BIODIVERSITAS
Volume 23, Number 3, March 2022
Pages: 1374-1382

ISSN: 1412-033X
E-ISSN: 2085-4722
DOI: 10.13057/biodiv/d230323

Molecular identification of *Eurycoma longifolia* Jack from Sumatra, Indonesia using *trnL-F* region

KUSUMADEWI SRI YULITA[1], ARIDA SUSILOWATI[2,♥], HENTI HENDALASTUTI RACHMAT[3], SUSILA[1], ASEP HIDAYAT[3], FIFI GUS DWIYANTI[4]

[1]Research Centre for Biology, National Agency for Research and Innovation. Jl. Raya Jakarta – Bogor Km. 46, Bogor 16911, West Java, Indonesia
[2]Faculty of Forestry, Universitas Sumatera Utara. Jl. Padang Bulan, Medan 20155, North Sumatra, Indonesia. Tel.: +62-61-8220605, ♥email: arida.susilowati.usu.ac.id
[3]Forest Research and Development Center, The Ministry of Environment and Forestry. Jl. Gunung Batu No. 5, Bogor, 16118, West Java, Indonesia
[4]Departement of Silviculture, Faculty of Foresty, Institut Pertanian Bogor. Jl. Ulin, IPB Dramaga Campus, Bogor 16680, West Java, Indonesia

Manuscript received: 25 January 2022. Revision accepted: 20 February 2022

Abstract. *Yulita KS, Susilowati A, Rachmat HH, Susila, Hidayat A, Dwiyanti FG. 2022. Molecular identification of* Eurycoma longifolia *Jack from Sumatra, Indonesia using* trnL-F *region. Biodiversitas 23: 1374-1382. Eurycoma longifolia* Jack (Simaroubaceae) or *pasak bumi* is a popular medicinal plant from Southeast Asia's rainforests that is used as an aphrodisiac, antimicrobial, anti-malaria antidiabetic, antiulcer, and anticancer agent. However, the increasing demand for this species for medicinal industries has led to illegal export in Indonesia. This study aimed to determine the specific genetic variation and develop DNA barcode using *trnL-F* region for *E. longifolia* originating from Sumatra, Indonesia. Twenty-two samples of the species were collected from four locations in Sumatra. An aligned sequence of the *trnL-F* was 960 bp with an A/T rich region (A: 30.2%, T: 34.5%, C: 16.7%, and G: 18.7%). The homology search using BLASTn of the GenBank NCBI showed that the nucleotide composition of the species was similar (99.9%) to the partial *trnL-F* region of *E. longifolia* MH751519 and *E. apiculata* GU593014. Close examination of the gene structure and composition showed that the DNA sequences have five nucleotides variations that were not possessed by the reference *E. longifolia*, and other taxa used. The obtained variations occurred mostly in the *trnL* intron region, and the phylogenetic analysis showed that the correct identity of the species of the samples by their position was at a similar clade as the other accessions of *E. longifolia*.

Keywords: DNA barcode, *Eurycoma longifolia*, medicinal plant, Sumatra, *trnL-F*

INTRODUCTION

Eurycoma longifolia Jack (Simaroubaceae) is a herbal medicinal plant from the rainforests of Southeast Asia that is commonly known as 'tongkat ali' or 'pasak bumi' (Susilowati et al. 2019). The species is distributed in South Myanmar, Vietnam, Laos, Cambodia, Thailand, Peninsular Malaysia, Singapore, Sumatra, Borneo, the Philippines and may occur in Bangladesh based on herbarium records of GBIF (Global Biodiversity Information Facility 2020). The roots of this species contained several active compounds, such as canthin, b-carboline alkaloids, a derivate of squalene tirucallane triterpenes, biphenylneolignans, and quassinoids (Abubakar et al. 2017). These compounds contain antimicrobial (Kong et al. 2014), antimalarial (Low et al. 2013), antidiabetic, antiulcer, anticancer (Bhatt and Karim 2010; Nhan and Loc 2017; Ismail et al. 2012, and biological activities that are commonly used as aphrodisiac agent (Ismail et al. 2012; Chen et al. 2014). Some of the products derived from these plants are currently found in the market as raw materials or packaged herbal products such as coffee, tea, capsules, tablets (Rehman et al. 2016), and sweets (Edwar 2015).

The species has been traded intensively in the domestic and international markets, but not included within the CITES appendices (Sihotang and Rahmawati 2019). In combination with the increasing market demand and low prices from farmers, make this species become a target for illegal exports in Indonesia. The extraction of plants is usually carried out by the destructive harvesting of root extracts in the wild population (Susilowati et al. 2021). The sustainability use of the species cannot be guaranteed when this continues to happen. Therefore, it is important to develop a system that can be used for tracking and tracing the source plants, including *E. longilofia* such as DNA barcode, HPLC (Abubakar et al. 2018), and bar-high resolution melting analysis (Fadzil et al. 2018).

DNA barcode is a standard tool for species identification by using a fragment of the sequence of certain genes/regions (CBOL 2009). Initially, this term was used in 2003 (Herbert et al. 2003) and has gained global attention in the scientific community (Chen et al. 2010). Recently, the use of DNA barcodes covers a wide range of studies from the discovery of new species, discriminating cryptic species, population diversity, food safety, and conservation (Hollingsworth 2011; Hartvig et al. 2015; Imtiaz et al. 2017). The CBOL proposed portions of two coding regions from the plastid (chloroplast) genome, namely *rbcL* and *matK* as a "core barcode" for plants were supplemented with additional regions as required (Hollingsworth 2011; Daniel et al. 2016). In this study, *trnL-F* was used as a marker of choice for DNA barcodes for *E. longifolia* and the region was widely used for molecular systematics, such as *Mangifera* (Fitmawati et al.

2017), *Taxus* (Coughlan et al. 2020), and *Prunus* (Sevendik et al. 2020). This is because of its sufficient mutations to detect variations at specific and infraspecific levels (Hocaoglu-Ozyigit et al. 2020).

