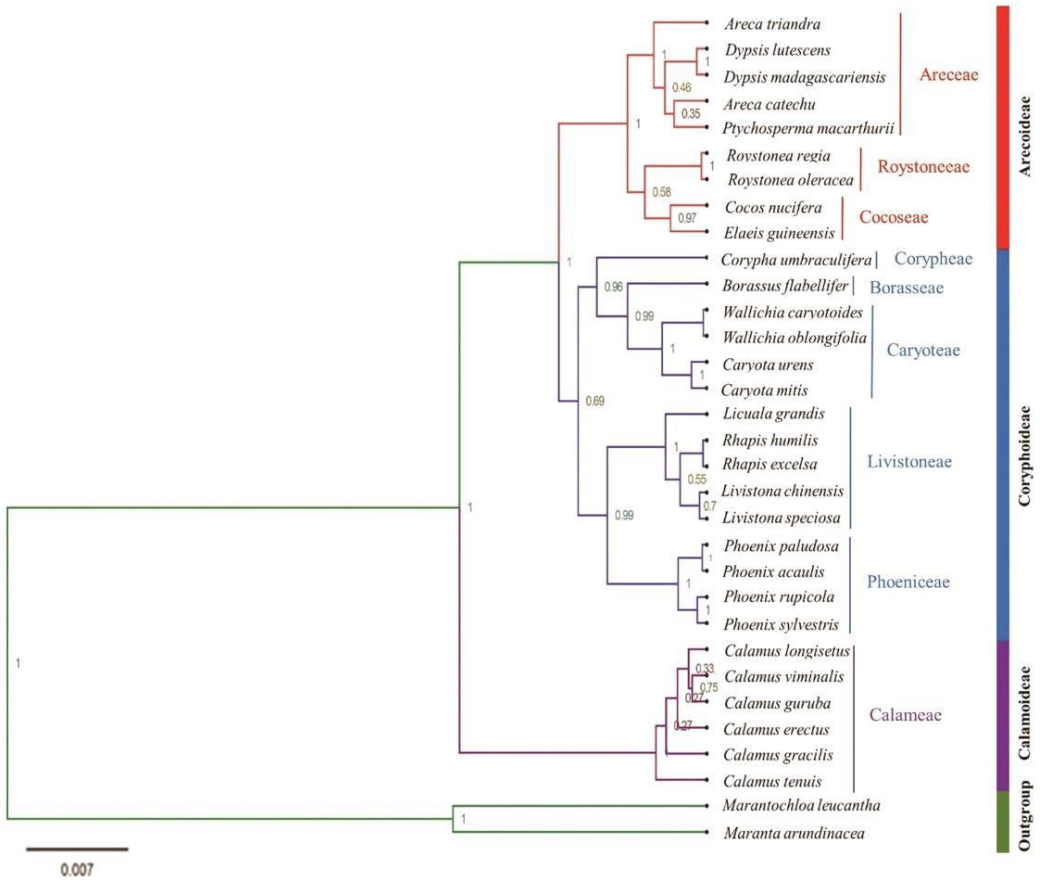

Fig. 7. Bayesian Inference analysis showing inter-relationships among three subfamilies with posterior probability values based on *matK* gene.

According to the geological time scale, the most ancient species among the six species in the genus *Calamus* is *C. longisetus* (3.18 MYA), whereas the most recently evolved taxa are *C. erectus*, *C. guruba* and *C. tenuis* (0.23 MYA). Among the three subfamilies, Calamoideae is the oldest (65.75 MYA) followed by Coryphoideae (39.00 MYA) and Arecoideae (23.82 MYA). Within the Arecoideae subfamily, the tribes Roystoneeae and Cocoseae originated earlier (22.18 MYA) during the Neogene period of the Cenozoic era, while the tribe Areceae originated later (14.07 MYA) in the same period of the Cenozoic era.

