HAYATI Journal of Biosciences January 2015
Vol. 22 No. 1, p 41-47
EISSN: 2086-4094

Available online at:
http://journal.ipb.ac.id/index.php/hayati
DOI: 10.4308/hjb.22.1.41

SHORT COMMUNICATION

DNA Barcoding of Sangihe Nutmeg (*Myristica fragrans*) using *mat*K Gene

TRINA EKAWATI TALLEI*, BEIVY JONATHAN KOLONDAM

Department of Biology, Faculty of Mathematics and Natural Sciences, University of Sam Ratulangi, Kampus UNSRAT Bahu, Manado 95115, Indonesia

Received September 3, 2014/Accepted December 22, 2014

Nutmeg (family: Myristicaceae) is a plant that originated from Banda islands and is widely cultivated in several places in the world. Secondary metabolites of this plant have a high value because of their benefits for the health, food, and beauty industries. This study aims at developing DNA barcode for nutmeg (*Myristica fragrans*) using standard recommended fragment of *mat*K (*maturase* K) gene. Universal *mat*K primer pairs were used to amplify 889 bp DNA fragment. BLAST search from NCBI site showed that Sangihe nutmeg has 100% identity with *Myristica fatua, M. maingayi,* and *M. globosa.* It also has 3 nucleotides difference with *Rivola sebifera* (identity 99.58%) and 4 nucleotides difference with *Knema laurina* (identity 99.43%). It can be inferred from this study that single locus of *mat*K gene cannot be used to differentiate species in *Myristica*; it can only be used to differentiate the genus level within family Myristicaceae.

Key words: Sangihe nutmeg, DNA barcode, *mat*K, *Myristica*

INTRODUCTION

Nutmeg family (Myristicaceae) comprises of many species, which is widely spread in tropical rain forests around the world. This plant has ecological and ethnobotanical significances (Steeves 2011), and is amongst older group within angiospermae, which consists of recently evolved species (Newsmaster & Ragupathy 2009). One member of this family is *Myristica fragrans* (Houtt.), which was originated from eastern part of Indonesia (North Sulawesi and Mollucas). Identification of species from nutmeg family is difficult considering the similarity of their leaves, so identification relies heavily on the characteristics of small flowers (1-4 mm in size), which can only be obtained when the plant is mature sufficiently (Newsmaster & Ragupathy 2009). Identification of nutmeg is needed to mitigate contamination, substitution, and falcification of herbal products, as WHO suggested that falcification of herbal products is a threat to consumer safety (Newsmaster *et al*. 2013).

A previous research about Myristicaceae showed that misidentification of spesies in this family was more than 50% in herbarium specimen and 25% in ecological plot (Steeves 2011). Due to the difficulty in identifying the morphology, especially for cryptic species, DNA diagnostic tool including plastid DNA barcode and multilocus genomic marker can be used (Schori & Showalter 2011). DNA barcode is a short sequence of a gene, usually less than 700 bp, which have been agreed upon and used to identify species based on the references contained in the DNA sequence database (Dick & Kress 2009).

One of the challenges in the selection of this barcode is to distinguish very closely related or newly developed species. A species cannot be precisely identified using DNA barcodes if the variation within barcode between species is low, or related species still retain ancestral polymorphism, or they have a history of hybridization (Dick & Kress 2009). Ideally, a DNA barcode should be routinely used using a primer pair, which can also be used for bidirectional sequencing, and provides the maximum level of discrimination between species (Janzen 2009).

Many genes used in plant systematics are in substitution rate spectrum, representing genes that evolve rapidly or slowly. Which genes to be used, is usually determined by the level of phylogenetic analysis conducted by researcher. Each region in a gene has strengths and weaknesses. Good quality sequences, for example, can be found in the *rbc*L (*ribulose-1, 5-bisphosphate carboxylase oxygenase large subunit*) and *atp*B (ATPase beta-subunit), but

*Corresponding author. Phone: +62-431-827932, Fax: +62-431-853715, E-mail: trinatallei@yahoo.com

these sequences have a low level of differentiation of species because they are highly conserved among plant groups, therefore its resolution is only good in the level of family and above. Sequences that have high degree of species differentiation are *trn*H-*psb*A (chloroplast intergenic spacer) and *mat*K (*maturase* K), because it evolves so quickly that provides enough character to analyze evolution below family level (Barthet 2006; Hollingsworth 2011).

