

## Short Communication: Weak delineation of *Styrax* species growing in North Sumatra, Indonesia by *matK* + *rbcL* gene

ARIDA SUSILOWATI<sup>1</sup>✉, HENTI HENDALASTUTI<sup>2</sup>, CUT RIZLANI KHOLIBRINA<sup>3</sup>, RINI RAMADHANI<sup>1</sup>

<sup>1</sup> Faculty of Forestry, Universitas Sumatera Utara. Jl. Tri Dharma Ujung No. 1, Kampus USU, Medan 20155, North Sumatra, Indonesia.

Tel./Fax.: + 62-61-8220605. ✉email: arida\_iswanto@yahoo.co.id

<sup>2</sup> Forest Research and Development Agency, Ministry of Environment and Forestry. Jl. Raya Gunung Batu. Bogor 16118, West Java. Indonesia

<sup>3</sup> Aek Nauli Forest Research Agency. Jl. Raya Parapat 10.5, Parapat, Simalungun 21174, North Sumatra, Indonesia

Manuscript received: 9 January 2017. Revision accepted: 31 July 2017.

**Abstract.** Susilowati A, Hendalastuti H, Kholibrina CR, Ramadhani R. 2017. Weak delineation of *Styrax* species growing in North Sumatra, Indonesia by *matK* + *rbcL* gene. *Biodiversitas* 18: 1270-1274. The aim of DNA barcoding is to enable precise identification of species from analysis of a unique DNA sequence of a target gene. The present study was undertaken to develop barcodes for different species of *Styrax* L. growing in North Sumatra, Indonesia. Four putative species of *Styrax* (each represented by specimens collected from two districts in North Sumatra) were evaluated using two regions in the plastid genome (*matK* and *rbcL*) in order to discriminate them at species level. Results showed that *matK* yielded 891 bp after alignment, however there was no precise identification to species level. The *matK* gene only differentiated the four recognized species into two groups, each group consisted of two of the species. The *rbcL* gene yielded 588 bp and showed no variation between species: that is, it determined all the currently recognized species to be of one and the same haplotype. Furthermore, combined *matK* + *rbcL* gave a similar grouping result to that for the *matK* gene considered alone. Considering the overall performance of these loci, we suggest *matK* + *rbcL* is not strong enough to determine *Styrax* growing in North Sumatra to the species level as distinguished on morphological grounds. These findings show the necessity of finding other candidate genes or markers that can potentially be helpful in delineating the various species of *Styrax* growing in North Sumatra, as well as other related *Styrax* genera.

**Keyword:** barcoding, *matK* + *rbcL* gene, *Styrax*, kemenyan

### INTRODUCTION

Species identification is of critical importance in conserving and utilizing biodiversity, but this is often hindered by a lack of professional knowledge of classification (Chase and Fay 2009). Tropical plant species are part of the world's flora that is especially threatened by human activities (Janzen 1988), and are in need of rapid identification methods to aid the development of reasonable protection strategies (Brooks et al. 2006). Taxonomy, using morphological classification methods, is a discipline that has been accepted for recognizing, naming, and classifying plants in order to avoid redundancy and ambiguity in the identification and naming of species. However, taxonomy is facing a crisis, in part because ever fewer specialists are available for the task. Fortunately, progress has recently been made in DNA-based technology that will facilitate the task in future and render its results more reliable.

The main aim of DNA barcoding is to establish a system that can be used for identification and taxonomic classification of biological organisms. The approach that has most success in animals is one based on using a portion of the *cytochrome oxidase 1* (CO1) mitochondrial gene, however establishing a standardized DNA barcoding system for plants has been more challenging

(Hollingsworth 2011). There are three important principles in DNA barcoding; namely, standardization, minimalism, and scalability. These principles imply a capability to choose one or few standard loci that can be sequenced routinely and reliably in very large and diverse sample sets, resulting in easily compared data that enable species to be distinguished from one another (Hollingsworth, 2011).

