INTERNATIONAL JOURNAL OF APPLIED BIOLOGY AND PHARMACEUTICAL TECHNOLOGY

www.ijabpt.com ISSN : 0976-4550

Volume-6, Issue-2, April-June-2015 Coden IJABFP-CAS-USA Copyrights@2015

Received: 9th Mar-2015 *Revised: 1st April -2015* *Accepted: 1st April-2015*

Research article

APPLYING *MATK* GENE FOR IDENTIFICATION OF LILIOPSIDA PLANT SPECIES FROM NORTH SULAWESI THROUGH BOLD SYSTEMS

Beivy Jonathan Kolondam

Department of Biology, Faculty of Mathematics and Natural Sciences, Sam Ratulangi University
Jl. Kampus UNSRAT FMIPA, Manado 95115, Indonesia
Corresponding e-mail: beivy.kolondam@unsrat.ac.id

ABSTRACT: DNA barcoding is a tool for species identification. For plant species identification, two of plastid genes (*rbcL* and *matK*) were used as standard barcodes. There are limitations of each gene marker but *matK* is considered to be the closest plant analogue to the *CO1* animal barcode. As a mega-biodiversity country, Indonesia has many plant species used for ornamental and/or medical purposes. This study aimed to assess the identification of 15 known ornamental Liliopsida plant species in North Sulawesi using *matK* gene as a single marker. There were three possibilities result for species identification in this study: species level identification by top hit (1 specimen), species level identification by highest similarities (7 specimens), and genera level identification by highest similarities (7 specimens). Species identification by the highest similarity is reliable only when the result is identical with the sequence stored in the database.
Key words: DNA Barcoding, *matK*, ornamental plants, North Sulawesi

INTRODUCTION

As a tool for species identification, DNA barcoding served many purposes. A proposal for using defined portions of the plastid genes (rbcL and matK) as standard barcodes for land plants were endorsed by the Plant Working Group of the Consortium for the Barcode of Life (CBOL) in 2009 (CBOL, 2009). Both marker and several other candidate markers have been tested in various taxonomic settings (Fazekas *et al*., 2008; Kress and Erickson, 2007; Lahaye *et al*., 2008, Seberg and Petersen, 2009). Plant barcodes are reliable for scientific and applied purposes which is potential for wider use such as for forensics and economic uses (Chase *et al*., 2005). There are limitations of each gene marker but matK is considered to be the closest plant analogue to the *CO1* animal barcode, disregarding the difficulties to PCR amplification using existing primer sets in non-angiosperms (Hollingsworth *et al*., 2011).

Indonesia is a mega-biodiversity country in the topical. Diversity in plant species reveals the use of plants as ornaments and/or for medicinal purposes. Several species were commercialized and has a remarkable price based on the varieties. This study aimed to assess the identification of several Liliopsida plants in Manado using *matK* gene as a single marker. Assessment will include accuracy of species identification using BOLD Systems database. Specimens were chosen from several popular ornamental plants.

MATERIALS AND METHODS

DNA Extraction

Total DNA from fresh leaf samples were extracted using innuPREP Plant DNA Kit (Analytik Jena) according to manual from the manufacturer. Modification of the protocol was made to gain higher concentration of chloroplast DNA by increasing the time for incubation in lysis solution and proteinase K to one hour.

Polymerase Chain Reaction

Amplification of matK gene occurred in 50 µl PCR reaction using 5X Firepol PCR Master Mix Ready-to-Load (Solis Biodyne). The final concentration contained 1x reaction buffer, 1.25 unit Taq polymerase, 200 µM dNTPs each, 1.5 mM $MgCl_2$, 10 pmol of each primer and 1 µl of DNA sample. For amplification and sequencing of matK gene, primer pairs 3F-r (5' CGT ACA GTA CTT TTG TGT TTA CGA G 3') and 1R-f (5'ACC CAG TCC ATC TGG AAA TCT TGG TTC 3') were used (Little and Stevenson, 2007). The cycling condition were started with 95°C 2 minutes of initial denaturation and continued with 35 cycles of 95°C 30 seconds, 50°C 30 seconds, and 72°C 50 seconds. Successfully amplified PCR products were separated by electrophoresis in 1% of agarose gel and visualized under UV light.

Sequencing

All of the PCR products were cleaned up and bi-directionally sequenced using BigDye® Terminator v3.1 cycle sequencing kit chemistry. The sequencing process was done by ABI PRISM 3730xl Genetic Analyzer (Applied Biosystems).