The *trnL-F* region is one of the recommended regions for DNA barcoding analysis (Ng et al. 2019; Hocaoglu-Ozygyt et al. 2020), since it meets all requirements identified by the CBOL. The loci were routinely retrievable with single primer pair, easy to obtain bidirectional sequence reads, and provided maximal discrimination among plant species (CBOL 2009). This study aimed to determine the specific genetic variation and develop DNA barcode using *trnL-F* region for *E. longifolia* originating from Sumatra, Indonesia.

MATERIALS AND METHODS

Plant materials

Twenty-four leaves samples were collected from Sumatra, Indonesia (Figure 1; Table 1) and dried in silica gel for DNA analysis. The identity was verified by collecting voucher herbarium specimens that were deposited at the University of Sumatra Utara, Medan and Forest Research and Development Center, Bogor (Indonesia). Meanwhile, the total genomic DNA was stored and the lab work was carried out at the Molecular Systematic Laboratoty, Research Center for Biology, The National Agency for Research and Innovation – Indonesia.

Isolation of DNA, amplification, and sequencing

The total DNA was isolated using Genomic DNA Mini Kit (Plant) from GeneAid following the manufacturer's protocol. Amplification of *trnL-F* region by the PCR technique was conducted using a universal pair primer of forward primer 'c' (CGAAATCGGTAGACGCTACG) and reverse primer 'f' (ATTTGAACTGGTGACACGAG) (Taberlet et al. 1991). The PCR mixture of a total volume of 12.5 µL consisted of 0.375 µL each of forward and reverse primer (5 pmol each), 0.25 µL Taq DNA polymerase, 2.5 µL of dNTP, 6.25 µL PCR Buffer, 1.75 µL of nuclease-free water, and 1 µL of DNA template (10 ng/µL). The reaction was performed in Sedi G Thermo Cycler (Wealtec) with the optimum condition of the following: a pre-denaturation at 94°C for 2 min, 30 cycles denaturation at 94°C for 30 s, annealing at 52°C for 30 s and extension at 72°C for 1 min, and a final extension at 72°C for 10 min.

The amplified bands were visualized on 1.5% agarose stained with GelRad. Meanwhile, electrophoresis was executed with 50 V for 60 min in 1x TBE buffer, and the target *trnL-F* bands were visualized under the UV light using Atto Bioinstrument. Sanger sequencing was used to sequence the amplicons at the First Base company, Singapore.

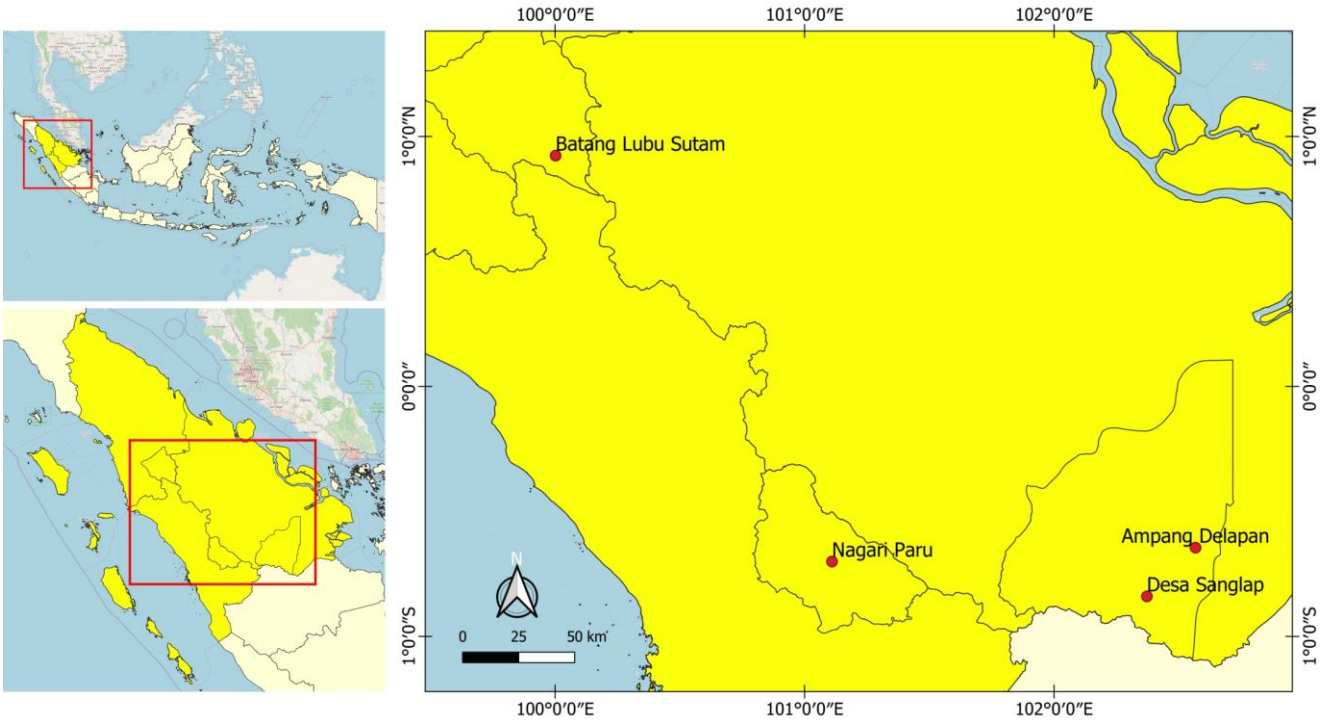

Figure 1. Source of *E. longifolia* from Sumatra, Indonesia

Table 1. Sources of *E. longifolia* from Sumatra, Indonesia and reference taxa were used for the phylogenetic analysis.