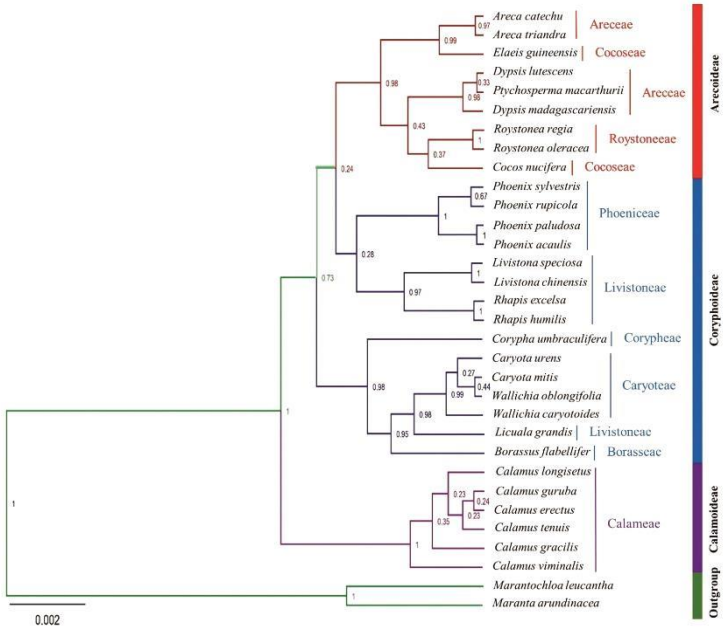

Fig. 8. Bayesian Inference analysis showing inter-relationships among three subfamilies of Arecaceae with posterior probability values based on *rbcL* gene.

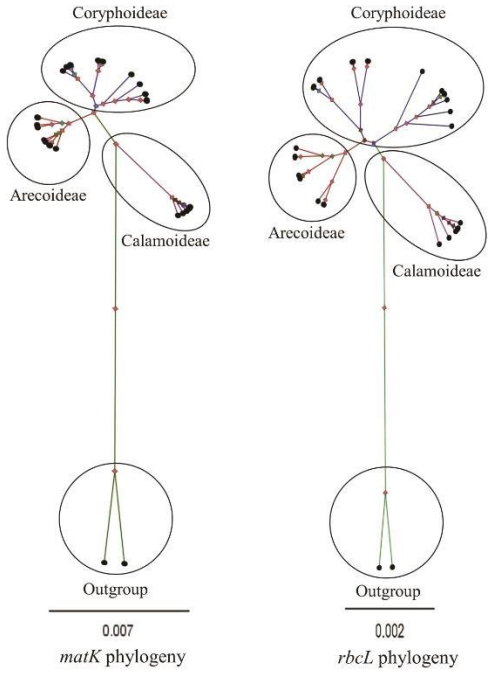

Fig. 9. Radial representation of divergence of three subfamilies of Arecaceae following Bayesian Inference analysis. Black circles are showing terminal nodes and square boxes are demonstrating internal nodes.