The standard animal CO1 gene fits all the criteria; however, finding similar universal markers for plants is more difficult and often requires multiple markers to obtain adequate species discrimination (Fazekas et al. 2008). Among the markers, there are seven plastid markers which, in various combinations, have been proposed for barcoding of plants: these are *rpoC1* + *rpoB* + *matK* or *rpoC1* + *matK* + *trnH-psbA* (Chase et al. 2007); *rbcL* + *trnH-psbA* (Kress and Erickson 2007) and *atpF-H* + *psbKI*. However, no agreement has yet been reached for elucidating the best combination, and none of the proposed barcodes is perfect in every respect (Pennisi 2007; Kress 2007; Erickson et al. 2008; Fazekas et al. 2008; Lahaye et al. 2008; Ledford 2008; Seberg and Peterson 2009).

To achieve a shared community resource for DNA barcoding of plants, it will be necessary to reach agreement on reliable, reproducible, barcoding procedures. Considering this situation, the CBOL Plant Working group has recommended a core-barcode consisting of a portion of two

plastid coding regions *rbcL* + *matK* to be supplemented with additional markers as required. The choice of *rbcL* + *matK* as a core barcode was based on the straight forward recovery of the *rbcL* region and the discriminatory power of the *matK* region. *matK* is known as one of the most rapidly evolving coding sections of the plastid genome (Hiluand Liang 1997), and is perhaps the closest plant analogue to the CO1 animal barcode.

In the study reported in this paper, we have attempted to apply the *rbcL* + *matK* barcoding approach to resolve the taxonomic status of some putative species of the genus *Styrax* L. that grow in North Sumatra, Indonesia. *Styrax* L. is a genus of about 130 species of large shrubs or small trees in the family Styracaceae. The genus *Styrax* is known around the world as the source of the commercial resin benzoin and is widespread but disjunct in its distribution, which includes the Americas, eastern Asia and the Mediterranean region (Fritsch 2001). Different species of *Styrax* can be found in some Southeast Asian countries such as Indonesia, Laos, Vietnam and Thailand. Within Indonesia, the province of North Sumatra has been reported as the largest distribution site for *Styrax*. Although some experts suggest that this species can also be found in South Sumatra and Papua, an accurate study confirming the extent of its Indonesian distribution has not yet been done.

*Styrax* species, known as ‘kemenyan’ in Indonesia, naturally grow along the western coast of the Sumatra mainland (Burkill 1953). Based on our primary data obtained during several field trips, kemenyan trees are economically important and have been massively cultivated for their resin for centuries in several sites of North Sumatra Province, on both private and community lands. The local community of North Sumatra recognize *Styrax* species under five local names based on morphological characters, namely kemenyan Toba, kemenyan durame, kemenyan minyak, kemenyan bulu, and kemenyan Batak. The species names that have been attributed to these (see for example Burkill 1935) are as follows: for kemenyan Toba, *Styrax sumatranus* J.J. Sm. (unresolved name); for kemenyan durame, *S. benzoin* Dryand. (unresolved name); for kemenyan bulu, *S. hiliferum* var. *hiliferum* (unresolved name); for kemenyan minyak, *S. paralleloneurum* Perkins (unresolved name); and for kemenyan Batak, *S. tonkinensis* Craib ex Hartwich (accepted name).

DNA barcoding is known to be useful to manage plant diversity inventories on a large scale, and to develop conservation strategies. This is indeed the case for

managing conservation and production strategy of species such as those of *Styrax* L. in Indonesia. However, there have been no DNA barcode studies from the tropical areas of Sumatra, Indonesia, which represent one of the biodiversity hotspots of the world. This study was conducted to determine whether the combined *matK* + *rbcL* can be effectively used for species delineation of the *Styrax* growing in North Sumatera, Indonesia.

## MATERIALS AND METHODS

### Study area

Samples of kemenyan tree originated from different localities including Humbang Hasundutan, Pakpak Bharat and North Tapanuli districts of North Sumatra Province, Indonesia (Figure 1). We chose Humbang Hasundutan, Pakpak Bharat and North Tapanuli because they represent different geological areas.

