Jurnal Biodjati 6(2):295–302, November 2021
e-ISSN : 2541-4208
p-ISSN : 2548-1606

JURNAL BIODJATI

http://journal.uinsgd.ac.id/index.php/biodjati

DEVELOPMENT OF DNA BARCODE FOR MAGNOLIOPSIDA AND LILIOPSIDA USING IN SILICO APPROACHES BASED ON mat-K SEQUENCES

Denia Dwi Citra Resmi[1], Topik Hidayat*[2], Siti Sriyati[3]

Selected paper from the 5th Seminar Nasional Biologi (SEMABIO), Bandung-Indonesia, July 15, 2021 (conference.bio.uinsgd.ac.id)
Received : 26 September, 2021
Accepted : 19 November, 2021

DOI: 10.15575/biodjati

[1, 2, 3]Department of Biology Education, Faculty of Mathematics and Science Education, Universitas Pendidikan Indonesia, Jl. Setiabudhi No 229 Bandung, 40154

e-mail:
[1]deniacitra2.dc@gmail.com,
*[2]topikhidayat@upi.edu,
[3]sriyati@upi.edu

**Corresponding author*

Abstract. *Indonesia has been estimated to contain 20,000 species of Magnoliophyta around the world. The current status of Indonesia's biodiversity shows that only 15.5% of the total flora in Indonesia has been identified. This is such a low percentage, requires researchers to obtain a rapid identification method, so that unidentified species can be grouped, at least at the level of the Magnoliopsida and Liliopsida classes. DNA barcoding is a technique that can be used to quickly identify species based on short sequences of specific regions in the genome. The purpose of this study was to analyze the relationship between Magnoliopsida and Liliopsida plants based on the mat-K marker and to obtain DNA barcodes for each of the Magnoliopsida and Liliopsida classes. This study used an in silico approach because the molecular data about these two selected classes with 101 species for samples are abundant in Genbank NCBI database. The primary design was carried out after analyzing the phylogenetic relationship between Magnoliopsida and Liliopsida. In silico analysis using BioEdit and PAUP to reconstructthe phylogenetic tree based on mat-K DNA showed results that were in line with previous studies. The phylogenetic tree using molecular data confirms that Magnoliopsida is the ancestor of Liliopsida. This study succeeded in obtaining two pairs of specific primers for Magnoliopsida and Liliopsida, which are cttcagtggtacggagtcaaat and gagccaaagttttagcacaagaa for Magnoliopsida, whereas cccatccatatggaaatcttggt and ttgaagccagaattgcttttcc for Liliopsida. These primers can later be used to distinguish the Magnoliopsida group from Liliopsida.*

Keywords: *DNA barcode, Liliopsida, Magnoliopsida, mat-K, Phylogenetics*

Citation
Resmi, D. D. C., Hidayat, T. & Sriyati, S. (2021). Development of DNA Barcode for Magnoliopsida and Liliopsida Using In Silico Approaches Based on mat-K Sequences. *Jurnal Biodjati*, 6(2), 295–302.

INTRODUCTION

Indonesia is a tropical archipelagic country with a total of 17,499 islands, 13,466 of them are verified islands, including 11,000 inhabited islands (Anonym, 2018). Indonesia consists of 2.1 million km of land area and about 5.8 million km of water area (Darajati et al., 2016). The abundance of biodiversity in Indonesia makes it known as a mega-biodiversity country. Furthermore, Indonesia is also part of the world's eighth most biodiverse country, besides Brazil, Colombia, Australia, Mexico, Madagascar, Peru, and

Jurnal Biodjati 6(2):295–302, November 2021

JURNAL BIODJATI

http://journal.uinsgd.ac.id/index.php/biodjati

China (Sutarno & Setyawan, 2015).

Around 25% or 20,000 species of Magnoliophyta (Anthophyta, Angiospermae) plants in the world are located in Indonesia, and 40% of them are Indonesian endemic plants. Although plant diversity in Indonesia is recorded in high numbers, the potential of genetic resources in plants has not been fully known (Kusmana & Hikmat, 2015). The latest status of Indonesia's biodiversity shows that only 15.5% of the total flora in Indonesia have been identified (Retnowati et al., 2019). For a long time, plant inventory, identification, and documentation activities have been a challenge for researchers. The main reason for finding a quick solution to the identification problem is the fear of the extinction of an organism (Cristescu, 2014).

Generally, plants are identified by experts based on limited fossil morphology and relatively inconsistent morphological characteristics because they are strongly influenced by the environment (Suparman, 2012). There is a technique that can be used to quickly identify species based on short sequences (DNA barcodes) for specific regions of the genome, known as DNA barcoding (Hebert et al., 2003). Until now, specific DNA barcodes have not been found for effectively identifying all plants because of the extremely high level of diversity in plants (Zhang & Jiang, 2020). The mat-K gene is widely used for taxonomic and phylogenetic studies in Magnoliophyta (Anthophyta, Angiospermae) (Samuel et al., 2006) and is commonly used as a barcode for land plants (Retnaningati, 2018). The sequence of mat-K gene attain a length of approximately 1600 bp, and that sequence is located between the trn-K introns of the chloroplast (Simpson, 2006). The mat-K gene has been compared with the rbcL gene, the substitution rate of the mat-K gene is three times higher specificallyat the nucleotide level, higher rate of evolution, and has more varied DNA sequences (Barthet & Hilu, 2007).