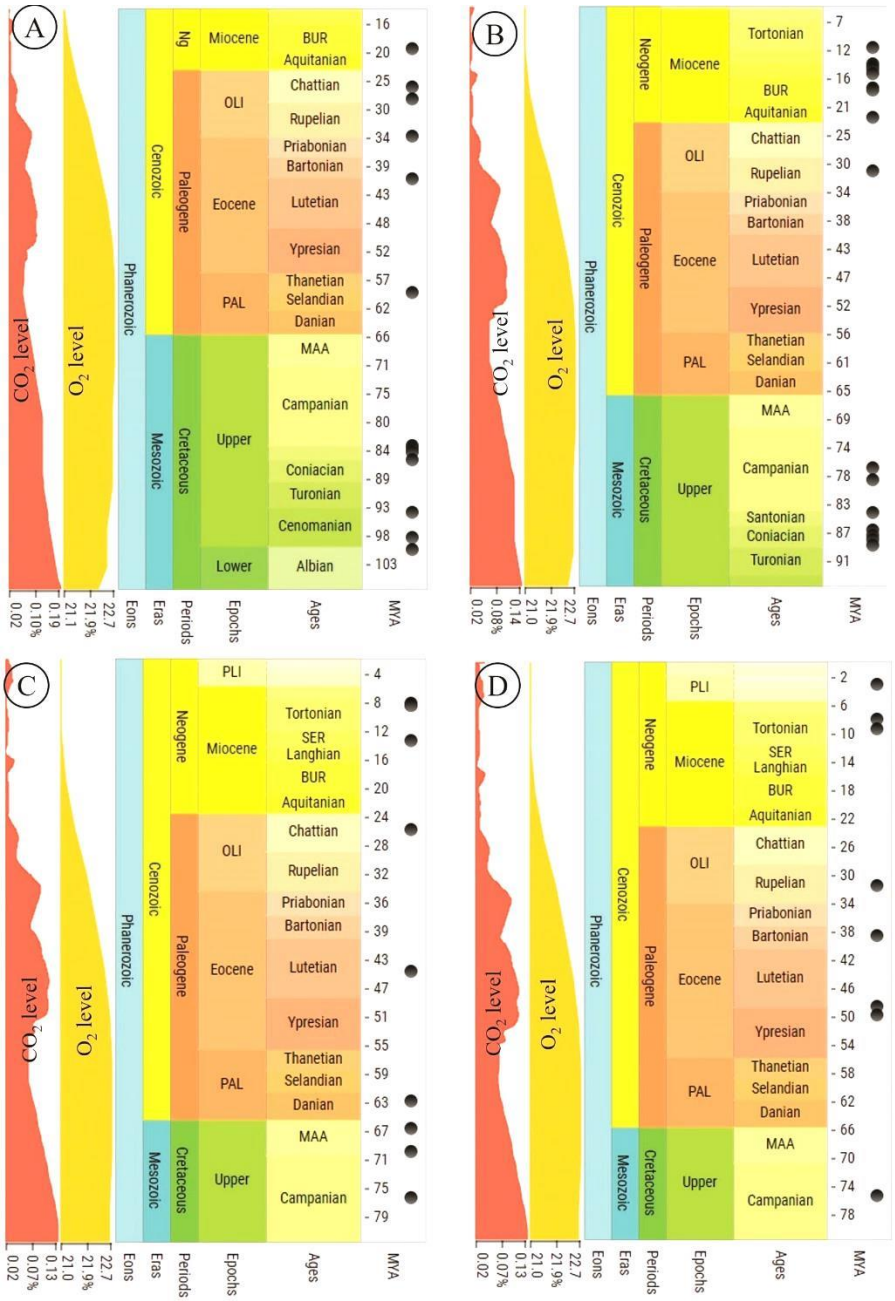

Fig. 10. Pairwise divergent times for different species of Arecaceae used in the calibration nodes. A. *Borassus flabellifer* and *Corypha umbraculifera*, median time: 84 MYA, Confidence Interval (CI): (33.8-85.8) MYA, adjusted time: not available; B. *Borassus flabellifer* and *Cocos nucifera*, median time: 22.3 MYA, CI: (14.4-83.8) MYA, adjusted time: 62.3 MYA; C. *Areca catechu* and *Roystonea regia*, median time: 35 MYA, CI: (8.1-69.9) MYA, adjusted time: 50 MYA; D. *Rhapis excelsa* and *Phoenix sylvestris*, median time: 35 MYA, CI: (8.0-49.8) MYA, adjusted time: 55 MYA.

In the subfamily Coryphoideae, the tribe Livistoneae diverged earlier (39.00 MYA) followed by the tribes Corypheae (34.17 MYA), Phoeniceae (32.53 MYA), Borasseae (25.18 MYA) and Caryoteae (12.62 MYA). Both the Corypheae and Phoeniceae tribes evolved during the Eocene epoch of the Paleogene period in the Cenozoic era (Fig. 11). Borasseae originated during the Oligocene epoch of the Paleogene period in the Cenozoic era, while Caryoteae evolved during the Miocene epoch of the Neogene period in the Cenozoic era.