### Procedure

#### Plant materials

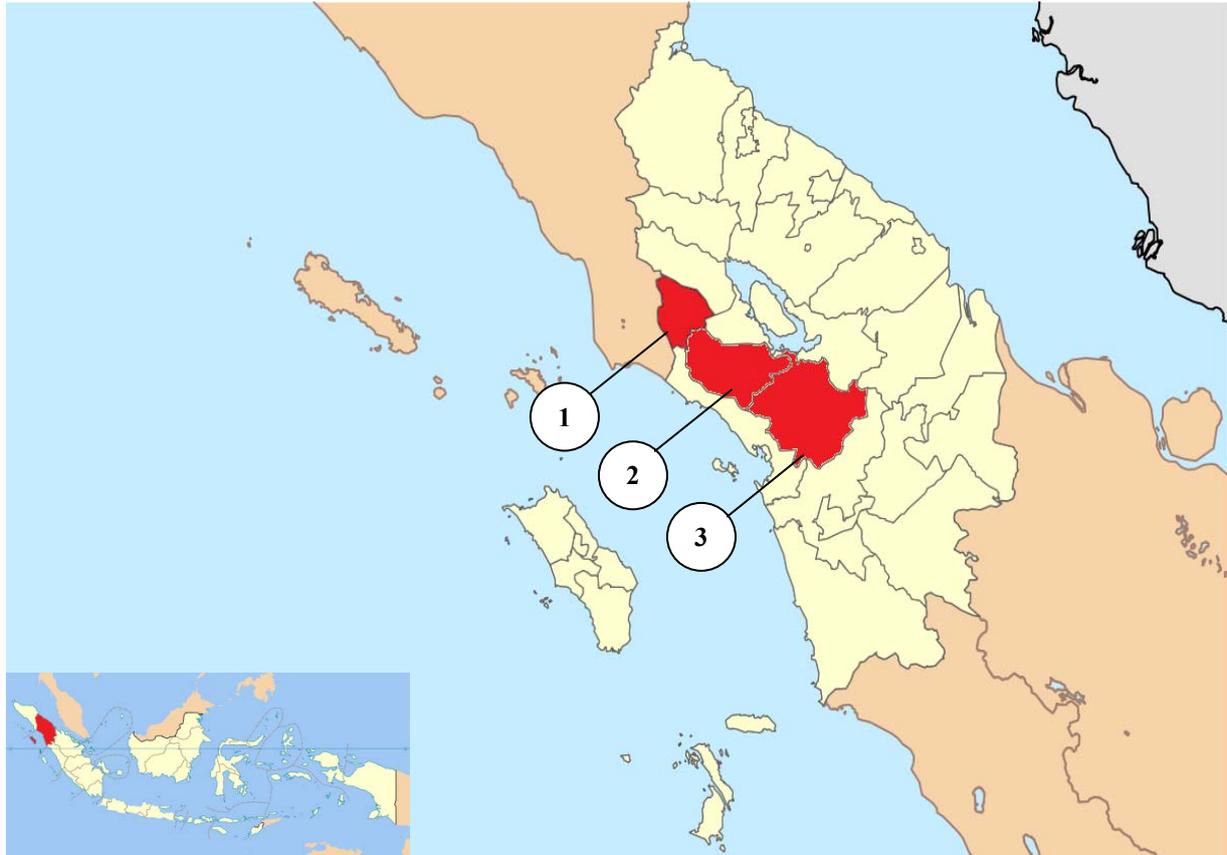
Four putative species of *Styrax* were analyzed (Table 1) where each species was represented by 2 individuals except for kemenyan Batak. Sampled leaves were dried with silica gel and used for DNA extraction. Total genomic DNA was extracted from leaf samples using a modified CTAB method (Murray and Thompson 1980).

#### DNA sequencing of *matK* and *rbcL* gene

*matK* gene was amplified by PCR using the *matK*\_390f dan *matK*\_132r as described by Cuenoud et al. (2002) while *rbcL* was amplified by primers described by Levin (2003) and Kress and Erickson (2007). PCR was performed in 20 µl of solution containing 10 ng of genomic DNA, 5 pmol of each forward and backward primer, and 10 µl of Go Taq® Hot Start Colourless Master Mix (Promega, Wisconsin, USA) according to the manufacturer’s instructions. Initial denaturation was performed at 95°C for 2 min, followed by 30-35 cycles of denaturation at 95°C for 1 min, annealing at 52°C and polymerization at 72°C for 2 min, and final extension at 72°C for 7 min. Prior to sequencing, the PCR products were purified. Automated sequencing was by Genetic Science (Singapore). DNA sequencing was performed for both strands with the primers used for the PCR amplifications.