The *mat*K gene is considered to evolve rapidly, due to the fact that the gene has a high degree of substitution and its sequence is more varied than other genes (Barthet 2006). However, a group of researcher at the Consortium for the Barcode of Life (CboL) recommends two loci combination, *rbc*L and *mat*K, as standard DNA barcode for plants. These two regions in chloroplast DNA were chosen because of having high degree of differentiation between species (Bafeel *et al*. 2011). Furthermore, these two genes play important role in phylogenetic reconstruction for land plants (Kuzmina *et al*. 2012).

Species identification in Myristicaceae family is important in order to distinguish invasive species from endangered species, or species that has an economic significance. This study aims at analyzing *mat*K gene as a candidate for DNA barcode in nutmeg family.

MATERIALS AND METHODS

DNA Extraction. Fresh leaf of Sangihe nutmeg was extracted using innuPrep DNA Micro Kit (Analytik Jena) according to the protocol given by the company. Five mg of fresh leaf was cut into small pieces and placed in a 1.5 mL Eppendorf tube. Into the tube, 200 µL Lysis Solution TLS and 20 µL Proteinase K were added. The tube was vortexed vigorously for 5 seconds and incubated at 50 °C for 30 min. Tube was centrifuged at 12,000 rpm for 1 min and supernatant was transferred into new 1.5 mL tube. Into the supernatant, 200 µL Binding Solution TBS was added to lyse the sample then vortexed.

Sample was applied to Spin Filter located in a 2.0 mL Receiver Tube and centrifuged at 12,000 rpm for 1 min. Into the Spin Filter, 400 µL Washing Solution was added and centrifuged at 12,000 rpm for 30 sec. Receiver Tube with filtrate in it was discarded and Spin Filter was placed in a new Receiver Tube. Into the Spin Filter, 750 µL Washing Solution MS was added and the tube was centrifuged at 12,000 rpm for 30 sec. The Receiver Tube containing filter was discarded and Spin Filter was again placed in a new Receiver Tube. The tube was centrifuged for maximum speed for 2 min to remove all traces of ethanol. The Receiver Tube was discarded. The Spin Filter was added to Elution Tube and 100 µL of Elution Buffer was added. The tube was incubated for 1 min at room temperature and centrifuged at 8,000 rpm for 1 min. The eluted DNA was used for DNA amplification.

DNA Amplification. Primers used for DNA amplification of *mat*K gene were MatK-1RKIM-f 5'-ACCCAGTC CATCTGGAAATCTTGGTTC-3' and MatK-3FKIM-r 5'-CGTACAGTACTTTTGT GTTTACGAG-3' designed by K. J. Kim from School of Life Sciences and Biotechnology, Korea University, Seoul, Korea (Kuzmina *et al*. 2012). DNA was amplified using 5x Firepol PCR Master Mix Ready-to-Load (Solis BioDyne). Total volume for amplification was 40 µL consisted of 2 µL DNA sample and 1.5 µL of each primer (10 µM). DNA amplification was carried out in PCR TPersonal (Bimetra) as follows: predenaturation at 95°C for 4 min followed by 35 cycles of denaturation at 95°C for 30 sec, annealing at 50 °C for 40 sec, elongation at 72 °C for 50 sec, and one cycle of final extention at 72 °C for 2 min. The PCR products were separated by 1.0% agarose gel electrophoresis. Amplicon was bidirectionally sequenced at First BASE (Malaysia).

Data Analysis. Sequence chromatogram was edited using Geneious 6.1.6 (Biomatters Ltd., Auckland, New Zealand) and assembled into bidirectional contig. Primer sequences were removed, reverse and complement process were conducted using primer MatK-3FKIM-r, and combined using MUSCLE (Multiple Sequence Comparison by Log-Expectation), which is integrated within Geneious.