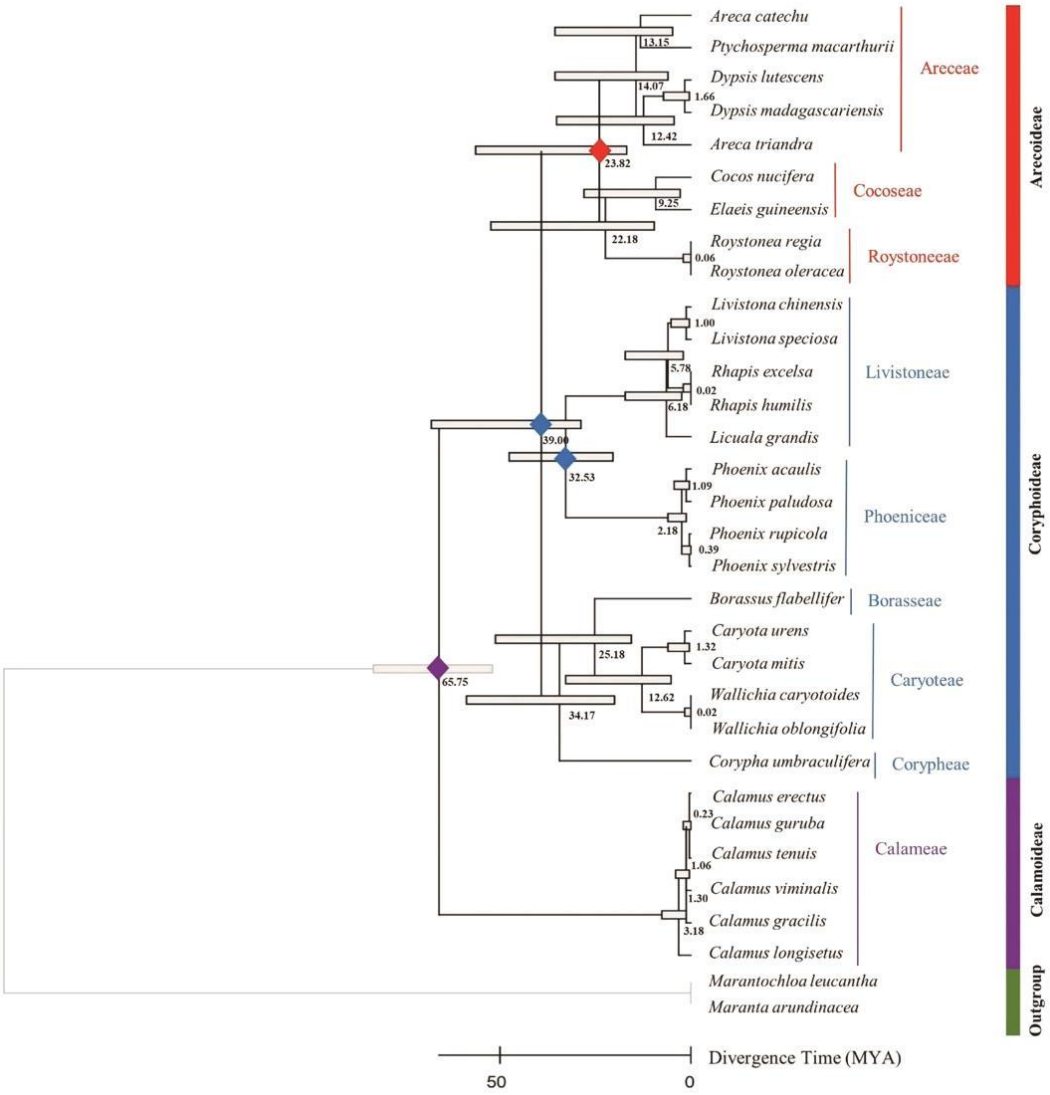

Fig. 11. Molecular dating assessment showing time tree for the three subfamilies of Arecaceae. Squares indicate calibration nodes used to construct the time tree.

The present investigation shed light on the molecular phylogeny of Arecaceae employing *matK* and *rbcL* barcodes of the chloroplast genome. The study only considered those taxa reported from Bangladesh based on the availability of sequence information in the NCBI nucleotide database. In Bangladesh, Arecaceae is represented by 20 genera and 40 species (Siddiqui *et al.*, 2007). Sequence information of both *matK* and *rbcL* genes was available for 30 species under 15 genera which was analyzed in the present study (Table 1). Application of NCBI public sequence data alone to resolve phylogenetic relationships was supported by several studies (Gholizadeh *et al.*, 2013; Ali *et al.*, 2020; Aykut, 2020). Several species including *Arenga pinnata* (Wurmb) Merr., *Calamus latifolius* Roxb., *Corypha taliera* Roxb., *Daemonorops jenkinsiana* (Griff.) Martius, *Didymosperma gracilis* Hook. f., *D. nanum* H. Wendl. & Drude, *Licuala peltata* Roxb., *L. spinosa* Jhun., *Nypa fruticans* Wurmb, and *Pinanga gracilis* Blume, were not included in the analysis due to their unavailability in the nucleotide database of NCBI. Arecaceae is globally represented by five subfamilies such as, Arecoideae, Calamoideae, Ceroxyloideae, Coryphoideae and Nypoideae (Asmussen *et al.*, 2006), however, we employed three subfamilies (e.g. Arecoideae, Calamoideae and Coryphoideae) in our study since the remaining two subfamilies are missing in the flora of Bangladesh.

