Life Science and Biotechnology
Volume 1, Number 1, Mei 2023
Pages: 21 – 26

Life Science Biotechnology

The *matK* Mini-barcode as a Potential Molecular Identification Tool for Medicinal Orchids

Mukhamad Su'udi[*1], Dwi Setyati[1], Gita Ayu Khoirunnisa[1], Yusfi Afidah[1], Fuad Bahrul Ulum[1]
[1]Laboratory of Botany, Biology Department, Jember University, City Jember, Indonesia
* Correspondence Author: msuudi52@gmail.com

Abstract

matK is one of the universal loci which is commonly used as a barcode in the molecular identification of plants. Up to now, the problem of *matK* as a single locus is the use of the entire gene region as barcodes. The long amplicon size of *matK* raises several problems in DNA barcoding. The aim of this study was to uncover a small region of the *matK* gene as a potential mini-barcode region for alternative sites to get more efficient and effective results. *In silico* studies performed include the collection of *matK* sequence of medicinal orchid from GenBank, the amplicon size from *matK* primer sets that are commonly used in the Orchidaceae family, and PCR analysis. Data analysis was done by comparing the results of PCR amplification from several primer sets: matK 390F-1326R, 743F-R2, 712F-1154R and 712F-1216R. The results show that all primer pairs are able to amplify the DNA template isolated from *Dendrobium purpureum*. However, the shorter DNA band obtained from *matK* 743F-R2, 712F-1154R and 712F-1216R primers sets were visualized as thicker than the longer amplicon band produced from *matK* primer 390F-1326R. This indicates that short *matK* sequence for DNA barcoding (mini-barcode) was more efficient in amplifying DNA templates compared to generally applied barcoding using full-length sequences.

Keywords: DNA barcoding, maturase K, orchids

Introduction

DNA barcoding approaches have been widely applied in species identification. The reason for applying this method is due to its ability to identify any samples quickly, precisely and accurately, and does not depend on the morphological or physiological conditions (Hebert & Gregory, 2005). The principle of DNA barcoding is to recognize species patterns at the molecular level using standardized molecular markers on DNA fragment sequences (Kress, & Erickson, 2008; Li et al, 2015). So far, there are two chloroplast loci which are recommended for molecular identification through DNA barcoding in plants, namely *matK* and *rbcL*. Based on the agreement of The Consortium Barcode of Life, the use of multi-locus (*matK*+*rbcL*) has a stronger ability to discriminate and identify plant species (CBOL Plant Working Group, 2009), however, it is also possible to use a single locus (Li et al, 2015).

The *matK* is considered to be one of the most informative locus to be used in determining plant phylogenetic relationships (Hilu et al, 2003). In addition, this locus has also been reported to have strong discrimination and identification abilities at almost all plant levels (Li et al, 2011; Yu et al, 2011). Therefore, the *matK* has been designated as one of the universal barcodes for plants. Besides the above-mentioned superiority belonging to *matK*, the use of full-length gene regions as barcodes brings about too large amplicon sizes which are considered inefficient, particularly at the amplification and sequencing stages. In addition, the costs required for sequencing are also higher (Dunning & Savolainen, 2010; Hollingsworth et al, 2011; Hajibaei & Mc Kenna, 2012). Based on these problems, designing the *matK* with a relatively small size (mini-barcode) for application in the molecular identification of plant species needs to be studied and developed.

Materials and Methods

The research was carried out experimentally and supported with bioinformatic analysis, data retrieval from databases and previous research literature. This research consisted of several stages: *in silico* study of medicinal orchids, exploration, collection and selection of *matK* primers, isolation of genomic DNA and PCR analysis as described as follows.

Sequences Collection

Sequence exploration of *matK* from medicinal orchids was retrieved from Genbank NCBI based

on the list of Asian medicinal orchid species reported by Teoh (2016). The sequences were collected and selected regarding the size of the full sequence and its open reading frame (ORF). For the primer design, information was collected from articles published in reputable journals. The primer name, nucleotide sequence, and information on amplicon size were collected.

Primer Design

The selection of the *matK* region for the mini-barcode was determined with a bioinformatics approach using *matK* sequences collected from selected orchids. The sequences were aligned using the Cluxtal X program to determine the position of the *matK* primers. The known position and length of the primers were used as the basic information in selecting the designed primer pairs. After determining the position and specifying the length of the amplicon size, the primer candidates were further analyzed to meet the criteria as an ideal primer. The software used for primer design includes Oligo Analyzer 1.0.2, NEB Tm Calculator (www.tmcalculator.neb.com) and Primer BLAST.