The *mat*K gene sequence was employed in Basic Local Alignment Search Tool (BLAST) (http://blast.ncbi.nlm.nih.gov/Blast.cgi) to query for highly similar sequence. Correct identification means identity percentage of BLAST is the highest searched sequences derived from expected species, or species from expected genus. In contrary, ambigous identification means identity percentage of BLAST is the highest of the searched sequenced derived not from expected species or genus, or family. A final alignment for *mat*K sequences was generated using Multalin V.5.4.1 developed by Corpet (1988) (http://multalin.toulouse.inra.fr/ multalin/). Sequences were trimmed at both ends of the alignment to avoid many missing data at the end of the sequences so the final characters were 707 sequences. Phylogenetic tree was constructed using software developed by Dereeper *et al*. (2008) (http://www.phylogeny.fr) and NCBI (http://ncbi.nlm.nih.gov). Similarity percentage matrix was calculated based on Clustal 2.1 (www.ebi.ac.uk)

RESULTS

We successfully amplified partial *mat*K gene from chloroplast genome of Sangihe nutmeg for the development of nutmeg DNA barcode (Figure 1).

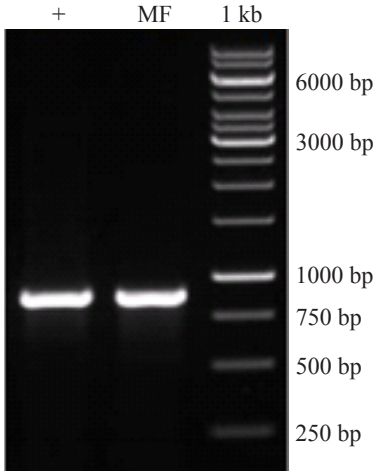

Figure 1. Agarose gel of amplified partial *mat*K gene (889 bp) of Sangihe nutmeg (*Myristica fragrans*) using MatK-1RKIM-f and MatK-3FKIM-r. (+) positive control, (MF) Sangihe nutmeg, and 1 kb marker.

Its sequence was used as query sequence in BLAST at NCBI to find similar sequence under the same or different genus within family Myristicaceae. The most highly similar identity sequences obtained from the GenBank are *M. fragrans, M. maingayi, M. fatua*, *M. globosa* (100%), *Virola sebifera* (99.58%), and *Knema laurina* (*M. laurina*) (99.43%) (Table 1). This finding confirms that Sangihe nutmeg is *M. fragrans*. It has similarity 99.58% with *Kneuma laurina* (synonym of *M. laurina*), which has 4 nucleotides difference with all *M. fragrans*. *Virola sebifera* has 3 nucleotides difference with all *M. fragrans* with 99.58% similarity (Table 2).

Reconstruction of phylogenetic tree showed that *mat*K locus could not differentiate Sangihe nutmeg from *M. fragrans* and other *Myristica* (*M. fatua*, *M. maingayi,* and *M. globosa*) (Figure 2). The tree was reconstructed based on the highest likelihood from sequence alignment. Horizontal dimension of the tree describes the number of genetic change. Value of 0.002 explains the length of the branch representing the number of nucleotide substitution (number of substitution per 100 nucleotide site). There was no

Table 1. Similarity percentage of Sangihe nutmeg (*Myristica fragrans*) with *Knema laurina* and *Virola sebifera* calculated using Clustal 2.1 (www.ebi.ac.uk)

Species	1	2	3	4	5	6	7
EU090507.1_*Virola_sebifera*	100.00	99.58	99.58	99.58	99.58	99.58	99.29
PalaSangihe_*Myristica_fragrans*	99.58	100.00	100.00	100.00	100.00	100.00	99.43
AJ966803.1_*Myristica_fragrans*	99.58	100.00	100.00	100.00	100.00	100.00	99.43
EU669472.1_*Myristica_fragrans*	99.58	100.00	100.00	100.00	100.00	100.00	99.43
GQ248165.2_*Myristica_fatua*	99.58	100.00	100.00	100.00	100.00	100.00	99.43
DQ401374.1_*Myristica_maingayi*	99.58	100.00	100.00	100.00	100.00	100.00	99.43
AY220450.1_*Knema_laurina*	99.29	99.43	99.43	99.43	99.43	99.43	100.00