The current study aimed to understand the molecular evolutionary relationships of the three subfamilies, viz. Arecoideae, Coryphoideae and Calamoideae, and to infer their molecular dating. Yao *et al.* (2023) proposed a plastome-based phylogenomic framework of Arecaceae, where Arecoideae, Coryphoideae and Calamoideae demonstrated monophyletic nature of origin. In the present investigation, the *matK* phylogeny of these three subfamilies was found congruent with the findings of Yao *et al.* (2023). Asmussen *et al.* (2006) proposed a new classification for these subfamilies based on plastid DNA sequences including *rbcL*, *trnL-trnF*, *matK* and *rps16*, and their findings revealed the divergence of Arecoideae and Coryphoideae from Calamoideae, which aligns with the results of our study (Figs 7 & 8). In a previous study, a close relationship was found between *Rhapis excelsa*, *Licuala kunstleri* and *Livistona chinensis*. Within the tribe Caryoteae, *Caryota mitis* clustered with *Wallichia distica*, and the tribe Phoeniceae exhibited a closer proximity to the tribe Livistoneae than to the tribe Caryoteae (Asmussen *et al.*, 2006). These findings were found congruent with our study, in particular, for the *matK* derived phylogeny (Figs 3, 4 & 7). Comer *et al.* (2016) carried out a phylogenetic study of the subfamily Arecoideae and its 14 tribes employing nuclear genes, where the tribe Roystoneeae clustered with the tribe Cocoseae. A similar phylogeny of Arecoideae using the chloroplast gene *matK* was reconstructed in the current investigation (Figs 3, 4 & 7). The PRK (Phosphoribulokinase) and RPB2 (RNA polymerase II, subunit B) genes of nuclear genome were analyzed to delineate phylogeny of the subfamily Arecoideae where subtribe Attaleinae (*Cocos nucifera*) clustered together with the subtribe Elaeidinae (*Elaeis guineensis*) under the same clade of the tribe Cocoseae (Baker *et al.*, 2011). Our *matK* phylogeny aligned with this finding and supported the position of tribe Cocoseae under the subfamily Arecoideae. However, the *rbcL* phylogeny of present study was not consistent with the findings of Baker *et al.* (2011).

In the present investigation, *matK* trees were found to be comparatively more consistent, accurate and congruent to segregate lineages of Arecoideae, Coryphoideae and Calamoideae than *rbcL* trees. We hypothesize that several factors are responsible for this variation between *matK* and *rbcL* phylogenies. The rate of evolution could be a predominant cause, where *rbcL* may get evolve at a faster or slower rate in some lineages within Arecaceae leading to more sequence variation and inconsistency in phylogeny. During the molecular clock test, *rbcL* alignment was not supported by the null hypothesis which further corroborates this supposition (DeBry, 1992; Huelsenbeck and Hillis, 1993). The function of the gene can also influence its consistency. As *rbcL* is involved in photosynthesis, a fundamental process, it may undergo different selective

pressures in different lineages of Arecaceae, affecting its sequence evolution. The length of the sequence employed in phylogenetic analysis can affect its reliability. When a gene offers a longer and more informative sequence, it has the potential to yield more dependable outcomes (Moreira and Philippe, 2000). As *matK* furnished longer sequences in contrast to *rbcL* in the present study, it showed more accuracy than *rbcL*. Asahina *et al.* (2010) used a similar protocol to the present investigation to resolve the phylogeny of medicinal *Dendrobium* species using *matK* and *rbcL* genes, where they reported *matK* to have better species discriminating power than *rbcL* which was further supported by our study. A combination of *matK* and *rbcL* has been used in several studies to resolve the phylogeny of plants which justifies our selection of these two plastid genes for Arecaceae (Goldman *et al.*, 2001; Bello *et al.*, 2009; Ortiz-Covarrubias *et al.*, 2022).