**Table 1.** Locations of *Styrax* populations sampled from North Sumatra, Indonesia

Local name	Origin	GPS position
Kemenyan durame	North Tapanuli	N: 01° 55' 46.7", E: 098° 58' 42.9"
	Humbang Hasundutan	N: 02° 11' 57.0", E: 098° 44' 04.3"
Kemenyan minyak	Pakpak Barat 01	N: 02° 30' 30.5", E: 098° 24' 21.5"
	Pakpak Barat 02	N: 02° 30' 29.0", E: 098° 24' 25.0"
Kemenyan Toba	Pakpak Bharat	N: 02° 29' 48.4", E: 098° 25' 10.2"
	North Tapanuli	N: 01° 53' 45.9", E: 098° 58' 39.0"
Kemenyan Batak	Pakpak Bharat	N: 02° 30' 31.0", E: 098° 24' 21.7"



**Figure 1.** Study site in North Sumatra Province, Indonesia. Note: 1. Pakpak Barat, 2. Humbang Hasundutan, 3. North Tapanuli

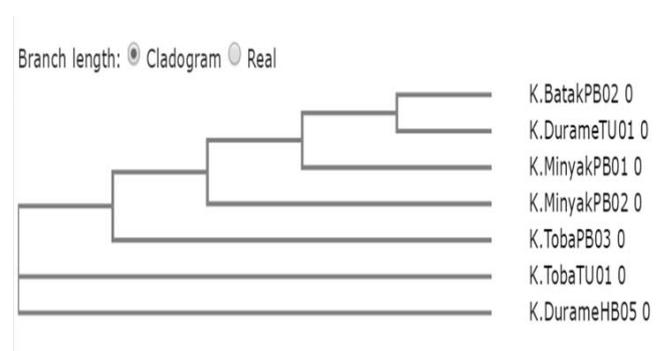
### Data analysis

DNA sequences were checked visually, and forward and backward traces were assembled using the ATGC program (Genetyx Corporation, Japan). Assembled sequences were exported into Bioedit for alignment to determine haplotypes. Further analysis using MUSCLE software was executed for constructing a dendrogram (Edgar 2004).

## RESULTS AND DISCUSSION

From our primary field trip observations and surveys, the identification of Kemenyan trees in North Sumatra was carried out according to morphological characteristics, based on local community knowledge. There are four different species groupings recognized by the communities. They assume that each grouping represents a different biological species. The four groupings are named kemenyan Toba, kemenyan minyak, kemenyan Batak, and kemenyan durame, identified by community knowledge as *S. sumatranus* J.J.Sm., *S. paralleloneurum* Perkins, *S. tonkinensis* Craib ex Hartwich and *S. benzoin* Dryand, respectively. In this study, we sought to reveal whether these morphologically based species identifications could be justified by *matK* and *rbcL* genetic analysis.

The direct nucleotide sequencing yielded 891 bp for *matK* and 588 bp for *rbcL* gene (see Figure S1-S3). *matK* separated the seven samples representing the four putative species into 2 haplotypes: Haplotype I consisted of a sequence common to individual samples from kemenyan minyak, kemenyan durame and kemenyan Toba; while Haplotype II consisted of a sequence common to individual samples from kemenyan minyak, kemenyan durame and kemenyan Batak.