Table 2. Similarity percentage and number of different nucleotides in *mat*K gene of Sangihe nutmeg (*Myristica fragrans*) and other species within Myristicaceae retrieved from GenBank

Species	Accession number	Identity (%)	Number of different nucleotides
Myrstica fragrans	EU669472.1	100.00	0
Myristica fatua	GQ248165.2	100.00	0
Myristica maingayi	DQ401374.1	100.00	0
Myristica globosa	GQ248166.1	100.00	0
Virola sebifera	EU090507.1	99.58	3
Virola michelii	JQ626468.1	99.43	4
Knema laurina	AY220450.1	99.43	4
Virola kwatae	JQ626460.1	99.29	5
Virola nobilis	GQ982126.1	99.29	5
Virola multiflora	GQ982125.1	99.29	5
Compsoneura debilis	EU090476.1	99.29	5
Coelocaryon preussii	AY743475.1	99.15	6
Staudtia kamerunensis	KC627748.1	99.15	6
Compsoneura ulei	EU090503.1	99.15	6
Compsoneura excelsa	EU090478.1	99.15	6
Compsoneura mexicana	EU090484.1	99.01	7
Iryanthera hostmanni	AY220449.1	99.01	7
Cephalosphaera usambarensis	AY220443.1	99.01	7
Iryanthera sagotiana	JQ626420.1	99.01	7
Coelocaryon oxycarpum	AY220444.1	99.01	7