DNA Isolation and Amplification

DNA was isolated from the leaves of *Dendrobium purpureum* orchid as a representative sample of medicinal orchids used in this study. The CTAB (Cetyl Trimethyl Ammonium Bromide) method was used for DNA isolation according to the protocol of Doyle & Doyle (1990). The fresh leaf sample weight used was 0.5 grams. DNA amplification was performed using several pairs of *matK* primers consisting of two primer pairs obtained from previously publised articles consisting of primers *matK* 390F: 5'-CGATCTATTCATTCAATATTTC-3', *matK* 1326R: 5'-TCTAGCACACGAAAGTCGAAGT-3' (Xu et al, 2015), *matK* 743F:5'-CTTCTGGAGTCTTTCTTGAGC-3' (Khew & Chia, 2011), *matK* R2:5'-CCCAATACAGTACAAAATTGAGC-3' (Batista et al, 2013) and two pairs of *matK* mini-barcode primers designed with a target sequence size of 400–500 bp (this study). DNA amplification was carried out as follows: pre-denaturation of 95ºC for 5 minutes, 35 cycles of denaturation of 95ºC for 30 seconds, annealing of 53ºC for 30 seconds, the extension of 72ºC for 1 minute 20 seconds, and final extension of 72ºC for 5 minutes. The PCR products were visualized using UV-transilluminator after electrophoresis in 1.25% agarose gel with EtBr staining and 100 bp marker plus DNA ladder (Bioneer, Korea).

Results and Discussion

matK Gene as Universal Barcode

The *matK*, a gene resident in chloroplasts, encodes a maturase protein which acts as a catalyst in the mRNA intron splicing process (Neuhaus & Link, 1987). Structurally, the *matK* gene is a coding area located between the two exons of the trnK gene (Figure 1). The overall length of the *matK* gene is about 1570 bp with an ORF product size of 500 aa in angiosperm plants (Neuhaus & Link, 1987; Hilu & Liang, 1997). Based on our in silico studies, the *matK* collected from several species of the Orchidaceae family possessed sequences size of more than 1500 bp (Table 1).

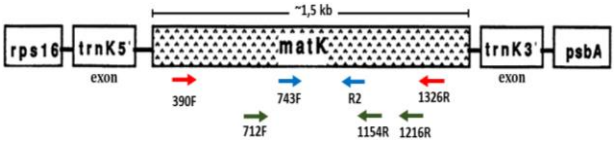

Figure 1. Gene structure and primer position of *matK* (Hilu & Liang, 1997)

The *matK* is considered one of the most informative loci for determining phylogenetic relationships (Hilu et al, 2003). This gene is also determined as the standard locus due to its universal properties and has a relatively high discriminatory ability for angiosperm plants including orchids (Lahaye et al, 2008). The universality of the *matK* is supported by the high level of nucleotide substitution in this region, leading the *matK* sequences appropriate to be used as barcode regions (Hollingsworth et al, 2011; Li et al, 2011). The level of *matK* nucleotide substitution is known to be three times higher than rbcL and 6 times higher at the amino acid level (Barthet & Hilu, 2007). The presence of *matK* in the chloroplast genome also makes *matK* considered very useful in reconstructing phylogenetic trees at various taxon levels owing to its relatively high level of evolution (Yu et al, 2011). Based on the phylogenetic studies, the *matK* gene provides satisfactory results for Orchidaceae members (Kim et al, 2014; Xu et al, 2015). The use of *matK* as a single or multi-locus with trnH-psbA was able to correctly identify >90% of orchid specimens (Lahaye et al, 2008). The ability of *matK* to discriminate in identifying the medicinal orchid species Dendrobium was higher than *rbcL* as proved by the phylogenetic tree that formed clades with the correct arrangement and was able to differentiate between species (Asahina et al, 2010). This ability is due to the inter-specific diversity of *matK* in the Orchidaceae family (Sosa et al, 2013).

However, despite the superior advantages of the *matK* gene, there is also a drawback that makes the *matK* less perfect in identification, i.e. the amplicon size produced is too big, which makes it less efficient in terms of time and cost (Hajibaei & Mc Kenna, 2012). In addition, the possibility of long mononucleotide repeats in genes able to trigger the formation of secondary structures that potentially inhibit the amplification and sequencing processes (Dunning & Savolainen, 2010; Yu et al, 2011). So far, the regions applied as barcodes by several existing *matK* primers still resulted in an average of > 500 bp amplicon size (Table 2).