Molecular dating analyses unveiled divergence periods and era for the three subfamilies, and the oldest point was recorded for the Late Cretaceous period of the Mesozoic era (65.75 MYA). The late Cretaceous period is significant for the rapid diversification and proliferation of angiosperms. This period witnessed the co-evolutionary "arms race" between angiosperms and insects. Many angiosperms developed specialized structures, such as flowers and nectar, to attract pollinators, like bees and butterflies. This co-evolutionary interaction contributed to the success of both groups, and shaped the modern biodiversity of flowering plants and insect pollinators (Batten, 1981). Cornejo *et al*. (2017) performed molecular dating with chloroplast genome data to resolve the phylogeny of the species *Stachys coccinea* employing a single calibration point. In our investigation, we have used four calibration points which corroborates the protocol more informative than Cornejo *et al.* (2017). The RelTime-ML module of MEGA has been used by several studies for molecular dating analyses which justifies our choice of using this package for molecular dating venture (Tokhmechi *et al.*, 2021; Kakhki *et al.*, 2023; Lyu *et al.*, 2023).

Arecaceae is an important angiosperm family that includes many medicinally and economically important species. Understanding the relationships among the member taxa of the family would clarify their systematic position and substantiate their molecular authentication based on genomic information derived from the plastome. Molecular dating information would provide additional phylogenetic support in relation to evolutionary divergence according to geological time scale. This approach would shed light further on the historical biogeography of Arecaceae by estimating their colonizing patterns throughout different regions of the world and speciation events in geological time scale. Until now, no efforts have been made to establish the phylogeny of Arecaceae taxa in Bangladesh using the *matK* and *rbcL* barcodes. Our study marks the inaugural endeavor to elucidate the phylogenetic relationships among Arecaceae species in Bangladesh, thereby validating the utility of two chloroplast DNA barcodes. This validation is accomplished through a comprehensive comparative analysis of phylogenetic relationships and evolutionary divergence, aligning with the geological time scale.

Acknowledgement

The authors extend their appreciation to the Researchers supporting project number (RSP2023R306), King Saud University, Riyadh, Saudi Arabia.

References

Ali, M.A., Rahman, M.O., Lee, J., Al Hemaid, F., Kambhar, S.V., Elangbam, M. and Gurung, A.B. 2020. Dissecting molecular evolutionary relationship of Krameriaceae inferred from phylotranscriptomic analysis. Bangladesh J Plant Taxon. **27**(2): 427-433.

Asahina, H., Shinozaki, J., Masuda, K., Morimitsu, Y. and Satake, M. 2010. Identification of medicinal *Dendrobium* species by phylogenetic analyses using matK and rbcL sequences. J. Nat. Med. **64**: 133-138.

Asmussen, C.B., Dransfield, J., Deickmann, V., Barfod, A.S., Pintaud, J.C. and Baker, W.J. 2006. A new subfamily classification of the Palm family (Arecaceae): Evidence from plastid DNA phylogeny. Bot. J. Linn. Soc. **151**(1): 15-38.

Aykut, Y. 2020. The importance in DNA barcoding of the regions which is covering rRNA genes and its sequences in the genus *Quercus* L. Bangladesh J. Plant Taxon. **27**(2): 261-271.

Baker, W.J., Norup, M.V., Clarkson, J.J., Couvreur, T.L., Dowe, J.L., Lewis, C.E., Pintaud, J., Savolainen, V., Wilmot, T. and Chase, M.W. 2011. Phylogenetic relationships among Arecoid Palms (Arecaceae: Arecoideae). Ann. Bot. **108**(8): 1417-1432.

Batten, D.J. 1981. Stratigraphic, palaeogeographic and evolutionary significance of Late Cretaceous and early Tertiary Normapolles pollen. Rev. Palaeobot. Palynol. **35**(2-4): 125-137.