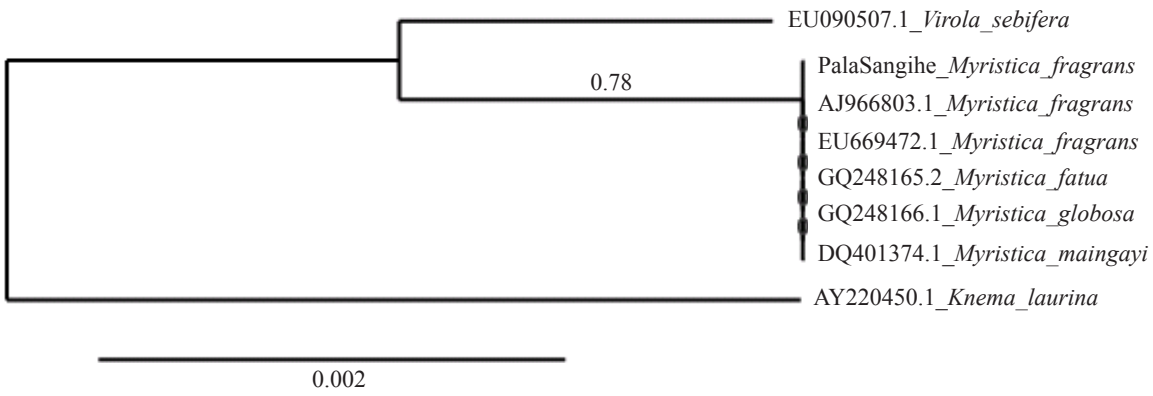

Figure 2. Phylogenetic tree of representative of genus *Myristica, Virola,* and *Knema* constructed based on likelihood phylogeny of nucleotide sequences of *mat*K gene, which were aligned using available online program at www.phylogeny.fr. Value of 0.002 shows the substitution rate per nucleotide site.

```
                51                                                  100
PalaSangihe     AATAGTCTCA TTACTCCAAA GAAATCCATT TCTCTTTTAA AAAAAGAGAA
V. sebifera     AATAGTCTCA TTACTCCAAA GAAATCCATT TCTCTTTTTA AAAAAGAGAA
K. laurina      AATAGTCTCA TTACTCCAAA GAAATCCATT TCTCTTTTAA AAAAAGAGAA
   Consensus    AATAGTCTCA TTACTCCAAA GAAATCCATT TCTCTTTTaA AAAAAGAGAA

                301                                                 350
PalaSangihe     ACCCTTTCAT ACATTATGTC AGATATCAAG GGAAATCCAT TCTGGCTTCA
V. sebifera     ACCCTTTCAT GCATTATGTC AGATATCAAG GGAAATCCAT TCTGGCTTCA
K. laurina      ACCCTTTCAT ACATTATGTC AGATATCAAG GGAAATCCAT TCTGTCTTCA
   Consensus    ACCCTTTCAT aCATTATGTC AGATATCAAG GGAAATCCAT TCTGgCTTCA

                401                                                 450
PalaSangihe     TTGGCAAAGT CATTTTTCCT TGTGGTGTCA ACCGGACAGG ATCCATATAA
V. sebifera     TTGGCAAAGT CATTTTTCCT TGTGGTGTCA ACCGGACAGG ATCCATATAA
K. laurina      TTGGCAAAGT CATTTTTCCC TGTGGTGTCA ACCGGACAGG ATCCATATAA
   Consensus    TTGGCAAAGT CATTTTTCCt TGTGGTGTCA ACCGGACAGG ATCCATATAA

                451                                                 500
PalaSangihe     ACCAATTATA CAATTATTCC TTCGATTTTA TGGGCTATCT TTCAAGTGTA
V. sebifera     ACCAATTATA CAATTATTCC TTCGATTTTC TGGGCTATCT TTCAAGTGTA
K. laurina      ACCAATTATA CAATTATTCC TTCGATTTTC TGGGCTATCT TTCAAGTGTA
   Consensus    ACCAATTATA CAATTATTCC TTCGATTTTc TGGGCTATCT TTCAAGTGTA

                601                                                 650
PalaSangihe     TGATTGGATC ATTGGCTAAA GCGAAATTTT GTAATGTATC TGGTCAACCC
V. sebifera     TGATTGGATC ATTGGCTAAA GCGAAATTTT GTAATGTATC TGGTCAACCC
K. laurina      TGATTGGATC ATTGGCTAAA GCGAAATTTT GTAATGTATC CGGTCAACCC
   Consensus    TGATTGGATC ATTGGCTAAA GCGAAATTTT GTAATGTATC tGGTCAACCC
```

Figure 3. Sequence alignment of *mat*K gene of *Myristica fragrans, Knema laurina* (*Myristica laurina*), and *Virola sebifera* using Multalin (available at http://multalin.toulouse.inra.fr/multalin) showing different number and position of nucleotide.

indel (insertion and deletion) found in *mat*K gene sequence of Sangihe nutmeg.

The highest similiraty (>99%) with other species from the genus *Myristica, Knema*, *Virola*, *Compsoneura, Coelocaryon, Staudtia, Iryanthera,* and *Cephalosphaerai*, proves that it belongs to family Myristicaceae. The similarity with genus *Magnolia* and *Manglietia* (Magnoliaceace) was around 92%. Position of different nucleotides in species with the same or different number of nucleotide difference were not the same in locus *mat*K as seen in alignment result amongst Sangihe nutmeg, *Virola sebifera*, and *Knema laurina* (Figure 3). Blast Tree View retrieved from NCBI can differentiate correctly the cluster of family Myristicaceae and Magnoliaceae based on *mat*K gene (data not shown).

DISCUSSION

Various molecular markers have been available in the literature for the purpose of identifying plant species. However, one marker alone is not sufficient to obtain the identification result, which is absolutely accurate (Dong *et al*. 2012). One locus on the

chloroplast genome, namely *mat*K, is widely used as a barcode though it is not always successfully applied for each plant.

Sangihe nutmeg (*M. fragrans* Houtt.), which has the synonym name of *M. officinalis* L.f., is an important herb and spice with economical significant due to having high aromatic content. Furthermore, its secondary metabolite has an important role in health, food, and cosmetics. In previous research, the same primer set of *mat*K used in this research (MatK-1RKIM-f and MatK-3FKIM-r) has been successfully used to amplify 35% of herbarium specimen and 45% of fresh specimen from 900 vascular plant specimen representing 312 species, 147 genus, 51 family, and 24 order (Kuzmina *et al*. 2012). Primer set of *mat*K- 3F-R and *mat*K-1R-F was successfully used to amplify *mat*K gene of Sulawesi local tomato (*L. esculentum*) plant. There is only one nucleotide difference between apple tomato and small fruit (cherry) tomato, thus the sequence similarity of these two kinds of tomatoes is 99.9%. Compared to other kind of tomatoes deposited in GenBank, the sequence difference is as high as 3 nucleotides therefore the sequence similiraty within species *L. esculentum* is 99.64% (Lawodi *et al.* 2013). Using *mat*K locus, sequence similarity of 99.43% can differentiate genus *Myristica* from *Knema*, and 99.25% from genus *Virola*, while sequence similarity of 99.64% places apple tomato and cherry tomato in the same species *L. esculentum*.