Data Analysis and Database Search

Chromatogram files were assembled and edited under Geneious v5.6 (Drummond *et al*., 2012) integrated platform software. All sequence data were aligned using MUSCLE (Edgar, 2004) and both primer sequences were trimmed out. To check for sequence reading error, the alignment of sequences were compared by its translated amino acid. Based on assumption that *matK* is a functional gene, there should be free of stop codons if translated *in silico* with one of three alternative amino acid reading frames. Some extra nucleotides that caused stop codon in translation were deleted.

The identification of plant samples achieved by employing BOLD Systems (Ratnasingham and Hebert, 2007) database. All searches in this database were done on December 7, 2014.

RESULT AND DISCUSSION

The fragment matK gene was successfully amplified from all samples using universal primers (Figure 1). Strong bands were observed by agarose gel electrophoresis. Chromatogram of the generated fragments varies in length from 887 bp to 908 bp (alignment not shown). There were 53 bp of primer sequences trimmed from both end of the fragments prior to identification using BOLD Systems.

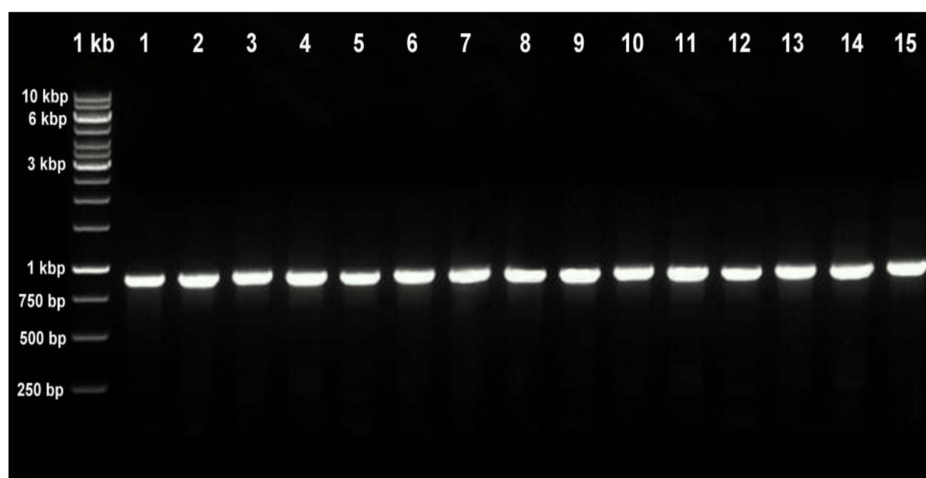

Figure 1. Electrophoresis of PCR Products in 1% agarose gel.

Based on the identification using BOLD Systems (Table 1), family names were all matched with the respective samples but the top hit was not completely reliable. The only sample that correctly identified by top hit was *Zamioculcas zamiifolia*. It also has 100% in similarity. This was understandable since the results of BOLD Systems were ranked based on the score, not the sequence similarity.

In the other hand, 7 out of 15 samples were be able to be identified correctly to the species level (*Eucharis amazonica, Caladium bicolor, Sansevieria trifasciata var. Hahnii, Sansevieria trifasciata var. Laurentii, Wittrockia superba, Canna indica, and Tradescantia spathacea*) by relying on the highest similarity for identification. *Canna indica* was the only result shown top hit that has the same genus. The other results have different genera names, although it shown the correct family. For *T. spathacea*, top hit identification showed only the order name (Commelinales). Both Sansevieria samples have identical DNA barcodes. Besides the inaccuracy for plant identification, the highest number in sequence similarity can be a better guide. There were also 7 out of 15 samples that have identification accuracy to the genera level. The samples were *Aglaonema* sp. var. Pride of Sumatera, *Aglaonema commutatum, Anthurium plowmanii,Dieffenbachia oerstedii, Dieffenbachia seguine, Cordyline terminalis,* and *Piper nigrum*.

If identification based on the highest similarity is reliable, it raises the question about its reliability. Result of 100% in sequence similarity (identical) is not a guarantee for identification some close-related species. For *W. superba* (99.75%) is accurate but for *C. terminalis* (99.76%) pointed to a different species. This implied that more consideration is required, even though the result showed more than 99% similarity with a species. Dissimilarity can also be the result of low quality reading in cycle sequencing or poor DNA editing.