Species	Locality	Sample code (for this study), GenBank accession (for reference taxa)	Reference
Eurycoma longifolia Jack	Batang Lubu Sutam, Padang Lawas, North Sumatra	1PEL	This study
E. longifolia Jack	Batang Lubu Sutam, Padang Lawas, North Sumatra	2PEL	This study
E. longifolia Jack	Batang Lubu Sutam, Padang Lawas, North Sumatra	4PEL	This study
E. longifolia Jack	Batang Lubu Sutam, Padang Lawas, North Sumatra	5PEL	This study
E. longifolia Jack	Batang Lubu Sutam, Padang Lawas, North Sumatra	6PEL	This study
E. longifolia Jack	Batang Lubu Sutam, Padang Lawas, North Sumatra	7PEL	This study
E. longifolia Jack	Batang Lubu Sutam, Padang Lawas, North Sumatra	8PEL	This study
E. longifolia Jack	Batang Lubu Sutam, Padang Lawas, North Sumatra	9PEL	This study
E. longifolia Jack	Batang Lubu Sutam, Padang Lawas, North Sumatra	10PEL	This study
E. longifolia Jack	Sijunjung, Kanagarian Paru, West Sumatra	02EL	This study
E. longifolia Jack	Sijunjung, Kanagarian Paru, West Sumatra	03EL	This study
E. longifolia Jack	Sijunjung, Kanagarian Paru, West Sumatra	04EL	This study
E. longifolia Jack	Sijunjung, Kanagarian Paru, West Sumatra	06EL	This study
E. longifolia Jack	Sijunjung, Kanagarian Paru, West Sumatra	07EL	This study
E. longifolia Jack	Sijunjung, Kanagarian Paru, West Sumatra	08EL	This study
E. longifolia Jack	Sijunjung, Kanagarian Paru, West Sumatra	09EL	This study
E. longifolia Jack	Sijunjung, Kanagarian Paru, West Sumatra	10EL	This study
E. longifolia Jack	Sanglap, Indragiri Hulu, Riau	ES1EL	This study
E. longifolia Jack	Sanglap, Indragiri Hulu, Riau	ES2EL	This study
E. longifolia Jack	Sanglap, Inragiri Hulu, Riau	ES3EL	This study
E. longifolia Jack	Ampang delapan, Indragiri Hulu, Riau	A81EL	This study
E. longifolia Jack	Ampang delapan, Indragiri Hulu, Riau	A83EL	This study
E. longifolia Jack	Kuala Lumpur, Malaysia	MH751519	Ng et al. (2019)
E. apiculata A.W.Benn.	Unknown	GU593014	Clayton et al. (2010 unpubl.)
Simaba morettii Feuillet	French Guiana	MG599405	Devecchi et al. (2018)
Odyendea gabunensis (Pierre) Engl.	Gabon	MG599427	Devecchi et al. (2018)
Simaba glabra Engl.	Mato Grosso, Brazil	MG599404	Devecchi et al. (2018)
Perriera_madagascariensis Courchet	Unknown	GU593020	Clayton et al. (2010 unpubl.)
Simaba sp. nov. ined. 1"*arborea*"	Brazil, Amazonas	MG599416	Devecchi et al. (2018)
Perriera madagascariensis Courchet	Madagascar, Toliara	MG599408	Devecchi et al. (2018)
Simaba suffruticosa Engl.	Brazil, Minas Gerais	MG599407	Devecchi et al. (2018)
Simarouba glauca DC.	Belize, Cayo	MG599406	Devecchi et al. (2018)
Simaba sp. nov. ined. 5 "*pumila*"	Brazil, Goiás	MG599403	Devecchi et al. (2018)
Simarouba amara Aubl.	Bolivia, Larecaja	MG599401	Devecchi et al. (2018)
Simarouba versicolor A.St.-Hil.	Brazil, Goiás	MG599400	Devecchi et al. (2018)

Data analysis

Contig editor on ATGC software package version 4.3.5 was used to assemble the *trnL-F* sequence (Genetyx Co., Japan). The forward and reverse sequences were observed to ensure there was no mismatch in the consensus produced. Furthermore, MEGA 7.0 software was used to evaluate the nucleotide composition of the *trnL-F* gene (Kumar et al. 2016). The homology and identity of samples were examined by using BLAST nucleotide on GenBank (https://blast.ncbi.nlm.nih.gov/Blast.cgi) available in Geneious (Geneious version 2021). Meanwhile, the data from the GeneBank were downloaded in FASTA format and aligned using Geneious.

The evolutionary history was inferred by using the Maximum Likelihood method based on the Kimura 2-parameter model (K2P) (Kimura 1980; Kumar et al. 2016). Furthermore, the bootstrap consensus tree inferred from 1000 replicates (Felsenstein 1985) represented the evolutionary history of the taxa analyzed. The branches corresponding to partitions reproduced in less than 50% bootstrap replicates were collapsed. The percentage of the replicated trees in which the associated taxa clustered together (1000 replicates) was shown next to the branches. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances. They were estimated using the Maximum Composite Likelihood (MCL) approach, and their topology was selected using a superior log-likelihood value. Furthermore, a discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 1.7815). Bootstrap support (BS) values of ≥85%, 75-84%, and 50-74% were considered to be strongly, moderately, and weakly supported, respectively; and values <50% were not indicated (Devecchi et al. 2018).

Bayesian inference was performed using MrBayes 3.6 (Hueselbeck and Ronguist 2001) as plugins in Geneious version 2021.0, and the substitution model of HKY85 with rate variation of distributed gamma was applied for this analysis. Two independent MCMC runs of four chains for

1,100,000 generations and sampling tree topologies are for every 200 generations. Burn-in periods were set to 100,000 generations according to the standard deviation of split frequency values (<0.01). Meanwhile, a consensus topology and nodal support estimated as posterior probability values were calculated from the remaining trees. Posterior Probability (PP) values ≥0.95 and < 0.95 were considered as strongly and weakly supported (Devecchi et al. 2018).

RESULTS AND DISCUSSION

The sequence homology and identity of *E. longifolia*

The amplicon size of the *trnL-F* chloroplast gene from the 24 samples was 931 bp consisting of 439 bp *trnL* (its exons and majority of the intron) and 492 *trnF* genes with the intergenic spacer between *trnL* and *trnF*. From the 439 *trnL* region, 409 bp were homologue to many sequences of the GenBank, 18 of which were having similarities of more than 99% (data not shown). Furthermore, thirteen reference species -that will be used as outgroup taxa were incorporated into this study to build a phylogenetic tree based on Maximum Likelihood and Bayesian analyses (Table 2). This was aimed to assess the phylogenetic position of *E. longifolia* samples from Sumatra, therefore determining its identity.

