Phylogenetic Analysis and Time Divergence of Genus *Musa* spp. Using S16 Protein Genes (Rps16) Intron Chloroplast

Farah Nahdia Noor *, Lailatus Sholichah

Biology Department, Science and Technology Faculty, Islamic State University Maulana Malik Ibrahim, Malang 65144, Indonesia

**ABSTRACT**

Family *Musaceae* is divided into three genera, *Musa*, *Musella*, and *Ensete*. The genus *Musa* became the genus with the most species cultivated in the Southeast Asian region. While the genus *Musella* and *Ensete* are monospecific genera that can only be found in certain regions. Distribution of banana nomenclature until now is still a debate. The taxonomic system of bananas is still based on conventional methods, namely by marking the genome. But unfortunately the nomenclature system is not suitable to be applied in the Southeast Asian region. Therefore be approached through molecular phylogenetic analysis used protein gen S16 (Rps16) chloroplast introns with Bayesian inference method and divergence time estimates. DNA sequences in alignment using MEGA 6.06 with the Clustal W method. Then highest priority density (HPD) is edited using the estimated distance of divergence time between species using BEAUti v1.8.4 software with the GTR (General Time Reversible) substitution model with the parameter G (Gamma) + I (Invariant Sites). The resulting phylogenetic tree is adjusted graphically in FigTree v1.4.3 software. The results show that the tree topology produced monophyletic lineage, in terms of the separation of the genus *Musa* from the *Musella lasiocarpa* and *Ensete ventricosum* Outgroups. Based on the results obtained, it can be concluded that phylogenetic *Musaceae* is based on the S16 protein gene (Rps16) Intron Chloroplast shows the results of a monophyletic tree topology, in terms of the separation of the genus *Musa* from the *Musella lasiocarpa* and *Ensete ventricosum* Outgroups.

**Keywords:** *Musa* spp, Phylogenetic, S16 Protein Genes, Chloroplast

**Introduction**

Bananas are one of the important plants of many cultivated plants. Banana is included in monocot that are widely cultivated in Southeast Asian region. Almost all cultivar of banana growing decay is thought to have existed and has been as cultivars in many regions [1]. Bananas grow in more than 120 countries in the world as well as being one of the advocates of food security for people in developing countries [2].

Bananas have three genera, namely *Musa*, *Ensete*, and *Musella*. These three genera of bananas come from the subclass *Commeliniidae*, Order *Zingiberales*, and include in Family *Musaceae* [3, 4, 2] the Southeast Asian region derived from the genus *Musa*. The genus *Musa* has areas including the plains of Sunda, the Philippines, the Wallace Region, and the Indo-Burma. While genus *Musella*, and *Ensete* are the monospecific genus whose spreading area includes South Sichuan, North Greece, Madagascar, Africa and Asia [4, 5].

The genus of *Musa* can be separated based on its morphological characteristic and genome characteristics [3]. Almost all consumed bananas today come from two species of wild banana *Musa acuminata* and *Musa balbisiana*. So that, the nomenclature of the genus *Musa* either from *Musa acuminata* and *Musa balbisiana* or the result of crossing two determined based on the genome.

**How to cite:**

of each species and the area of origin of bananas [6, 7]. From a taxonomic point of view, this raises problems in the nomenclature system of banana cultivars from the genus Musa and the grouping of banana species cultivars [4].

History of banana was first coined by Linnaeus in 1753. In his book Species Plantarum, the Origin of the Modern Botanical Nomenclature, Linnaeus have succeeded publish a group of bananas which contain lots of starch so they must be cooked before being consumed. This species called Musa paradisiaca Linn. Then in 1759, in his book Systema Naturae, Linneaus published Musa sapientum Linn., a group of desert banana that has a sweet taste and can directly eaten when it ripe. In Latin America and West Africa both species are more familiar with the term ‘banana’ and ‘plantain’. The nomenclature system initiated by Linnaeus lasted for almost two centuries [8].

But unfortunately, the nomenclature is not appropriate if applied in the Southeast Asian region. This is because there are too many banana cultivars scattered in the area that became the center of origin of Banana [9, 10] with the diversity of bananas they have in various regions. Therefore, in 1955 a change in the classification and nomenclature system of banana cultivars was proposed by Simmonds and Shepherd. Simmonds and Shepherd’s banana nomenclature system uses genome-based nomenclature and has been consensual determined for the nomenclature of banana cultivars in Southeast Asia in [11, 8].

Genome-based banana nomenclature system is based on the assumption that banana cultivars that exist today are descended from Musa acuminata and Musa balbisiana [10]. The cultivar bananas found today have a genomic composition derived from the two wild banana elders Musa acuminata with the genome symbol "A" and Musa balbisiana with the symbol "B". The crossing between the two wild banana elders will produce diploid (AB), triploid (AAB and ABB) and tetraploid hybrids (AAAB, AABB, and ABBB) [12] so that cultivar bananas have now evolved from banana first which generally has seeds into seedless bananas.