**Figure 2.** Grouping of the seven samples representing four putativekemenyan species, assigned by combined *matK* + *rbcL* gene

There was no variation observed among individual samples sequenced for the *rbcL* gene: all individuals were assigned to just one haplotype. As a consequence of the non-informative *rbcL* gene used in this study, species delineation could only be determined by the *matK* gene (Figure 2). The results also indicated that the use of combined *matK* + *rbcL* sequences for attempted DNA barcoding of the *Styrax* species is too weak to delineate the North Sumatran populations to species level. Species discrimination with plant barcodes is typically lower than with CO1 in animals. As reported previously (Pennisi 2007; Kress 2007; Erickson et al. 2008; Fazekas et al. 2008; Lahaye et al. 2008; Ledford 2008; Seberg and Peterson 2009), delineating higher plant species by using DNA based methods often requires a multi-genic approach for which there is as yet no universal consensus. In our case, the combination of *matK* + *rbcL* was not successful in clearly differentiating kemenyan species. The results showed that the four different types of Kemenyan considered as separate species on the basis of local community knowledge, were not assigned into separate clusters by our genetic analysis. In situations like this, where barcoding does not provide a unique species identification, it instead identifies 'species groups' (typically, local groups of closely related congeners).

The addition of more extensive samples for each of putative Kemenyan species might produce different results as it may generate unique DNA stretches that can be used as species identifiers. According to the CBOL Plant Working Group (2009) standards, an ideal barcode should be a short segment that allows easy amplification, be amenable to bi-directional sequencing with little requirement for manual editing of sequence traces, and possess sufficient variation among sequences to distinguish between species. For bisexual species, such as Dipterocarps, chloroplast DNA markers are the best choice for identification. However, our study has demonstrated that use of *matK* and *rbcL* is inadequate as a barcode for discriminating between the kemenyan populations we sampled from North Sumatra. There is the need to add more candidates for DNA barcoding of *Styrax* spp. Genomic regions such as *trnH-psbA*, *trnL-trnF* and others could be worth considering as candidate genes for *Styrax* barcoding purposes.

### Acknowledgments

We are grateful to Prof. Iskandar Z. Siregar for letting us use his laboratory for DNA extraction and PCR amplification, and to Dr. Fifi Gus Dwiyanti for generous help in accommodating our laboratory work. Funding was provided by BOPTN USU 2016.