Nucleotide composition and variations

The aligned DNA sequence of *trnL-F* is 960 bp. The *trnL-F* is an A/T rich region, composed of 34.5 % thymine (T), 16.7% cytosine (C), 30.2% adenine (A), and 18.7% guanine (G) (Table 3). The single-base substitution (point mutations) found in the 931 bp of the *trnL-F* region were 5, located at positions 209, 534, 822, 907, and 910, and one mutation at position 272, was found only in *E. longifolia* of the reference taxon (Table 3). Meanwhile, the first base mutation was observed in position 209, a trans-version of G --> C in five samples from West Sumatra. The second was a transversion from C --> G in position 272 as an autapomorph nucleotide for *E. longifolia* from Malaysia (MH751519), and the third point mutation was another transversion from C--> A observed in the same five samples at position 534. Furthermore, there was another point mutation in transition A --> G of position 822 belonging to samples from Riau. Another transition of G --> A at position 907 was recorded from samples Sanglap Riau (ES1, 2, 3EL, Table 3), while a base substitution (T) was also recorded in some of the outgroups observed. Finally, a transition from C--> T was observed in the intergenic spacer between *trnL* and *trnF* gene (position 910). These mutations were observed in the same three samples from West Sumatra and most samples from North Sumatra.

Phylogenetic tree

The results of the phylogenetic tree reconstruction using Maximum Likelihood showed that a monophyletic group with high (99%) bootstrap support (BS) was formed between the *E. longifolia* samples, the reference, and *E. apiculata* (Figure 2, clade A). Samples from Riau islands (F) formed a group together with low support from the bootstrap (66%) within the ingroup, while samples from West and North Sumatra were placed randomly and some were grouped in mixed groups (E) with 66% BS. Furthermore, five samples of West Sumatra were united in group G with 64% BS, and two from the North Sumatra populations remained unresolved (9PEL and 10PEL). This ingroup was similar to *Simarouba* spp. with high support from the bootstrap (B, 92%) (Figure 2). The remaining reference taxa considered as outgroup were located at the basal lineages, and they formed subgroup C with 91% BS and subgroup D with 87% BS or remained single with unresolved position (*Simaba* sp. nov. ined. 1"*arborea*" MG599416, *Simarouba glauca*, and *Perriera madagascarensis*).

A similar topology was obtained by Bayesian analysis, where the ingroup samples were *E. longifolia* from Sumatra, Kuala Lumpur (Malaysia), and *E. apiculata* (Figure 3). The phylogenetic position of *E. longifolia* from Kuala Lumpur (Malaysia), *E apiculata*, and 2 samples of *E. longifolia* from the North Sumatra remained unresolved. The related taxa of the ingroup were two species of *Simarouba* as shown in the ML tree. Primarily, the difference between the Bayesian and ML topology existed in the phylogenetic position of *Pierreria madagascarensis* placed at the basal lineages in the Bayesian topology, instead of *Odyendea gabunensis* (Figure 2). The remaining outgroup taxa were located at the basal clade C (Figure 3).

Discussion

Determining the sample size for DNA barcode studies helped to generate sufficient and consistent nucleotide variations that can be used as identities for certain species, such as Single Nucleotide Polymorphisms (SNPs) and haplotypes. Meanwhile, a small sample size leads to low information content due to the high presence of sequence artifacts, such as indels events (Liu et al. 2012; Jarret et al. 2019).

Table 2. Species accession number, pairwise identity, and identical sites of *E. longifolia* references

Accession number	Species	% pairwise identity	% identical sites
GU593014	*Eurycoma apiculata*	99.9	99.9
MG599427	*Odyendea gabunensis*	98.5	98.5
MG599405	*Simaba morettii*	99.1	99.1
MH751519	*Eurycoma longifolia*	99.9	99.9
GU593020	*Perriera madagascariensis*	98.7	98.7
MG599404	*Simaba glabra*	98.8	98.8
MG599416	*Simaba* sp. nov. ined. 1"arborea"	98.4	98.4
MG599408	*Perriera madagascariensis*	98.9	98.9
MG599407	*Simaba suffruticosa*	98.4	98.4
MG599406	*Simarouba glauca*	98.8	98.8
MG599403	Simaba sp. nov. ined. 5 "pumila"	98.7	98.7
MG599401	*Simarouba amara*	98.9	98.9

Table 3. Nucleotide composition and variation found in the *trnL-F* sequence of *E. longifolia* from the reference accessions

Species	Nucleotide content (%)				Position of point mutation					
	T(U)	C	A	G	209	272	534	822	907	910
Reference nucleotide					G	C	C	A	G	C
Eurycoma longifolia 02EL	34.6	16.6	30.1	18.7	.	.	.	.	.	T
E. longifolia 03EL	34.6	16.6	30.1	18.7	.	.	.	.	.	T
E. longifolia 04EL	34.6	16.6	30.1	18.7	.	.	.	.	.	T
E. longifolia 06EL	34.5	16.6	30.2	18.7	C	.	A	.	.	.
E. longifolia 07EL	34.5	16.6	30.2	18.7	.	.	A	.	.	.
E. longifolia 08EL	34.5	16.7	30.2	18.6	C	.	A	.	.	.
E. longifolia 09EL	34.5	16.7	30.2	18.6	C	.	A	.	.	.
E. longifolia 10EL	34.2	16.9	30.1	18.7	C	.	A	.	.	.
E. longifolia 1PEL	34.6	16.6	30.1	18.7	.	.	.	.	.	T
E. longifolia 2PEL	34.5	16.7	30.2	18.6	.	.	.	.	.	T
E. longifolia 4PEL	34.6	16.6	30.1	18.7	.	.	.	.	.	T
E. longifolia 5PEL	34.6	16.6	30.1	18.7	.	.	.	.	.	T
E. longifolia 6PEL	34.6	16.6	30.1	18.7	.	.	.	.	.	T
E. longifolia 7PEL	34.6	16.6	30.1	18.7	.	.	.	.	.	T
E. longifolia 8PEL	34.6	16.6	30.1	18.7	.	.	.	.	.	T
E. longifolia 9PEL	34.6	16.6	30.1	18.7	.	.	.	.	.	.
E. longifolia 10PEL	34.5	16.7	30.1	18.7	.	.	.	.	.	.
E. longifolia ES1EL	34.5	16.7	30.1	18.7	.	.	.	G	A	.
E. longifolia ES2EL	34.5	16.7	30.1	18.7	.	.	.	G	A	.
E. longifolia ES3EL	34.5	16.7	30.1	18.7	.	.	.	G	A	.
E. longifolia A81EL	34.5	16.7	30.0	18.8	.	.	.	G	.	.
E. longifolia A83EL	34.5	16.7	30.0	18.8	.	.	.	G	.	.
E. longifolia MH751519	34.3	16.7	30.1	18.8	.	G	.	.		.
E. apiculata GU593014 shrub or small tree	34.5	16.7	30.1	18.7	.	.	.	.	.	.
Simaba sp. nov. ined. 5 “pumila” MG599403 geophytic subshrub	34.5	16.6	30.4	18.5	.	.	.	.	.	.
Perriera madagascariensis GU593020 tree	34.6	16.5	30.3	18.7	.	.	.	.	.	.
P. madagascariensis MG599408 tree	34.7	16.6	30.1	18.6	.	.	.	.	.	.
Simarouba amara MG599401 tree	34.4	16.8	30.3	18.6	.	.	.	.	T	.
Simarouba glabra MG599404 shrub	34.5	16.6	30.4	18.5	.	.	.	.	T	.
Simaba glauca MG599406 tree	34.6	16.7	30.2	18.5	.	.	.	.	T	.
Simaba morettii MG599405 tree	34.3	16.6	30.3	18.7	.	.	.	.	T	.
Simaba sp. nov. ined. 1“arborea" MG599416 unknown	34.3	16.7	30.5	18.5	.	.	.	.	.	.
Simaouba versicolor MG599400 tree	34.4	16.7	30.2	18.7	.	.	.	.	.	.
Odyendea gabunensis GU593019 tree	34.1	16.5	30.1	19.2	.	.	.	.	.	.
Average	34.5	16.7	30.2	18.7						