Table 1. Species Identification based on matK Gene Marker

No.	Specimen	Family	Identification through BOLD Systems (December 7, 2014)			
			Top Hit	Similarity (%)	Highest Similarity	Similarity (%)
1	*Eucharis amazonica*	Amaryllidaceae	*Lycoris chinensis*	*98.71*	*Eucharis amazonica*	100
2	*Aglaonema* sp. var. Pride of Sumatera	Araceae	*Anchomanes difformis*	98.33	*Aglaonema crispum*	99.75
3	*Aglaonema commutatum*	Araceae	*Anchomanes difformis*	98.21	*Aglaonema crispum*	99.75
4	*Anthurium plowmanii*	Araceae	*Anthurium schlechtendalii*	98.98	*Anthurium ravenii*	99.1
5	*Caladium bicolor*	Araceae	*Xanthosoma mafaffa*	98.1	*Caladium bicolor*	100
6	*Dieffenbachia oerstedii*	Araceae	*Dieffenbachia spruceana*	98.69	*Dieffenbachia aglaonematifolia*	99.75
7	*Dieffenbachia seguine*	Araceae	*Dieffenbachia spruceana*	98.57	*Dieffenbachia oerstedii*	99.25
8	*Zamioculcas zamiifolia*	Araceae	*Zamioculcas zamiifolia*	100	*Zamioculcas zamiifolia*	100
9	*Cordyline terminalis*	Asparagaceae	*Cordyline australis*	99.53	*Cordyline cannifolia*	99.76
10	*Sansevieria trifasciata* var. Hahnii	Asparagaceae	*Dracaena angustifolia*	99.06	*Sansevieria trifasciata*	100
11	*Sansevieria trifasciata* var. Laurentii	Asparagaceae	*Dracaena angustifolia*	99.06	*Sansevieria trifasciata*	100
12	*Wittrockia superba*	Bromeliaceae	*Ochagavia carnea*	*99.04*	*Canistrum superbum* (synonym of *Wittrockia superba*)	99.75
13	*Canna indica*	Cannaceae	*Canna paniculata*	99.76	*Canna indica*	100
14	*Tradescantia spathacea*	Commelinaceae	Commelinales (no specific name)	97.24	*Tradescantia spathacea*	100
15	*Piper nigrum*	Piperaceae	Piperales (no species name)	99.04	Piper (no species name)	99.04

CONCLUSION

There were three possibilities for species identification in this study: (1) species level identification by top hit (1 out of 15 specimens) (2) species level identification by highest similarities (7 out of 15 specimens) (3) genera level identification by highest similarities (7 out of 15 specimens). Species identification by the highest similarity is reliable only when the result is identical with the sequence stored in the database.

REFERENCES

CBOL (2009). A DNA Barcode for Land Plants. Proc. Natl. Acad. Sci.: Vol. 106, 31, 12794–12797.

Chase, M.W. Salamin, N. Wilkinson, M. (2005). Land plants and DNA barcodes: short-term and long-term goals. Philos. Trans. R. Soc. Lond. B. Biol. Sci.: Vol. 29, 360, 1889-1895.

Drummond AJ, Ashton B, Buxton S, Cheung M, Cooper A, Duran C, Field M, Heled J, Kearse M, Markowitz S, Moir R, Stones-Havas S, Sturrock S, Thierer T, and Wilson A (2012). Geneious v5.6. New Zealand.

Edgar, R. C. (2004). MUSCLE: Multiple Sequence Alignment with High Accuracy and High Throughput. Nucl Acids Res.: Vol. 32, 5, 1792-1797.

Fazekas A.J. Burgess, K.S. Kesanakurti P.R., Graham, S.W. Newmaster, S.G. (2008). Multiple Multilocus DNA Barcodes from the Plastid Genome Discriminate Plant Species Equally Well. PLoS ONE: Vol. 3, 7, e2802.

Hollingsworth PM, Graham SW, Little DP (2011). Choosing and Using a Plant DNA Barcode. PLoS ONE: Vol. 6, 5, e19254.

Kress, W. J. and Erickson, D. L. A (2007). Two-locus Global DNA Barcode for Land Plants: The Coding rbcL Gene Complements the non-coding trnH-psbA spacer region. PLoS ONE: Vol. 2, 6, e508.

Lahaye, R., van der Bank, M., Bogarin, D. (2008) DNA Barcoding the Floras of Biodiversity Hotspots. Proc. Natl. Acad. Sci.: Vol. 105, 8, 2923–2938.

Little, D. P. & Stevenson, D. W. (2007). A comparison of algorithms for the identification of specimens using DNA barcodes: examples from gymnosperms. *Cladistics*: Vol. 23, 1, 1-21.

Ratnasingham, S. and Hebert, P. D. N. (2007). BOLD: The Barcode of Life Data System (www.barcodinglife.org). Molecular Ecology Notes: Vol. 7, 3, 355-364.

Seberg, O. and Petersen, G. (2009). How Many Loci Does It Take to Barcode a Crocus? *PLoS ONE*: Vol. 4, 2, e4598.

ISSN : 0976-4550

INTERNATIONAL JOURNAL OF APPLIED BIOLOGY AND PHARMACEUTICAL TECHNOLOGY

Email : ijabpt@gmail.com

Website: www.ijabpt.com