### REFERENCES

- Brooks TM, Mittermeier RA, da Fonseca GA, Gerlach J, Hoffmann M, Lamoreux JF, Mittermeier CG, Pilgrim JD, Rodrigues ASL. 2006. Global biodiversity conservation priorities. *Science* 313: 58-61.
- Burkill, I.H. 1935. A Dictionary of the Economic Products of the Malay Peninsula (Vol II (1-Z)). Published on behalf of the Governments of the Strait Settlements and Federated Malay States, London.
- CBOL Plant Working Group. 2009. A DNA barcode for land plants. *Proc Natl Acad Sci USA* 106 (31): 12794-12797.
- Cuenoud P, Savolainen V, Chatrou LW, Powell M, Grayer RJ, Chase MW. 2002. Molecular phylogenetics of Caryophyllales based on nuclear *18S rDNA* and plastid *rbcL*, *atpB*, and *matK* DNA sequences. *Amer J Bot* 89: 132-144.
- Chase MW, Cowan RS, Hollingsworth PM, van den Berg C, Madrinan S, Petersen J, Seberg O, Jørgensen T, Cameron KM, Carine M, Pedersen N, Hedderson TAJ, Conrad F, Salazar GA, Richardson JA, Hollingsworth ML, Barraclough TG, Kelly L, Wilkinson M. 2007. A proposal for a standardized protocol to barcode all land plants. *Taxon* 56: 295-299.
- Chase MW, Fay MF. 2009. Barcoding of plants and fungi. *Science* 325: 682-683. DOI: 10.1126/science.1176906
- EdgarRC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32 (5): 1792-97.
- Erickson DL, Spouge J, Resch A, Weigt LA, Kress JW. 2008. DNA barcoding in land plants: developing standards to quantify and maximize success. *Taxon* 57: 1304-1316.
- Fazekas AJ, Burgess KS, Kesanakurti PR, Graham SW, Newmaster SG, Huban BC, Percy DM, Hajibabaei M, Barret SCH. 2008. Multiple multilocus DNA barcodes from the plastid genome discriminate plant species equally well. *PLoS ONE* 3 (7): e2802. DOI: 10.1371/journal.pone.0002802
- Fritsch PW. 2001. Phylogeny and biogeography of the flowering plant genus *Styrax* (Styracaceae) based on chloroplast DNA restriction sites and DNA sequences of the internal transcribed spacer region. *Mol Phylogenet Evol* 19 (3): 387-408.
- Hilu KW, Liang H. 1997. The *matK* gene: sequence variation and application in plant systematics. *Amer J Bot* 84: 830-839.
- Hollingsworth ML, Clark AA, Forrest LL, Richardson J, Pennington RT, Long DG, Cowan R, Chase MW, Gaudeul M, Hollingsworth PM. 2009. Selecting barcoding loci for plants: Evaluation of seven candidate loci with species-level sampling in three divergent groups of land plants. *Mol Ecol Res* 9: 439-457.
- Hollingsworth PM, Sean WG, Damon PL. 2011. Choosing and using a plant DNA barcode. *PlosONE* 6 (5): e19254. DOI: 10.1371/journal.pone.0019254
- Janzen DH. 1988. Tropical ecological and biocultural restoration. *Science* 239: 243-244.
- Kress WJ, Erickson DL. 2007. A two-locus global DNA barcode for land plants: The coding *rbcL* gene complements the non-coding *trnH-psbA* spacer region. *PLoS ONE* 2 (6): e508. DOI: 10.1371/journal.pone.0000508.
- Lahaye R, Van der Bank M, Bogarin D, Warner J, Pupulin F, Gigot G, Maurin O, Duthoit S, Barraclough TG, Savolainen V. 2008. DNA barcoding the floras of biodiversity hotspots. *Proc Natl Acad Sci USA* 105: 2923-2928.
- Ledford H. 2008. Botanical identities: DNA barcoding for plants comes a step closer. *Nature* 451: 616.
- Levin RA, Wagner WL, Hoch PC, Nepokroeff M, Pires JC, Zimmer EA, Sytoma KJ. 2003. Family-level relationships of Onagraceae based on chloroplast *rbcL* and *ndhF* data. *Amer J Bot* 90: 107-115
- Murray HG, Thompson WF. 1980. Rapid isolation of high molecular weight DNA. *Nucleic Acids Res* 8: 4321-4325.
- Pennisi E (2007) Taxonomy. Wanted: A barcode for plants. *Science* 318: 190-191.
- Seberg O, Petersen G. 2009. How many loci does it take to DNA barcode a crocus? *PLoS ONE* 4 (2): e4598. DOI: 10.1371/journal.pone.0004598.

**Figure S1.** Nucleotide sequence of *matK* gene for kemenyan Batak PB02, kemenyan minyak PB01 and kemenyan durame TU01

ATATTTCCTTTTTTAGAAAGAAAAATTTTCATATTTAAATTATGTGTTAGATATACGAATACCCCACTCCATTCATCTGGAAG  
 TTTTGGTTCAAACCTCTTCGCTACTGGGTAAAAGATGCCTCTTCTTTGCATTTATTACGATTCTTTCTCTACGAGCATTGTAA  
 TTGGAATAATCTTACTACTCCAAAGAAATCTAGTTTTTTCTTTTTCAAAAAGAAATCAAAGATTATTATTGTTTCTATATAAT  
 TCTCATGTATATGAATACGAATCTATCTTTGTTTTTCTTCGAAACCAATCTTCTCATTACGATCAATATCTTCTGGAACCC  
 TTCTTGAACGAATATATTTCTATGGAAAAATAGACCATTTTGGAGAAGTCTTTACTAAGGATTTTCAGGCCGTTCTATGGTT  
 GTTCAAGGATCTTTTCATGCATTATGTTAGGTATCAAGGAAAATCCATTTTGGCTTCAAAGGGACGTTTTTTTTTAATGAAT  
 AAATGGAAATATTACCTTGTCAATTTCTGGCAATGTTATTTTTACAAGTGGTCTCAACCAGGAAGGATCCATATACAATTAT  
 CCAATCATTCTCTCGACTTTCTGGGCTATCTTTCAAATGTGCGCCTAAACCCTTCAATGGTCCGGAGTCAAATGCTAGAAAA  
 TGCATTTTTTAATAGGTAATGCTATTAAGAAGTTTGATACCCTAGTTCTAATTAGAGGAATAATTGGATCATTGTATAAAAA  
 AAAATTTGTAACGCGTTAGGACATCCCATTAGTAAGCCAGCCTGGGCCGATTTATCAGATTCTGATATTATTGATCGATTCC  
 GGTGTATATATAGAAATCTTTCTCATTATCATAGCGGATCCTCAAAAAAATGAGTTTGTATCGAATAAAG