Table 1. Full-length sequences collection of *matK* from medicinal orchids

No	Species name	Accession Code	Size (bp)	Amino acid (aa)
1	*Dendrobium fimbriatum*	AB519776.1	1875	447
2	*Dendrobium moniliforme*	AB519775.1	1877	446
3	*Dendrobium nobile*	AB519772.1	1878	446
4	*Dendrobium pulchellum*	AB519778.1	1870	446
5	*Dendrobium tosaense*	AB519771.1	1879	446
6	*Phaleonopsis deliciosa*	AB217749.1	1790	516
7	*Phaleonopsis wilsonii*	AB217751.1	1788	517
8	*Habenaria linguella*	MF945416.1	1655	511
9	*Geodorum recurvum*	KF673833.1	1652	518
10	*Bulbophyllum cylindraceum*	KX455832.1	1554	517
11	*Calanthe alismifolia*	KF673787	1688	519
12	*Calanthe cardioglossa*	KF673794.1	1689	519
13	*Calanthe davidii*	KF673796	1692	519
14	*Calanthe densiflora*	KF673799	1662	516
15	*Holcoglossum quasipinifolium*	HQ452924	1768	517
16	*Liparis cordifolia*	KJ459312.1	1557	518
17	*Mycaranthes pannea*	KY239608	1740	508
18	*Spiranthes sinensis*	AB040206	1530	504
19	*Vanda concolor*	JX202679	1629	517
20	*Cymbidium goeringii*	KF673831	1619	517

Mini-barcode DNA

The development in the molecular field has led to many new methods for assisting in the field of life science research, one of which is DNA barcodes. Currently, the term mini-barcoding is introduced as a DNA barcoding method using a small portion of a gene region or a short nucleotide sequence to identify an organism (Hajibaei et al, 2006). The mini-barcode is proposed as a new application for solving the problems arising from full barcodes in general. The use of full-length barcodes is considered to be ineffective for identifying specimens that underwent DNA degradation, for example, dry specimens or specimens with certain treatments that led to reduced quality of DNA (Meusnier et al, 2008). The use of mini-barcodes in PCR analysis allows for increased amplification results compared to the use of full barcodes. However, one thing should be considered that the discriminative ability of mini-barcodes decreases as the number of nucleotides decreases (Little, 2011; Meusnier et al, 2008).

The limitations of mini-barcoding in discriminating species can be overwhelmed by designing specific primers to precisely identify the species (Gao et al, 2019). Determination of the length and position of the primer is very important in determining its ability to discriminate between species (Hajibaei et al, 2006). The results of designing *matK* primer candidates as potential primers for mini-barcodes have been determined to test their effectiveness, three primers were obtained, namely *matK*712F: 5'-CTGTTTCTTCGTAAAACAGTCTTC-3', *matK* 1154R: 5'-GAGAAATAACTGGCTATGG-3' and *matK* 1216R: 5'-CGGTTTACTAATAGGATGACCC-3' with cohort partners namely 712F-1154R and 712F-1216R. Two primers retrieved from the literature (*matK* 390F-1326R and 743F-R2) were selected as comparison parameters in PCR effectiveness

Tabel 2. Primer sequences of *matK* applied for DNA barcoding in Orchidaceae family

No	Nucleotide sequences	Primer name	Amplicon size (bp)	Reference
1	5'-CGATCTATTCATTCAATATTTC-3'	390 F	722-930	Kim et al,2014; Xu et al,2015
	5'-TCTAGCACACGAAAGTCGAAGT-3'	1362 R		
2	5′-CGTTCTGACCATATTGCACTATG-3′	19 F	1548	Khew dan Chia, 2011
	5′-GGCAACAAAACTTCCTATATCC-3'	56 F		
	5′-CTTCTGGAGTCTTTCTTGAGC-3′	743 F		
	5′-CGGATAATGTCCAAATACCAAATA-3′	1520 R		
	5′-CAGAAAAGAGAAGGAATAATTGG-3′	1027 R		
	5′-GACTCCAAAACCTTCTGATACC-3′	336 R		
3	5′-TAATTTACGATCAATTCATTC-3′	Xf	784-1071	Siripiyasing et al, 2012; Tanee et al, 2012
	5′-GTTCTAGCACAAGAAAGTCG-3′	5r		
4	5'-CTAATACCCCATCCCATCCAT-3'	F2	630	Batista et al, 2013
	5'-CCCAATACAGTACAAAATTGAGC-3'	R2		
5	5'-ATATCCGCTACTCCTTCAGGAG-3'	KUf	1146	Srikulnath et al, 2015
	5'-CGAGCCAAAGTTCTAGCACACG-3'	Kur		

Mini-barcode in *matK* Gene

The position of each primer pair designed in this study is illustrated in Figure 1. The amplicon size of each primer is ± 900 bp (390F-1326R), ± 400 bp (743F-R2 and 712F-1154R), and ± 500 bp (712F-1216R). The primer sets were then tested using DNA samples isolated from *D. purpureum* as a template in the PCR process. Based on the PCR results, all of the *matK* primer pairs successfully amplify the DNA template in accordance with the predicted amplicon size. As the results shown in the visualization, the thickness of the DNA bands produced by the primers with long amplicon sizes is thinner compared to the results of shorter amplicon primers (Figure 2).