Consequently, a lack of resolution was obtained to accurately identify certain specimens. It is a common consensus that 5-10 samples for each species were sufficient to conduct a DNA barcode study (Jarret et al. 2019). Liu et al. (2012) considered that relatively small sample sizes were adequate to recover sequence variation in slowly evolving genes (two or three sequences per species population for *matK*), while higher numbers are necessary for rapidly evolving markers (minimum of 10, 8, and 6 individuals per population for *trnH-psbA*, *trnL-F*, and ITS, respectively). A sample size of 8-10 individuals per species was further recommended in the geographic area to be sufficient for the *Taxus* barcode. In this current study, 22 samples of *E. longifolia* representing a geographic range in Sumatra were considered sufficient to produce consistent nucleotide polymorphism in the *trnL-F* region. Therefore nucleotide variation and composition from this study are sufficiently achieved. In addition, the composition of *E. longifolia* from this study is similar to those reference sequences (Table 2) as well as in *Prunus armeniaca* (Rosaceae) (Sevindik et al. 2020).

Sumatra shares many of its species with peninsular Malaysia and Borneo. The three landmasses were once part of a single, larger landmass during the last ice age when the sea level was more than 100 m lower than now. Consequently, these forests share many similarities in their flora and fauna. The geological similarity of the *E. longifolia* from North Sumatra and Peninsular Malaysia is due to their similar location. The North Sumatra basin is one of three back-arc basins formed during the Tertiary (Early Oligocene), on the Eurasian plate or the Sunda Shelf (Inger and Voris 2001; Rosid et al. 2020). The southwestern Sumatra is bordered by the Bukit Barisan Mountains uplifted in the Middle Miocene, in the southeast by the Asahan Arc. In the northeastern of Sumatra, it borders Peninsular Malaysia, and in the northern, the basin opens to the Andaman Sea. Hence the reference species (MH 751519) that is the genome sequence of *E. longifolia* from Kuala Lumpur Malaysia is placed in the same clade as the samples studied (Figures 2 and 3).