Bello, M.A., Bruneau, A., Forest, F. and Hawkins, J.A. 2009. Elusive relationships within order Fabales: Phylogenetic analyses using *matK* and *rbcL* sequence data. Syst. Bot. **34**(1): 102-114.

Bremer, K. 2000. Early Cretaceous lineages of monocot flowering plants. Proc. Natl. Acad. Sci. **97**(9): 4707-4711.

Bremer, K., Friis, E. and Bremer, B. 2004. Molecular phylogenetic dating of Asterid flowering plants shows early Cretaceous diversification. Syst. Biol. **53**(3): 496-505.

Christenhusz, M.J. and Byng, J.W. 2016. The number of known plants species in the world and its annual increase. Phytotaxa **261**(3): 201-217.

Comer, J.R., Zomlefer, W.B., Barrett, C.F., Stevenson, D.W., Heyduk, K. and Leebens-Mack, J.H. 2016. Nuclear phylogenomics of the Palm subfamily Arecoideae (Arecaceae). Mol. Phylogenet. Evol. **97**: 32-42.

Cornejo, A.M.B., Lindqvist, C., Balch, E.M.P.M. and Delgado, M.E.S. 2017. Phylogeny of the *Stachys coccinea* (Lamiaceae) complex based on molecular and morphological data. Syst. Bot. **42**(3): 484-493.

DeBry, R.W. 1992. The consistency of several phylogeny-inference methods under varying evolutionary rates. Mol. Biol. Evol. **9**(3): 537-551.

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

Drummond, A.J., Suchard, M.A., Xie, D. and Rambaut, A. 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. Mol. Biol. Evol. **29**(8): 1969-1973.

Gholizadeh, S., Djadid, N.D., Nouroozi, B. and Bekmohammadi, M. 2013. Molecular phylogenetic analysis of *Anopheles* and *Cellia* subgenus Anophelines (Diptera: Culicidae) in temperate and tropical regions of Iran. Acta Trop. **126**(1): 63-74.

Goldman, D.H., Freudenstein, J.V., Kores, P.J., Molvray, M., Jarrell, D.C., Whitten, W.M., Cameron, K.M., Jansen, R.K. and Chase, M.W. 2001. Phylogenetics of Arethuseae (Orchidaceae) based on plastid *matK* and *rbcL* sequences. Syst. Bot. **26**(3): 670-695.

Hahn, W.J. 2002. A molecular phylogenetic study of the Palmae (Arecaceae) based on *atpB*, *rbcL*, and 18S nrDNA sequences. Syst. Biol. **51**(1): 92-112.

Harley, M.M. 2006. A summary of fossil records for Arecaceae. Bot. J. Linn. Soc. **151**(1): 39-67.

Hedges, S.B., Dudley, J. and Kumar, S. 2006. TimeTree: A public knowledge-base of divergence times among organisms. Bioinformatics **22**(23): 2971-2972.

Huelsenbeck, J.P. and Hillis, D.M. 1993. Success of phylogenetic methods in the four-taxon case. Syst. Biol. **42**(3): 247-264.

Kakhki, A.N., Schweizer, M., Lutgen, D., Bowie, R.C., Shirihai, H., Suh, A., Schielzeth, H. and Burri, R. 2023. A phylogenomic assessment of processes underpinning convergent evolution in open-habitat chats. Mol. Biol. Evol. **40**(1): msac278.

Katoh, K., Rozewicki, J. and Yamada, K.D. 2019. MAFFT online service: Multiple sequence alignment, interactive sequence choice and visualization. Brief. Bioinformatics **20**(4): 1160-1166.

Lyu, S., Mei, Q., Liu, H., Wang, B., Wang, J., Lambers, H., Wang, Z., Dong, B., Liu, Z. and Deng, S. 2023. Genome assembly of the pioneer species *Plantago major* L. (Plantaginaceae) provides insight into its global distribution and adaptation to metal-contaminated soil. DNA Res. **30**(4): dsad013.

Marshall, D.C., Hill, K.B., Moulds, M., Vanderpool, D., Cooley, J.R., Mohagan, A.B. and Simon, C. 2016. Inflation of molecular clock rates and dates: Molecular phylogenetics, biogeography, and diversification of a global Cicada radiation from Australasia (Hemiptera: Cicadidae: Cicadettini). Syst. Biol. **65**(1): 16-34.