**Figure S2.** Nucleotide sequence of *matK* gene for kemenyan durame HB05, kemenyan minyak PB02, and kemenyan Toba PB03 and TU01

ATATTTCCTTTTTTAGAAAGAAAAATTTTCATATTTAAATTATGTGTTAGATATACGAATACCCCACTCCATTCATCTGGAAG  
 TTTTGGTTCAAACCTCTTCGCTACTGGGTAAAAGATGCCTCTTCTTTGCATTTATTACGATTCTTTCTCTACGAGCATTGTAA  
 TTGGAATAATCTTACTACTCCAAAGAAATCGAGTTTTTTCTTTTTCAAAAAGAAATCAAAGATTATTCTTGTTCCTATATAAT  
 TCTCATGTATATGAATACGAATCTATCTTTGTTTTTCTTCGAAACCAATCTTCTCATTACGATCAATATCTTCTGGAACCC  
 TTCTTGAACGAATATATTTCTATGGAAAAATAGACCATTTTGGAGAAGTCTTTACTAAGGATTTTCAGGCCGTTCTATGGTT  
 GTTCAAGGATCTTTTCATGCATTATGTTAGGTATCAAGGAAAATCCATTTTGGCTTCAAAGGGACGTTTTTTTTTAATGAAT  
 AAATGGAAATATTACCTTGTCAATTTCTGGCAATGTTATTTTTACAAGTGGTCTCAACCAGGAAGGATCCATATAAACCAAT  
 TATCCAATCATTCTCTCGACTTTCTGGGCTATCTTTCAAATGTGCGCCTAAACCCTTCAATGGTCCGGAGTCAAATGCTAGA  
 AAATGCATTTTTTAATAGGTAATGCTATTAAGAAGTTTGATACCCTAGTTCTAATTAGAGGAATAATTGGATCATTGTATAAAA  
 AAAAAAATTTGTAACGCGTTAGGACATCCCATTAGTAAGCCAGCCTGGGCCGATTTATCAGATTCTGATATTATTGATCGAT  
 TCGGGTGTATATATAGAAATCTTTCTCATTATCATAGCGGATCCTCAAAAAAATGAGTTTGTATCGAATAAAGTATATACT  
 TCACTT

**Figure S3.** Nucleotide sequence of *rbcl* gene for all kemenyan assessed in this study

CAGAGACTAAAGCAAGTGTGGATTCAAAGCTGGTGTAAAGATTACAAATTGACTTATTATACTCCTGAATATGAAACCAA  
 AGATACTGATATCTTGGCAGCATTTCGAGTAACGCCTCAACCTGGAGTTCGCGCTGAAGAAGCGGGGGCCGCGGTAGCTGCC  
 GAATCTTCTACTGGTACATGGACAACGTGTGTGGACCGATGGACTTACTAGCCTTGATCGTTACAAAGGGAGATGCTACCACA  
 TTGAGCCCGTTGCTGGAGAAGAAAATCAATATATTGCTTATGTAGCTTATCCTTTAGACCTTTTTGAAGAAGGTTCTGTTAC  
 TAACATGTTTACTTCCATTGTTCGGTAATGTATTTGGGTTCAAAGCTCTTCGCGCTCTACGTCTGGAAGATCTGCGAATTCT  
 CCTGCGTATGTTAAAACCTTTCCAAGGACCGCCTCATGGCATCCAAGTTGAAAGAGATAAATTAATAAGTACGGTTCGTCCTCC  
 TGTGGGATGTACTATTAACCTAAATTTGGGGTTATCTGCTAAAAACTACGGTAGAGCAGTTTATGAATGTCTTCGCGGTGG  
 ACTTGATTTTTACAA