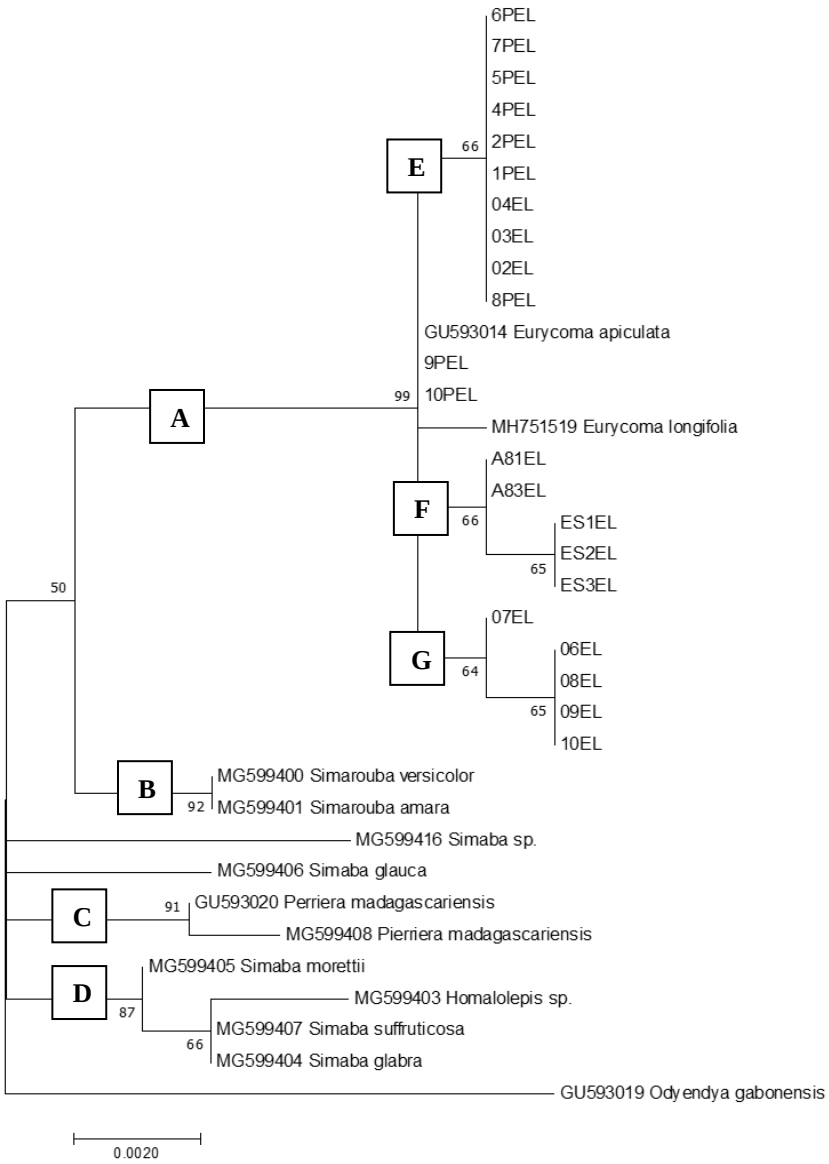

Figure 2. A phylogenetic tree generated by Maximum Likelihood method using 35 *trnL-F* sequences. The tree with the highest log likelihood (-1607.01) was shown, and branch supports were inferred using 1000 bootstrap replicates. The values are shown in the branches, and the tree is drawn to scale, with lengths measured in the number of substitutions per site

Samples sequence from North Sumatra did not have a single substitution, since the sequence compositions are similar to West Sumatran and Riau populations. This idea was confirmed by the unresolved phylogenetic position of the two North Sumatran samples (9PEL and 10 PEL) and reference samples from Kuala Lumpur, which were consistent in both topologies (Figure 2 and 3). In addition, of the five single base substitutions possessed by the samples from Sumatra, at least two base substitutions may be unique to Riau. The pattern seemed to be consistent in all studied samples even though this deduction was derived only from one region (*trnL-F*) particularly referring to the point mutations recorded only from Riau.

The two phylogenetic trees using Maximum Likelihood and Bayes resulted in a similar topology. Meanwhile, the *E. longifolia* from Indonesia, the reference species from Malaysia as well as *E. apiculata* formed a monophyletic group. This clade has received strong branches support in the ML and the Bayesian topology. Therefore, the identity of *E. longifolia* from Sumatra was correctly inferred, and the nesting of *E. apiculata* was interesting due to its high similarity to the region. The incorporation of more regions may provide different results. Meanwhile, no phylogenetic study has been conducted to highlight the genus *Eurycoma*. However, this study suggested that this genus is related to *Simarouba*. Another result has been reported from an earlier phylogenetic study of Simaroubaceae (with an emphasis on the genus *Simaba*) based on combined morphology and molecular biology datasets (Devecchi et al. 2018).

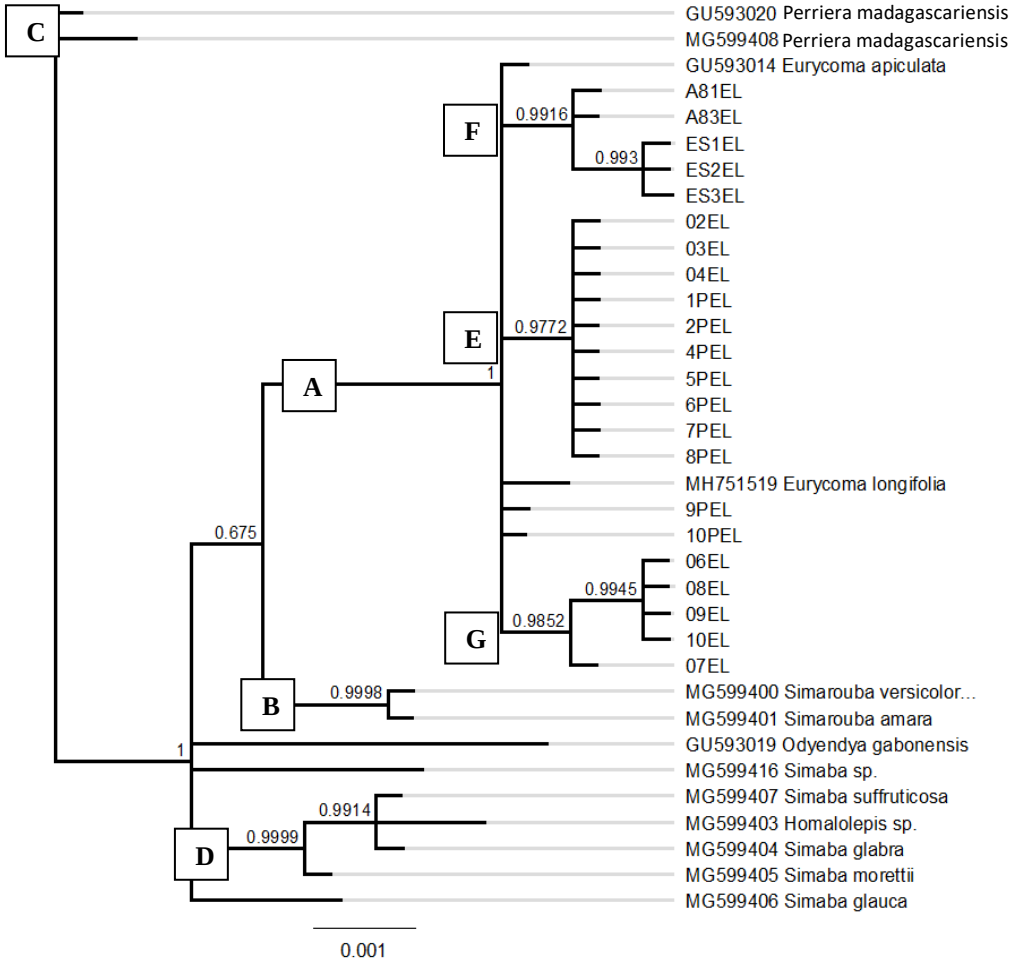

Figure 3. A phylogenetic tree generated by 35 *trnL-F* sequences of 12 species within family Simaroubaceae based on Bayesian inference (Figure created from Geneious version 2021.0 created by Biomatters. Available from: https://www.geneious.com). The number above branches are Posterior Probability (PP) values. Capital case in boxes refers to the same placement in node as in Figure 2

The result classified *E. longifolia* (sample was from Thailand), *Pierreodendron africanum*, and *Gymnostemon zaizou* into a similar group, and *Simarouba* was nested within the genus *Simaba*, which is a paraphyletic genus. Both *Pierreodendron africanum* and *Gymnostemon zaizou* are endemic genera to Africa and form a large tree compared to *E. longifolia* that resembles palm-like treelets. All the included outgroup (reference taxa) have different habitus as *E. longifolia*. *Eurycoma longifolia* has palm-like treelets, while the other species have tree, shrub, and geophytic sub-shrub (Devecchi et al. 2018) The phylogenetic placement of *E. longifolia* as a sister to both genera is due to the inclusion in a similar sample species. Therefore, a more detailed study on the phylogenetic position of *Eurycoma* spp. should be conducted to clarify the evolutionary relationships among the genera of the Simaroubaceae family, particularly within the paleotropic species.