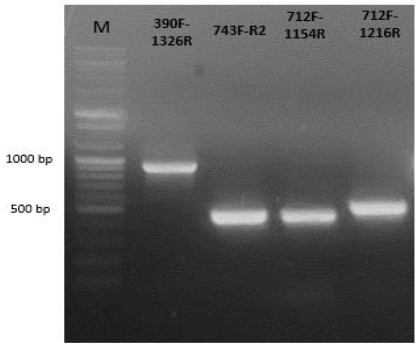

Figure 2. PCR product of *matK* obtained from designed primer sets visualized under UV-transilluminator. Marker (M) used 100 bp plus DNA ladder (Bioneer). DNA source as template for PCR: *Dendrobium purpureum*.

In accordance with the previously reported experiment, this data underlines that the shorter region selected as the barcode area, the faster and easier to amplify than the full sequence barcode. Mini-barcodes are also considered to have the ability to identify species under certain conditions such as DNA underwent degradation (Little, 2014; Gao et al, 2019). However, the challenge faced by the use of mini-barcodes is the selection of the right area to produce an accurate identification through full consideration during primer design (Lahaye et al, 2008).

Based on the results of PCR analysis, the region of 700-1216 bp which was chosen as the primer attachment position, was able to amplify DNA well. This is consistent with previous reports which stated that the *matK* gene region in the 600-1400 bp region is recommended as a region that has the potential to be a universal barcode since it is considered to represent information of the entire gene (Li et al, 2011). In addition, the 600-800 bp region has high variability and is stable, for use as a barcode (Li et al, 2011).

In silico studies show that the use of a short part of the *matK* gene regions is more efficient and effective for amplification than the full-length or whole barcode. The ability of mini-barcode *matK* for determination and species identification, followed by the advantages of the mini-barcode deliver this method more profitable for future application, especially in terms of identification and phylogenetic studies of the Orchidaceae family. The selection of an appropriate short region as a mini-barcode is highly recommended for amplification effectiveness in the 600-800 bp of the *matK* gene.

References

Asahina, H., Shinozaki, J., Masuda, K., Morimitsu, Y., & Satake, M. 2010. Identification of medicinal Dendrobium species by phylogenetic analyses using matK and rbcL sequences. Journal of Natural Medicines, 64(2): 133–138.

Barthet, M. M., & Hilu, K. W. 2007. Expression Of matK: Functional and Evolutionary Implications. 94(8): 1402–1412.

Batista, J. A. N., Borges, K. S., de Faria, M. W. F., Proite, K., Ramalho, A. J., Salazar, G. A., & van den Berg, C. 2013. Molecular phylogenetics of the species-rich genus Habenaria (Orchidaceae) in the New World based on nuclear and plastid DNA sequences. Molecular Phylogenetics and Evolution, 67(1): 95–109.

CBOL Plant Working Group. 2009. A DNA barcode for land plants. Proceedings of the National Academy of Sciences of the United States of America, 106(31): 12794–12797.

Doyle Jeff J, Doyle, J. L. 1990. A Rapid DNA Isolation Procedure for Small Quantities of Fresh Leaf Tissue. Phytochemical Bulletin, 19(1): 11–15.

Dunning, L. T., & Savolainen, V. 2010. Broad-scale amplification of matK for DNA barcoding plants, a technical note. Botanical Journal of the Linnean Society, 164(1): 1–9.

Gao, Z., Liu, Y., Wang, X., Wei, X., & Han, J. 2019. DNA Mini-Barcoding: A Derived Barcoding Method for Herbal Molecular Identification. Frontiers in Plant Science, 10: 1–26.

Hajibaei, M., & Mc Kenna, C. 2012. DNA mini-barcoddes. Methods in Molecular Biology, 858: 339–353. Retrieved from http://link.springer.com/10.1007/978-1-61779-591-6.

Hajibaei, M., Smith, A. M., Janzen, D. H., Rodriguez, J. J., Whitefield, J. B., & Hebert, P. D. N. 2006. Barcoding: A minimalist barcode can identify a specimen whose DNA is degraded. Molecular Ecology Notesology Notes, 6: 959–964.