The development of DNA barcode is very important for plant identification and to retain plant's identity. Genetic diversity assessment is also important of species that are endemic, rarely found, or endangered, because it helps in plant conservation. Newsmaster and Raguphaty (2009) reports that *mat*K has significant variation and can be used for DNA barcode in nutmeg family. However, considering Sangihe nutmeg has 100% similarity with three other mentioned species, it is suggested that the *mat*K gene cannot be used to differentiate *Myristica* species. While 3 nucleotides difference in *mat*K gene can already place *Virola sebifera* as different species from *M. fragrans*, the same nucleotide difference in the same gene still places apple tomato and small fruit (cherry) tomato under the same species *L. esculentum* (Lawodi 2013). Ideally, for a single locus DNA barcode, minimum difference ranged from 0.0 to 0.27% for diferentiating species, or similarity of 99.74 to 100% to place organisms in the same species (Purushothaman *et al*. 2014), although in *L. esculentum*, the similarity within the species is 99.64% (Lawodi 2013). Therefore, minimum requirement for species differentiation is 99.64% similarity.

DNA barcode can differentiate cryptic species *T. cope* (Sunai pul) from different species with similar morphological character *T. wightii* (Kattai pul). Although there was no difference in *rbc*L sequences for these two cryptic species, *mat*K and *trnH-psb*A sequences show differences, 2 and 1 nucleotide, respectively. If these two loci are considered as two-gene approach, the interspecific variation will be large and there is no intraspecific variation (Ragupathy *et al*. 2009). We did not find in this study inter- and intra-specific variation in *Myristica* species.

Intergenic spacer *trn*H-*psb*A has sufficient sequence variation in *Compsoneura* species, but it is not so in *Virola*, although both are members of the Myristicaceae family. Both of these species are difficult to identify morphologically and has many sister species (Steeves 2011). Based on intra- and interspecific genetic divergence from the same genus (congeneric species), *mat*K divergences are roughly three times higher than the *rbc*L at the level of infraspecific. Its evolution consistently shows the level of differentiation amongst species in angiospermae (Janzen 2009; Saarela *et al*. 2013).

Newsmaster *et al*. (2007) show that variation at the locus *mat*K is moderate while the *trn*H-*psb*A is more varied, so as to provide results with good resolution within *Compsonuera* species. The use of both loci minimizes intraspecific variation relative to interspecific divergence distribution, with 95% of the specimen can be identified correctly in sequence identification analysis. Schroeder *et al*. (2012) state that the highest level of differentiation of species *Populus* was obtained using combination of two intergenic spacers (*trn*G-*psb*K, *psb*K-*psb*L) and coding region *rpo*C. Meanwhile coding regions *mat*K and *rbc*L, which are often recommended in the DNA barcode project only show moderate variability in *Populus* genus and are not efficient to be used in species differentiation.

A slightly different primer pairs (one or two nucleotides) given by Plant Working Group (PWG-CBOL), which is designed for *mat*K, namely matK2.1-Myristicaceae (5′-CCTATCCATCTGGATA TCTTGG-3′) and matK5-Myristicaceae (5′-GTTCTAGCACACG AAAATCG-3′) were used to amplify partial *mat*K gene with the length of approximately 760 bp in *Compsoneura*. From this, 10 varied sites and two indels were found. This variation, however, cannot differentiate species, eventhough there are intra- and interspecific variation (Newsmaster *et al*. 2007). Of 1355 plant specimen

collected, *mat*K can identify correctly 80% from 1079 species and 96% from 409 genera within Fabaceae family. Thus *mat*K is a valuable marker for plant species in this family (Gao *et al*. 2011).