In conclusion, a total of 931 bp *trnL-F* sequence used, five-point mutations were specifically discovered in the samples of *E. longifolia* from Sumatra. Four of the nucleotides were found in the *trnL* intron and one in the intergenic spacer between the *trnL* and *trnF* genes. Furthermore, two point mutations were observed from some and all samples from West Sumatra and Riau respectively. Therefore, the Riau clades were the only groups of *E. longifolia* that were resolved in the phylogenetic analysis. In addition, the *trnL-F* region can be used as one of the potential markers for establishing DNA barcodes for *E. longifolia*. Addition DNA barcode markers were recommended to be used with a similar mutation rate as *trnL-F* to complement these present results. The use of genomic data and the inclusion of more samples from Borneo should provide better resolution to the data set and establish DNA barcode for *E. longifolia* from Indonesia.

ACKNOWLEDGEMENTS

The authors are grateful to the Directorate of Research and Community Service, Ministry of Research and Technology and National Research and Innovation Agency, Republic of Indonesia (Number: 12/E1/KP.PTNBH/2021), through *Penelitian Dasar*

Unggulan Perguruan Tinggi (PDUPT) scheme for providing a research grant. The Research Centre for Biology of the National Agency for Research and Innovation, Indonesia is also appreciated for providing the facility in the Molecular Systematics Laboratory and the Forestry Research and Development Centre of the Ministry of Environment and Forestry for supporting the fieldwork. Contribution of the authors are as follow: KSY conducted laboratory work, data analysis, established the first draft, edit and write the manuscript; HHR collected samples, data analysis, established the first draft, edit and write manuscript; AS generated funding, collected samples, established the first draft, edit and write manuscript; AH collected samples, established the first draft, edit and write the manuscript; and FGD to established the first draft, edit and write the manuscript. The kind assistance of Mrs. Susila for the technical laboratory work is also acknowledged.

REFERENCES

Abubakar BM, Mohd SF, Omar MSS, Wagiran A. 2017. DNA barcoding and chromatography finger prints for the authentication of botanicals in herbal medicinal products. Evid Based Compl Altern Med 2017: 1352948. DOI: 10.1155/2017/1352948.

Abubakar BM, Salleh FM, Omar MSS, Wagiran A. 2018. Assessing product adulteration of *Eurycoma longifolia* (tongkat ali) herbal medicinal product using DNA barcoding and HPLC analysis. Pharm Biol 56 (1): 368-377. DOI: 10.1080/13880209.2018.1479869.

Bhat R, Karim A. 2010. Tongkat ali (*Eurycoma longifolia* Jack): A Review on its ethnobothany and pharmacological importance. Fitoterapia 81: 669-679. DOI: 10.1016/j.fitote.2010.04.006.

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

Chen S, Yao H, Han J, Liu C. 2010. Validation of the ITS2 region as a novel DNA barcode for identifying medicinal plant species. PLoS One 5: e008613. DOI: 10.1371/journal.pone.0008613.

Chen S, Pang X, Song J, Shi L, Yao H, Han J, Leon CA. 2014. Renaissance in herbal medicine identification: from morphology to DNA. Biotechnol Adv 32: 1237-1244. DOI: 10.1016/j.biotechadv.2014.07.004.

Clayton JW, Muellner AN, Pell S, Weeks A, Buerki S, Chiang YC, Cody S. 2010. Molecular phylogenetics of the eudicot order Sapindales based on plastid rbcL, atpB and trnL-trnF DNA sequences. Unpublished. https://www.ncbi.nlm.nih.gov/nuccore/GU593020.1

Coughlan P, Carolan JC, Hook ILI, Kilmartin L, Hodkinson TR. 2020. Phylogenetics of *Taxus* using the internal transcribed spacers of nuclear ribosomal DNA and plastid *trnL-F* Regions. Hortic 6: 19. DOI: 10.3390/horticulturae6010019.

Daniel H, Lin CS, Yu M, Chang WJ. 2016. Chloroplast genomes: Diversity, evolution, and applications in genetic engineering. Genome Biol 17: 134. DOI: 10.1186/s13059-016-1004-2.

Devecchi MF, Thomas WW, Plunkett GM, Pirani JR. 2018. Testing the monophyly of *Simaba* (Simaroubaceae): Evidence from five molecular regions and morphology. Mol Phylogenet Evol 120: 63-82. DOI: 10.1016/j.ympev.2017.11.024.

Edwar F. 2015. Candy and jelli as innovation products from pasak bumi (*Eurycoma longifolia* Jack). Jurnal Litbang Industri 5: 45-52. DOI: 10.24960/jli.v5i1.659.45-52. [Indonesian]

Fadzil NF, Wagiran A, Salleh FM, Abdullah S, Izham NHM. 2018. Authenticity testing and detection of *Eurycoma longifolia* in commercial herbal products using bar-high resolution melting analysis. Genes 9: 408. DOI: 10.3390/genes9080408.

Felsenstein J. 1985. Confidence limits on phylogenies: An approach using the bootstrap. Evolution 39: 783-791. DOI: 10.1111/j.1558-5646.1985.tb00420.x.

Fitmawati, Harahap SP, Sofiyanti. 2017. Phylogenetic analysis of mango (*Mangifera*) in northern Sumatra based on gene sequences of cpDNA trnL-F intergenic spacer. Biodiversitas 18 (2): 715-719. DOI: 10.13057/biodiv/d180239.

Geneious version 2021.0 created by Biomatters. Available from: https://www.geneious.com.

