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

SHEIKH SUNZID AHMED, M. OLIUR RAHMAN¹, M. AJMAL ALI², FAHAD AL-HEMAID² AND JOONGKU LEE³

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

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

Abstract

A molecular phylogenetic investigation was undertaken for 30 species belonging to 15 genera of the palm family Arecales 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 monophly 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 Arecales or Palm 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 (Christenhuz 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,
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 \textit{matK} and \textit{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 \textit{matK} and \textit{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 \textit{et al.}, 2007). Based on availability, a total of 30 taxa belonging to 15 genera were chosen and both \textit{matK} and \textit{rbcL} gene sequences were downloaded in FASTA format. In addition, two species of Marantaceae, viz. \textit{Maranta arundinacea} L. and \textit{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 \textit{matK} and \textit{rbcL} sequences for these two species. The FASTA files were accumulated together using Notepad of Windows 10 to create two separate multifasta files for \textit{matK} and \textit{rbcL} sequences.

**Sequence alignment**

All the \textit{matK} and \textit{rbcL} sequences were subjected to Multiple Sequence Alignment (MSA) following sequence retrieval. Multifasta \textit{matK} and \textit{rbcL} files were uploaded to the MAFFT server (Katoh \textit{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 \textit{et al.}, 2021). Transition-transversion bias was calculated for \textit{matK} and \textit{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 \textit{matK} and \textit{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.

<table>
<thead>
<tr>
<th>No.</th>
<th>Taxon</th>
<th>Subfamily</th>
<th>matK Accession</th>
<th>rbcL Accession</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td><em>Areca catechu</em> L.</td>
<td>Arecoideae</td>
<td>KX783635.1</td>
<td>MK753924.1</td>
</tr>
<tr>
<td>2</td>
<td><em>A. triandra</em> Roxb. ex Buch.-Ham.</td>
<td>Arecoideae</td>
<td>MK705059.1</td>
<td>MK753941.1</td>
</tr>
<tr>
<td>3</td>
<td><em>Borassus flabellifer</em> L.</td>
<td>Coryphoideae</td>
<td>MK705088.1</td>
<td>MK753416.1</td>
</tr>
<tr>
<td>4</td>
<td><em>Calamus erectus</em> Roxb.</td>
<td>Calamoideae</td>
<td>JQ041985.1</td>
<td>MK754002.1</td>
</tr>
<tr>
<td>5</td>
<td><em>C. gracilis</em> Roxb.</td>
<td>Calamoideae</td>
<td>JQ041982.1</td>
<td>JQ042033.1</td>
</tr>
<tr>
<td>6</td>
<td><em>C. guruba</em> Buch.-Ham. ex Mart.</td>
<td>Calamoideae</td>
<td>JQ042013.1</td>
<td>JQ042064.1</td>
</tr>
<tr>
<td>7</td>
<td><em>C. longisetus</em> Griff.</td>
<td>Calamoideae</td>
<td>JX185542.1</td>
<td>JX185534.1</td>
</tr>
<tr>
<td>8</td>
<td><em>C. tenuis</em> Roxb.</td>
<td>Calamoideae</td>
<td>MK705230.1</td>
<td>JX502779.1</td>
</tr>
<tr>
<td>9</td>
<td><em>C. urens</em> L.</td>
<td>Coryphoideae</td>
<td>MK705128.1</td>
<td>JF344863.1</td>
</tr>
<tr>
<td>10</td>
<td><em>Caryota mitis</em> Lour.</td>
<td>Coryphoideae</td>
<td>MK705154.1</td>
<td>MK753393.1</td>
</tr>
<tr>
<td>11</td>
<td><em>Corypha umbraculifera</em> L.</td>
<td>Coryphoideae</td>
<td>MK705182.1</td>
<td>MK753724.1</td>
</tr>
<tr>
<td>12</td>
<td><em>Cocos nucifera</em> L.</td>
<td>Arecoideae</td>
<td>KX783653.1</td>
<td>MK753840.1</td>
</tr>
<tr>
<td>13</td>
<td><em>Dypsis lutescens</em> (H. Wendl.) Beentje &amp; J. Dransf.</td>
<td>Arecoideae</td>
<td>MK704536.1</td>
<td>MK753964.1</td>
</tr>
<tr>
<td>14</td>
<td><em>Licuala grandis</em> (T. Moore) H. Wendl.</td>
<td>Arecoideae</td>
<td>OL354144.1</td>
<td>OL536973.1</td>
</tr>
<tr>
<td>15</td>
<td><em>Livistona chinensis</em> (Jacq.) R. Br. ex Mart.</td>
<td>Coryphoideae</td>
<td>KX783705.1</td>
<td>GUI35214.1</td>
</tr>
<tr>
<td>16</td>
<td><em>Ptychosperma macarthurii</em> (H. Wendl. ex H. J. Veitch) H. Wendl. ex Hook. f.</td>
<td>Arecoideae</td>
<td>MK704980.1</td>
<td>MK753661.1</td>
</tr>
<tr>
<td>17</td>
<td><em>Rhapis excelsa</em> (Thunb.) A. Henry</td>
<td>Coryphoideae</td>
<td>KX783766.1</td>
<td>MK753583.1</td>
</tr>
<tr>
<td>18</td>
<td><em>R. humilis</em> Blume</td>
<td>Coryphoideae</td>
<td>MK704592.1</td>
<td>MK753575.1</td>
</tr>
<tr>
<td>19</td>
<td><em>Roystona oleracea</em> (Jacq.) O. F. Cook</td>
<td>Arecoideae</td>
<td>MK704872.1</td>
<td>MK753867.1</td>
</tr>
<tr>
<td>20</td>
<td><em>W. caryotoides</em> Roxb.</td>
<td>Coryphoideae</td>
<td>MK705141.1</td>
<td>MK753528.1</td>
</tr>
<tr>
<td>21</td>
<td><em>W. oblongifolia</em> Griff.</td>
<td>Coryphoideae</td>
<td>MK705143.1</td>
<td>MK753495.1</td>
</tr>
<tr>
<td>1</td>
<td><em>Maranta arundinacea</em> L.</td>
<td>Marantaceae</td>
<td>JQ588311.1</td>
<td>JQ592612.1</td>
</tr>
<tr>
<td>2</td>
<td><em>Marantochloa leucantha</em> (K. Schum.) Milne-Redh.</td>
<td>Marantaceae (family)</td>
<td>OL690061.1</td>
<td>OL536989.1</td>
</tr>
</tbody>
</table>
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, Corypheeae 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 excelsa*, *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 Corypheeae tribes were represented by a single species, namely *Borassus flabellifer* (Borasseae) and *Corypha umbraculifera* (Corypheeae). 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, Corypheeideae 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.