Moreira, D. and Philippe, H. 2000. Molecular phylogeny: Pitfalls and progress. Int. Microbiol. **3**(1): 9-16.

Muellner-Riehl, A.N., Weeks, A., Clayton, J.W., Buerki, S., Nauheimer, L., Chiang, Y.C., Cody, S. and Pell, S.K. 2016. Molecular phylogenetics and molecular clock dating of Sapindales based on plastid *rbcL*, *atpB* and *trnL-trnF* DNA sequences. Taxon **65**(5): 1019-1036.

Naro-Maciel, E., Le, M., FitzSimmons, N.N. and Amato, G. 2008. Evolutionary relationships of marine turtles: A molecular phylogeny based on nuclear and mitochondrial genes. Mol. Phylogenet. Evol. **49**(2): 659-662.

Ortiz-Covarrubias, Y.C., Orozco-Sifuentes, M.M., Mendoza-Rodríguez, D.V., Villlarreal-Quintanilla, J.A., Martínez, O., Hernández-Godínez, F., Jáuregui-González, M.D.J. and Reyes-Valdés, M.H. 2022. Phylogeny, origin and diversification of the *Dasylirion* genus based on *matK* and *rbcL* sequences. Plant Genet. Resour. **20**(2): 108-115.

Palmer, J.D., Jansen, R.K., Michaels, H.J., Chase, M.W. and Manhart, J.R. 1988. Chloroplast DNA variation and plant phylogeny. Ann. Mo. Bot. Gard. **75**(4): 1180-1206.

Reid, N.M. and Carstens, B.C. 2012. Phylogenetic estimation error can decrease the accuracy of species delimitation: A Bayesian implementation of the general mixed Yule-coalescent model. BMC Evol. Biol. **12**(1): 1-11.

Roger, A.J. and Hug, L.A. 2006. The origin and diversification of eukaryotes: Problems with molecular phylogenetics and molecular clock estimation. Philos. Trans. R. Soc. Lond. B. Biol. Sci. **361**(1470): 1039-1054.

Rose, P.P. and Korber, B.T. 2000. Detecting hypermutations in viral sequences with an emphasis on G →A hypermutation. Bioinformatics **16**(4): 400-401.

Siddiqui, K.U., Islam, M.A., Ahmed, Z.U., Begum, Z.N.T., Hassan, M.A., Khondker, M., Rahman, M.M., Kabir, S.M.H., Ahmad, M., Ahmed, A.T.A., Rahman, A.K.A. and Haque, E.U. (Eds). 2007. Encyclopedia of Flora and Fauna of Bangladesh, Vol. **11**. Angiosperms: Monocotyledons (Agavaceae-Najadaceae). Asiatic Society of Bangladesh, Dhaka. pp. 99-138.

Soltis, D.E., Kuzoff, R.K., Mort, M.E., Zanis, M., Fishbein, M., Hufford, L., Koontz, J. and Arroyo, M K. 2001. Elucidating deep-level phylogenetic relationships in Saxifragaceae using sequences for six chloroplastic and nuclear DNA regions. Ann. Mo. Bot. Gard. **88**(4): 669-693.

Tamura, K., Stecher, G. and Kumar, S. 2021. MEGA11: Molecular evolutionary genetics analysis version 11. Mol. Biol. Evol. **38**(7): 3022-3027.

Tokhmechi, K., Abadkhah, M. and Koolivand, D. 2021. Emerging and population analysis of *Grapevine Pinot gris virus* isolates from Iran. 3 Biotech **11**: 368.

Wattoo, J.I., Saleem, M.Z., Shahzad, M.S., Arif, A., Hameed, A. and Saleem, M.A. 2016. DNA Barcoding: Amplification and sequence analysis of *rbcL* and *matK* genome regions in three divergent plant species. Adv. Life Sci. **4**(1): 03-07.

Yao, G., Zhang, Y.-Q., Barrett, C., Xue, B., Bellot, S., Baker, W.J. and Ge, X.-J. 2023. A plastid phylogenomic framework for the palm family (Arecaceae). BMC Biol. **21**: 50.

(Manuscript received on 7 August 2023; revised on 3 December 2023)