Hebert, P. D. N., & Gregory, T. R. 2005. The promise of DNA barcoding for taxonomy. Systematic Biology, 54(5): 852–859.

Hilu, K. W., Borsch, T., Müller, K., Soltis, D. E., Soltis, P. S., Savolainen, V., Chatrou, L. W. 2003. Angiosperm phylogeny based on matK sequence information. American Journal of Botany, 90(12): 1758–1776.

Hilu, K. W., & Liang, H. 1997. The matK gene sequence variation and application in plant systematics. American Journal of Botany, 84(6): 830–839.

Hollingsworth, P. M., Graham, S. W., & Little, D. P. 2011. Choosing and Using a Plant DNA Barcode. PLoS ONE, 6(5): 1-13.

Khew, G. S. W., & Chia, T. F. 2011. Parentage determination of vanda miss joaquim (orchidaceae) through two chloroplast genes rbcL and matK. AoB PLANTS, 11(1): 1–12.

Kim, H. M., Oh, S. H., Bhandari, G. S., Kim, C. S., & Park, C. W. 2014. DNA barcoding of Orchidaceae in Korea. Molecular Ecology Resources, 14(3): 499–507.

Kress, W. J., & Erickson, D. L. 2008. DNA barcodes: Genes, genomics, and bioinformatics. PNAS, 105(8): 2761–2762.

Lahaye, R., Bank, M. Van Der, Bogarin, D., Warner, J., Pupulin, F., Gigot, G., Savolainen, V. 2008. DNA barcoding the floras of biodiversity hotspots. Proceedings of the National Academy of Sciences of the United States of America, 105(8): 2923–2928.

Li, M., Cao, H., But, P. P. H., & Shaw, P. C. 2011. Identification of herbal medicinal materials using DNA barcodes. Journal of Systematics and Evolution, 49(3): 271–283.

Li, X., Yang, Y., Henry, R. J., Rossetto, M., Wang, Y., & Chen, S. 2015. Plant DNA barcoding: from gene to genome. Biological Reviews of the Cambridge Philosophical Society, 90(1): 157–166.

Li, Y., Gao, L. M., Poudel, R. C., Li, D. Z., & Forrest, A. 2011. High universality of matK primers for barcoding gymnosperms. Journal of Systematics and Evolution, 49(3): 169–175.

Little, D. P. 2011. DNA barcode sequence identification incorporating taxonomic hierarchy and within taxon variability. PLoS ONE, 6(8): e20552.

Little, D. P. 2014. A DNA mini-barcode for land plants. Molecular Ecology Resources, 14(3): 437–446.

Meusnier, I., Singer, G. A. C., Landry, J. F., Hickey, D. A., Hebert, P. D. N., & Hajibabaei, M. 2008. A universal DNA mini-barcode for biodiversity analysis. BMC Genomics, 9: 4–7.

Neuhaus, H., & Link, G. 1987. The chloroplast tRNALys(UUU) gene from mustard (Sinapis alba) contains a class II intron potentially coding for a maturase-related polypeptide. Current Genetics, 11(4): 251–257.

Siripiyasing, P., KAenratana, K., Mokkamul, P., Tanee, T., Sudmoon, R., & Chaveerach, A. 2012. DNA barcoding of the Cymbidium species (Orchidaceae) in Thailand. African Journal of Agricultural Research, 7(3): 393–404.

Sosa, V., Mejía-Saules, T., Cuéllar, M. A., & Vovides, A. P. 2013. DNA Barcoding in Endangered Mesoamerican Groups of Plants. Botanical Review, 79(4): 469–482.

Srikulnath, K., Sawasdichai, S., Jantapanon, T. K., Pongtongkam, P., & Peyachoknagul, S. 2015. Phylogenetic relationship of Dendrobium species

in Thailand inferred from chloroplast matK gene and nuclear rDNA ITS region. Horticulture Journal, 84(3): 243–252.

Tanee, T., Chadmuk, P., Runglawan, S., Chaveerach, A., & Noikotr, K. 2012. Genetic Analysis for identification, genomic template stability in hybrids and barcodes of the Vanda species (Orchidaceae) of Thailand. African Journal of Biotechnology, 11(55): 11772–11781.

Teoh, E. (2016). Medical Orchid of Asia. Springer International Publishing.

Xu, S., Li, D., Li, J., Xiang, X., Jin, W., Huang, W., Huang, L. 2015. Evaluation of the DNA barcodes in Dendrobium (Orchidaceae) from mainland Asia. PLoS ONE, 10(1), 1–12.

Yu, J., Xue, J. H., & Zhou, S. L. 2011. New universal matK primers for DNA barcoding angiosperms. 49(3): 176–181.