So far, to find out the nomenclature and grouping of bananas based on the genome, many use morphological approaches. Research using morphological (phenetic) characters on banana cultivars grouping in Indonesia has been widely done [13]. However, the results of these studies are often inaccurate because they are not explained in detail about the characteristics that can affect the grouping between varieties, are subjective, and can be influenced by environmental factors [14]. Therefore a molecular based approach is needed [15, 16] to obtain more valid results [14].

Approaches through molecular markers are believed to have a higher level of accuracy in banana cultivation grouping compared to morphological approaches [16, 17]. Some molecular markers that are often used in plant groupings include RAPD [18] RFLP [19, 20] and AFLP [21]. The molecular markers have been assessed, being able to distinguish genotypes among individual plants with a high degree of accuracy [22].

Some DNA sequences that are often used for identification of bananas are ITS [15-16], rbcl, trn-L [23] MatK [24] and atpB [14]. However, specific areas of ITS are specific DNA sequences that are most suitable for analysis of bananas, especially for genome identification or clarification, phylogenetic analysis, evolutionary history and biogeography [4].

Phylogenetic analysis in an effort to describe the kinship of a species is not only able to contribute to evaluating flora diversity, but also to develop efficient plant breeding systems. In plant breeding must have definitive identification of both selected plant cultivars [25]. For this reason, it is necessary to do an ilogenetic analysis of teh Musaceae family, especially the genus Musa to find out more about the genus of Musa ‘taxonomy system. With the current technological advancements, it has made it easy to sort the DNA, which later succeeded in revolutionizing the methodology of phylogenetic analysis. The most frequent target for this type of phylogenetic analysis is extra core DNA, mitochondria, Internal Transcribed Tracers (ITS), and chloroplast gene [26].

Mitochondria genes are derived from paternal genes, while chloroplast genes are derived from maternal genes in bananas [27, 28] so that these markers have the potential to analyze the origin of cultivars. In a previous study the RPSL6 intron protein gene was the ideal complement for the ITS study conducting inter/intrageneric phylogenetic analysis in the Sileneae, Anacardiaceae families [29, 30]. Chloroplast rps16 gene with a
sequence length of 790–887 bp produce similarity scores of 85–86% in analyzing the Poaceae, and Solanaceae families [30]. This instructs that this sequences is good for phylogenetic studies at the family level and below [30]. The phylogenetic analysis of the genus Musa will be very interesting to study, so the usage of the S16 Protein Genes (Rps16) intron chloroplast expected to provide phylogenetic information that received.

Ribosomal protein S16 is one of the proteins from the small ribosomal subunit. S16 proteins have about 100 amino-acid residues. There are two paralogues in Arabidopsis thaliana, RPS16-1 (chloroplastic) and RPS16-2 (targeted to the chloroplast and the mitochondrion) [31].

Molecular sequences have revolutionized phylogenetic analysis. In vascular plants, most are based on sequencing of molecular phylogenetic studies depending on the area of plastid genomic DNA, and on internal (or external) transcribed spacers (ITS/ETS) from 18S–5.8S–26S ribosomal cistron nuclear. The chloroplast use trnL–F, ndhF, rps16, ITS and ETS.

Phylogenetic analysis is not only able to contribute to evaluating flora diversity, but can also be used to evaluate geological history through the study of estimating the time of genetic divergence of an organism [32]. Phylogenetic analysis and estimated time for divergence in the Genus Musa are very interesting to study, so the use of the S16 (Rps16) Intron chloroplasts expected to provide accurate phylogenetic information. Therefore, the study entitled “Phylogenetic Analysis and Estimated Time of Divergence of genus Musa spp. Using S16 Protein Genes (Rps16) Intron Chloroplast” is important to do in evaluating the phylogenetic tree and knowing the estimated time divergence of the Genus Musa.

### Material and Methods

#### Data Sequences

Sequencing data collection is a sequence of Ensete (Ensete ventricosum) species and a species of Musella species (Musella lasiocarpa) as an out group. The species of Musa is in the group (Table 1). Sequence data is obtained from NCBI. The sequences used are DNA fragments. From the chloroplast genome which are inherited maternally, namely rps16 [27].

<table>
<thead>
<tr>
<th>No.</th>
<th>Species</th>
<th>GenBank accession number</th>
<th>Distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Musa salaccensis</td>
<td>FJ428112</td>
<td>Indo-Malaysian countries</td>
</tr>
<tr>
<td>2</td>
<td>Musa hirta</td>
<td>FJ428117</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Musa Monticola</td>
<td>FJ428119</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Musa Ingens</td>
<td>FJ428118</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Musa Textilis</td>
<td>FJ428121</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Musa Maclayi</td>
<td>FJ428122</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Musella lasiocarpa</td>
<td>FJ428123</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Ensete ventricosum</td>
<td>FJ428125</td>
<td>Asia</td>
</tr>
<tr>
<td>9</td>
<td>Musa Rosea</td>
<td>FJ428131</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Musa siamensis</td>
<td>FJ428134</td>
<td>Indo-Malaysian countries</td>
</tr>
<tr>
<td>11</td>
<td>Musa Velutina</td>
<td>FJ428141</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>Musa acuminata</td>
<td>FJ428140</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>Musa Yunnanensis</td>
<td>FJ428143</td>
<td>Indo-Malaysian and SE</td>
</tr>
<tr>
<td>14</td>
<td>Musa Nagensium</td>
<td>FJ428144</td>
<td>Asian countries to North</td>
</tr>
<tr>
<td>15</td>
<td>Musa Balbisiana</td>
<td>FJ428145</td>
<td>Australia</td>
</tr>
<tr>
<td>16</td>
<td>Moses basjoo</td>
<td>FJ428147</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>Moses exotica</td>
<td>FJ428151</td>
<td>Indo-Malaysian countries</td>
</tr>
</tbody>
</table>