DNA barcode single locus cannot provide adequate level of differentiation therefore combination of two loci was evaluated with 4 available markers to determine its ability to differentiate species. DNA barcode 2-loci recommended by CBOL, *mat*K and *rbc*L, can only provide 90% level of differentiation in *Cassia* species. Although PCR amplification was successful for all markers, intraspecific divergence is 0 for all *Cassia* species studied. By assuming that 1.0% divergence is the minimum requirement to differentiate 2 species, thus 4 markers can only differentiate 15-65% species studied when used separately. The use of 2-loci *mat*K+ITS and *rbc*L+*trn*H-*psb*A shows 100% differences between species (Purushosthaman *et al*. 2014).

Sequence recovery for *rbc*L (93%) is higher than *mat*K (81%), and *rbc*L (92%) is easier to recover than *mat*K (77%) from herbarium specimen in arctic plants. Distance-based and sequence-similarity analysis of *rbc*L + *ma*tK loci combination can differentiate 97% genus, 56% species, and 7% intraspecific (Saarela *et al*. 2013). Based on study on sequence quality and level of species differentiation, Janzen (2009) recommends two loci combination of *rbc*L and *mat*K as DNA barcode for plants. These two DNA barcodes can provide universal framework to use DNA sequence data to identify specimens routinely.

Newsmaster and Ragupathy (2009) reported that *mat*K gene alone can differentiate *Vachellia* species from other *Acacia* species, thus it can be used for taxa separation at the genus level. Especially for nutmeg, two-loci (*mat*K and *trn*H-*psb*A) has significant variation so that it can be used as a DNA barcode. However, the level of variation for the combination of the two loci is not significantly higher than *trn*H-*psb*A alone (Newsmaster *et al*. 2007). A study conducted by Burgess *et al*. (2011) shows that marker *rpo*C1 has the lowest resolution amongst loci evaluated (*mat*K > *atp*F-*atp*H > *rbc*L > *trn*H-*psb*A > *rpo*C1).

According to the Plant List, the family Myristicaceae has approximately 21 genera and 1,114 species with 206 (18.0%) synonyms. Therefore, the review of the naming of the genus and species in the family Myristicaceae needs to be done using molecular markers, such as DNA barcode. The use of DNA barcode in plants is not only important as a means of identification, but will also provide taxonomic resolution and insight in the diversity of plants. In this study, we found that eventhough primer pairs MatK-1RKIM-f and MatK-3FKIM-r can amplify partial *mat*K gene of Sangihe nutmeg (*Myristica fragrans*), the area cannot distinguish intra- and interspecific. Thus, re-testing is needed to be done using a combination of 2 or more loci to obtain DNA barcode that can be applied in *Myristica* species differentiation.

REFERENCES

Bafeel SO, Arif IA, Bakir MA, Khan HA, Al Farhan AH, Al Homai AA, Ahamed A, Thomas J. 2011. Comparative evaluation of PCR success with universal primers of maturase K (*mat*K) and ribulose-1, 5-bisphosphate carboxylase oxygenase large subunit (rbcL) for barcoding of some arid plants. *Plant Omics J* 4:195-198.

Barthet MM. 2006. Expression and function of the chloroplast-encoded gene *mat*K [Dissertation]. Virginia: Virginia Polytechnic Institute.

Burgess KS, Fazekas AJ, Kesanakurti PR, Graham SW, Husband BC, Newsmaster SG, Percy DM, Hajibabaei M, Barett SCH. 2011. Discriminating plant species in a local temperate flora using the *rbcL+ matK* DNA barcode. *Meth Ecol Evol* 2:333-340. http://dx.doi.org/10.1111/j.2041-210X.2011.00092.x

Corpet F. 1988. Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Res*16:10881-10890. http://dx.doi.org/10.1093/nar/16.22.10881

Dereeper A, Guignon V, Blanc G, Audic S, Buffet S, Chevenet F, Dufayard J-F, Guindon S, Lefort V, Lescot M, Claverie JM, Gascuel O. 2008. Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Res* 36:465-469. http://dx.doi.org/10.1093/nar/gkn180