![Diagram of substitution sites across the matK region obtained from 30 species of Arecaceae using Marantochloa leucantha as reference](image)

**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 (*Roystoneae*) 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 Bootstrap 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).](image_url)
The ML tree of \textit{rbcL} taxa unraveled a tree topology very similar to that of the NJ-\textit{rbcL} tree (Fig. 6). In comparison to bootstrap support, the ML-\textit{rbcL} tree was found to have weaker support than the NJ-\textit{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. \textit{Phoenix} clade was the underlying reason for the polyphyletic nature of Coryphoideae. Due to \textit{Licuala grandis}, tribe Livistoneae became polyphyletic while tribe Caryoteae, Borasseae and Coryphea remained monophyletic. In Arecoideae, the tribes Areceae and Roystoneae 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 \textit{Calamus}.

Fig. 3. Neighbor Joining tree showing inter-relationships among three subfamilies of Arecaceae using 1000 bootstrap replicates based on \textit{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 Coryphaeae, Borasseeae 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.
MOLECULAR PHYLOGENETICS AND DATING OF ARECACEAE

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 Roystoneae 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.
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).

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 (Wurbm) Merr., Calamus latifolius Roxb., Corypha taliera Roxb., Daemonorops jenkinsiana (Griff.) Martius, Didymasperma gracilis Hook. f., D. nanum H. Wendl. & Drake, Licuala pelata Roxb., L. spinosa Jhun., Nypa fruiticans 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, Ceroxylloideae, 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 Phoenixaceae 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 Roystoneaceae 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 Dendrobiuim 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; Kakshi 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


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