DNA sequences in alignment using MEGA 6.06 with the Clustal W method. The gap at the beginning and end of the sequence was removed and a homologous sequence along 781 bp was
obtained. Then highest priority density (HPD) is edited using the estimated distance of divergence time between species using BEAUti v1.8.4 software with the GTR (General Time Reversible) substitution model with the parameter G (Gamma) + I (Invariant Sites). Calculation of MCMC (Monte Carlo Markov Chain) for every 10,000,000 generations, and sample trees every 1,000 generations and runs for 1,000,000 generations. The logfile output is checked with the Tracer software to confirm convergence. Tree maximum clade credibility and related posterior probabilities are calculated using Tree Annotator software [4]. The resulting phylogenetic tree is adjusted graphically in FigTree v1.4.3 software.

Results and Discussion

A Multiple Sequence Alignment (MSA) is a sequence of alignment of three or more biological sequences including proteins, DNA, or RNA. This is used to calculate the best match for the chosen sequence, and align it so that identity, similarities and differences can be seen. The input set of query sequences is assumed to have an evolutionary lineage relationship and is derived from the same ancestor. If a portion of the sequence is included in the dataset, the correlated region (suitable region) is used for phylogenetic analysis. There are several popular programs for MSA such as Clustal W [33].

Maximum likelihood and the Bayes method are the most widely used methods, because this method introduces an evolutionary model outlined based on the statistical method. The Bayesian method is a characteristic state method for inferring a phylogenetic tree based on posterior probabilities below the best predicted model [34].

The Bayesian method uses the concept of possibility and search for a reasonable set of trees. The Bayesian method requires prior distribution information on model parameters, such as substitution model parameters, branch length, and tree topology. Posterior probability obtained by exploring the tree space using a sampling technique, called the Markov Monte Carlo (MCMC) chain algorithm. The Bayesian approach has become popular because of advances in computing speed and is one of the most widely used methods now [35].

![Figure 3](Image)

Figure 3. Chronogram of the divergence time of the Genus Musa based on the S16 Protein Gen sequence (Rps16) Intron Chloroplast. The tree is reconstructed based on the uncorrelated log of the normal 'relaxed clock' and Bayesian Inference with 95% credible intervals.
The time difference for Musaceae was estimated at 51.9 Ma (61.2–45.6 Ma 95% of the highest priority density (HPD)), indicating the origin of the early Eocene. Likewise, the separation between Ensete and Musella twin genera was also located at the beginning of the Eocene at 44.7 (48.2–43.1 Ma 95% HPD; the calibration point used was 43 Ma). The initial radiation from Ensete occurred in the Oligocene at an estimated average age of 28.5 Ma (42.1-16.9 Ma 95% HPD). Diversification of Moses began at the end of the Eocene (average estimated age, 37.9 Ma; 50.5–24.5 Ma 95% HPD). Clade I in the genus Moses which contains diversified species of Ingentimuma, Australimusa and Callimusa in the Oligocene at 26.3 Ma (38.9–16.0 Ma 95% HPD), while Clade II (represented by species from the Moses and Rhodochlamys) began to emit 6 million years later in the early Miocene (estimated average age, 20.9 Ma; 30.4–13.3 Ma 95% HPD [12].

The taxonomic implications of phylogeny based on tree reconstruction results (Figure 3) confirm the Musaceae family in general, and genus Musa specifically becomes monophyletic. The genus Musella lasiocarpa appears as a sister species close to Ensete ventricosum. Phylogenetic relationships obtained in Musaceae are mostly in conformity with tree yields [21]. The Musaceae form a monophyletic clone in which genera currently recognized (Musa, Ensete and Musella) form a separate monophyletic lineage.

After the results of the phylogenetic tree, the tree should be evaluated in certain points. First, if two or more species are located separately, the tree does not represent the original phylogeny of the species. Second is evolutionary distance. The length of the tree branch reflects the evolutionary distance of each branch [34].

**Conclusion**

Based on the results obtained, it can be concluded that phylogenetic Musaceae is based on the S16 protein gene (Rps16) Intron Chloroplast shows the results of a monophyletic tree topology, in terms of the separation of the genus Musa from the Musella lasiocarpa and Ensete ventricosum Outgroups. The group of genus Musa separated and formed two clades. Estimated time of divergence shows the time of separation from Musa in the early Eocene. This also supports the hypothesis that the genus Musa originated from the same genus as the monsoon and ethsete, which later broke into several genera.

**References**