Dick CW, Kress JW. 2009. Dissecting tropical plant diversity with forest plots and a molecular toolkit. *Bio Science* 59:745-755. http://dx.doi.org/10.1525/bio.2009.59.9.6

Dong W, Liu J, Yu J, Wang L, Zhou S. 2012. Highly variable chloroplast markers for evaluating plant phylogeny at low taxonomic levels and for DNA barcoding. *PLoS ONE* 7:e35071. http://doi:10. 1371/journal.pone.0035071

Gao T, Sun H, Yao J, Song Y, Zhu ZM, Chen S. 2011. Identification of Fabaceae plants using the DNA barcode *mat*K. *Planta Med* 77:92-94. http://dx.doi.org/10.1055/s-0030-1250050

Hollingsworth PM. 2011. Refining the DNA barcode for land plants. *Current Issue* 108:19451-19452. http://doi: 10.1073/pnas.1116812108. http://dx.doi.org/10.1073/pnas.1116812108

Janzen DH. 2009. A DNA barcode for land plants. *PNAS* 106:12794-12797. http://dx.doi.org/10.1073/pnas.0905845106

Kuzmina ML, Johnson KL, Barron HR, Herbert PDN. 2012. Identification of vascular plants of Churchill, Manitoba, using a DNA barcode library. *BMC Ecology* 12:1-11. http://dx.doi.org/10.1186/1472-6785-12-25

Lawodi EN, Tallei TE, Mantiri FR, Kolondam BJ. 2013. Variasi genetik tanaman tomat dari beberapa tempat di Sulawesi berdasarkan gen *mat*K. *Pharmacon* 2:114-121.

Newsmaster SG, Fazekas AJ, Steeves RAD, Janovec J. 2007. Testing candidate plant barcode regions in the Myristicaceae. *Molecular Ecology Resources* 8:480-490. http://doi: 10.1111/j.1471-8286.2007.02002.x

Newsmaster SG, Grguric M, Shanmughanandhan D, Ramalingam S, Ragupathy S. 2013. DNA barcoding detects contamination and substitution in North American herbal products. *BMC Medicine* 11:1-13.

Newsmaster SG, Ragupathy S. 2009. Testing plant barcoding in a sister species complex of pantropical *Acacia* (Mimosoideae, Fabaceae). *Molecular Ecology Resources* 9:172-180. http:// doi: 10.1111/j.1755-0998.2009.02642.x

Purushothaman N, Newsmaster SG, Ragupathy S, Stalin N, Suresh D, Arunraj DR, Gnanasekaran G, Vassou SL, Narasimhan D, Parani M. 2014. A tiered barcode authentication tool to differentiate medicinal *Cassia* species in India. *Genet Mol Res* 13:2959-2968. http://dx.doi.org/10.4238/2014.April.16.4

Ragupathy S, Newsmaster SG, Murugesan M, Balasubrmaniam V. 2009. DNA barcoding discriminates a new cryptic grass species revealed in an ethnobotany study by the hill tribes of the Western Ghats in southern India. *Molecular Ecology Resources* 9:164–171. http:// doi: 10.1111/j.1755-0998.2009.02641.x

Saarela JM, Sokoloff PC, Gillespie LJ, Consaul LL, Bull RD. 2013. DNA Barcoding the Canadian crctic flora: core plastid barcodes (*rbcL* + *matK*) for 490 vascular plant species. *PLoS ONE* 8: e77982. http://doi: 10.1371/journal.pone.0077982

Schori M, Showalter AM. 2011. DNA barcoding as a means for identifying medicinal plants of Pakistan. *Pak J Bot* 43:1-4.

Schroeder H, Hoeltken AM, Fladung M. 2012. Differentiation of *Populus* species using chloroplast single nucleotide polymorphism (SNP) markers essential for comprehensible and reliable poplar breeding. *Plant Biol (Stuttg)*. 14:374-381. http://doi: 10.1111/j.1438-8677.2011.00502.x

Steeves RAD. 2011. An intrageneric and intraspecific study of morphological and genetic variation in the neotropical *Compsoneura* and *Virola* (Myristicaceae) [Thesis]. Canada: The University of Guelph.