Global Biodiversity Information Facility. 2020. https://www.gbif.org/search?q=eurycoma%20longifolia [12 October 2020]

Hartvig I, Czako M, Kjær ED, Nielsen LR, Theilade I. 2015. The use of DNA barcoding in identification and conservation of rosewood (*Dalbergia* spp.). PLoS One 10: e0138231. DOI: 10.1371/journal.pone.0138231.

Herbert PD, Cywinska A, Ball SL, deWaard JR. 2003. Biological identifications through DNA barcodes. Proc Biol Sci 270: 313-321. DOI: 10.1098/rspb.2002.2218.

Hocaoglu-Ozyigit A, Ucar B, Altay V, Ozyigit II. 2020. Genetic diversity and phylogenetic analyses of turkish cotton (*Gossypium hirsutum* L.) lines using ISSR markers and chloroplast *trnL*-F regions. J Nat Fibers. DOI: 10.1080/15440478.2020.1788493.

Hollingsworth PM. 2011. Refining the DNA barcode for land plants. Proc Natl Acad Sci USA 108: 19451-19452. DOI: 10.1073/PNAS.1116812108.

Hueselbeck JP, Ronquist F. 2001. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics 17 (8): 754-755. DOI: 10.1093/bioinformatics/17.8.754.

Imtiaz A, Nor SAM, Naim DM. 2017. Review: Progress and potential of DNA barcoding for species identification of fish species. Biodiversitas 18: 1394-1405. DOI: 10.13057/biodiv/d180415.

Inger RF, Voris HK. 2001. The biogeographical relations of the frogs and snakes of Sundaland. J Biogeogr 28: 863-891. DOI: 10.1046/j.1365-2699.2001.00580.x.

Ismail SB, Mohammad W, Zahiruddin WM, George A, Nik HNH, Mustapha KZM, Liske. 2012. Randomized clinical trial on the use of physta freezedried water extract of *Eurycoma longifolia* for the improvement of quality of life and sexual well-being in men. Evid Based Complement Altern Med 2012: 429268. DOI: 10.1155/2012/429268.

Jarret D, Phillips, Daniel J, Gillis, Robert H, Hanner. 2019. Incomplete estimates of genetic diversity within species: Implications for DNA barcoding. Ecol Evol 9: 2996-3010. DOI: 10.1002/ece3.4757.

Kimura MA. 1980. Simple for estimating evolutionary rate of base substitutions through comparative Studies of nucleotide sequences. J Mol Evol 16: 111-120. DOI: 10.1007/BF01731581.

Kong C, Yehye WA, Rahmat NA, Tan MW, Nathan S. 2014. Discovery of potential anti-infectives against *Staphylococcus aureus* using a *Caenorhabditis elegans* infection model. BMC Compl Altern Med 14: 2-17. DOI: 10.1186/1472-6882-14-4.

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

Liu J, Provan J, Gao LM, Li DZ. 2012. Sampling strategy and potential utility of indels for DNA barcoding of closely related plant species: A case study in *Taxus*. Intl J Mol Sci 13: 8740-8751. DOI: 10.3390/ijms13078740.

Low BS, Das PK, Chan KL. 2013. Standardized quassionoid-rich *Eurycoma longifolia* extract improved spermatogenesis and fertility in male rats via the hypothalamicpituitary-gonadal axis. J Ethnopharmacol 149: 706-714. DOI: 10.1016/j.jep.2012.11.013.

Ng WL, Lee SY, Yeap SK. 2019. Characterization of the complete chloroplast genome of an important Southeast Asian medicinal plant, *Eurycoma longifolia* (Simaroubaceae). Mitochondrial DNA Part B 4: 128-129. DOI: 10.1080/23802359.2018.1540263.

Nhan NH, Loc NH. 2017. Production of *Eurycoma* from cell suspension culture of *Eurycoma longifolia*. Pharm Biol 55: 2234-2239. DOI: 10.1080/13880209.2017.1400077.

Rehman SU, Choe K, Yoo HH. 2016. Review on a traditional herbal medicine, *Eurycoma longifolia* Jack (tongkat ali): Its traditional uses, chemistry, evidence-based pharmacology and toxicology. Molecules 21 (3): 331. DOI: 10.3390/molecules21030331.

Rosid MS, Tullailah NN, Wibowo RA. 2020. Identification of potential hydrocarbon traps in the "X" area of pre-tertiary sediments of the North Sumatra basin using gravity and seismic data. Jurnal Fisika 10: 8-21. [Indonesian]

Sevindik E, Murathan ZT, Sevindik M. 2020. Molecular genetic diversity of *Prunus armeniaca* L. (Rosaceae) genotypes by RAPD, ISSR-PCR, and chloroplast DNA (cpDNA) *trnL*-F sequences. Intl J Fruit Sci 20 (S3): S1652-S1661 DOI: 10.1080/15538362.2020.1828223.

Sihotang VBL, Rahmawati K. 2019. The politics of stake collection and trading business of pasak bumi: The strategy of Alam Lestari Trading

Business and community of Pasak Bumi collectors in Bahorok District, North Sumatra. Proceedings of the National Seminar on Conservation and Utilization of Wild Plants and Animals: Research as the foundation for the conservation and sustainable use of wild plants and animals. Research Center for Biology, Indonesian Institute of Sciences, Bogor. [Indonesian].

Susilowati A, Rachmat HH, Elfiati D, Hasibuan MH. 2019. The composition and diversity of plant species in pasak bumi's (*Eurycoma longifolia*) habitat, Batang Lubu Sutam Forest North Sumatera. Biodiversitas 20 (2): 413-418. DOI: 10.13057/biodiv/d200215.

Susilowati A, Rachmat HH, Yulita KS, Elfiati D, Ginting IM. 2021. Distribution and association pattern of pasak bumi (*Eurycoma longifolia*) in Batang Lubu Sutam District, Padang Lawas, North Sumatra. 2021. IOP Conf. Series: Earth and Environmental Science. 912: 012056. DOI: 10.1088/1755-1315/912/1/012056.

Taberlet P, Gielly L, Patou G, Bouvet J. 1991. Universal primers for amplifification of three noncoding regions of chloroplast DNA. Plant Mol Biol 17: 1105-1109. DOI: 10.1007/BF00037152.