Currently, bioinformatics is one of the trends to process molecular biology data using software on the computers for biological science or biotechnology, such as identifying genus or species, looking for a relationship (phylogenetic), and/or searching for candidates of the new genes (Witarto & Sajidan, 2010). Research activities using an in silico approach requires relatively a short time by utilizing software on the computer and tend to spend low costs compared to other research approaches (Witarto & Sajidan, 2010). The low percentage of plants that have been identified gives the push to all plant researchers to find rapid method for plant identification, so that plants can have an identity and be one of the plant group members, at least at the class level of Magnoliopsida and Liliopsida. In this study, the development of DNA barcode for Magnoliopsida and Liliopsida plants using in silico approaches based on mat-K sequences from chloroplast genomes was obtained and potentially be an initial reference for identifying plants by grouping them into two classes, namely Magnoliopsida and Liliopsida before grouping at the next level of more specific taxa.

MATERIALS AND METHODS

The research was conducted from January to April 2021. Table 1 shows the tools and Table 2 shows the materials used in this study. The study began with preparing all the tools and the materials. The essential preliminary step to phylogenetic reconstruction and analysis was conducting the alignment of the sample sequences based on mat-K DNA using ClustalX 1.83. The most common method and also used in

this study was multiple sequence alignment (MSA). The next step was reconstructing the phylogenetic tree of Magnoliopsida and Liliopsida by analyzing it using PAUP 4.0b10 based on bootstrap values, constant characters, uninformative characters, informative characters, consistency index (CI), and retention index (RI) (Hidayat et al., 2021). Phylogenetic analysis used the maximum parsimony method. The phylogenetic tree file from PAUP 4.0b10 was also visualized using The TreeView program to make its structure clearer.

Table 1. The tools used in this study

Tools	Specification	Utility
Laptop	ASUS VivoBook Flip 14 TP410U, Intel® Core™ i3-7100U CPU	Process and analyze research data
Wi-Fi	Indihome 10 Mbps	Facilitates internet access in the research process

Table 2. The materials used in this study

Materials	Utility
Sequence of 101 plant samples of Magnoliopsida, Liliopsida, and outgroup using "matK complete sequence" as a keyword after write the name of species (NCBI GenBank https://www.ncbi.nlm.nih.gov/)	Using the samples for the study
Microsoft Excel	Forming the table information of 101 samples
NotePad	Creating the lists of DNA sequences in FASTA format
ClustalX 1.83	Aligning the whole sample DNA sequences
PAUP 4.0b10	Trimming, analyzing the cluster, and reconstructing the phylogenetic tree
TreeView	Visualizing the phylogenetic tree
BioEdit	Forming a consensus DNA sequences for Magnoliopsida and Liliopsida
FastPCR 6.7.77	Designing the specific primers for Magnoliopsida and Liliopsida, and doing the trials for both of primers

The next stage was how to get the candidates of primer for each class of Magnoliopsida and Liliopsida. Before designing some specific primers for each class, it is important to form consensus DNA sequences for Magnoliopsida and Liliopsida using BioEdit program. It is necessary to re-examine, check, and confirm these positions in the sequence electropherograms (Datta, 2018). In this study, re-examination, checking, and confirmation of nucleotide positions were performed in the BioEdit program manually based on the most common nucleotides in each position. After forming the consensus,

then the DNA consensus was processed using the FastPCR 6.7.77 program (Kalendar, Lee, & Schulman, 2014). PCR Primer Design in FastPCR 6.7.77 program is the featured used for designing primer. Furthermore, there is another feature used for in silico PCR trials using candidate of primers and DNA sample sequences based on the mat-K marker, named In silico PCR (Kalendar et al., 2014). One by one, all of the candidate primers were tested on DNA sample sequences. The results of the in silico PCR were presented in the Result Report column, it proved by the appearance or absence of a DNA segment called sample amplicon. The candidates of forward and reverse primer were paired one by one to be a primer pair for in silico PCR. The output from amplification using the candidates of primer pair were recorded in the Excel table, and percentage of the success rate was calculated. The results of this study were the pair of primer candidates with the highest percentage and specific for each class that could be used in PCR trials for plants of the class Magnoliopsida and Liliopsida.

RESULTS AND DISCUSSION

The results of the phylogenetic analysis in the PAUP 4.0b10 program, as many as1700 characters were produced, including 128 constant characters (8%), 278 uninformative characters (16%), and 1294 informative characters (76%). Moreover, there were 16 phylogenetic trees with a consistency index (CI) of 0.341 and a retention index (RI) of 0.649.

The most parsimonic tree which has the fewest number of the DNA base evolutions was used for this study (Campbell et al., 2008). The interpretation of bootstrap values in the phylogenetic tree (Figure 1) shows the high and moderate categories (75-100%). The phylogenetic reconstruction (Figure 1) was related to previous studies by Cronquist. In this study, the phylogenetic tree was built by molecular data, furthermore, the phylogenetic reconstruction was confirmed that Magnoliopsida is the ancestor of Liliopsida, and there seems to be no separation between Magnoliopsida and Liliopsida. Magnoliopsida hasformed the subclass Magnoliidae, and the lineage of Liliopsida comes from the basal of subclass Magnoliidae (Hussain, Verma, & Abdin, 2008).

The next stage was formed the consensus sequences to obtain candidates of primer for each class of Magnoliopsida and Liliopsida. From the result, there were other forms of the codes (besides ATGC) of nucleotides in the consensus sequences of Magnoliopsida and Liliopsida (Figure 2). However, the code was included in the IUPAC list of nucleotide codes. At the beginning of the experiment, consensus sequences with other codes were processed in the FastPCR program to obtain candidates of primers, but it turned out that these primers could not amplify most of the samples. Each candidate of Magnoliopsida primer was tested, and the results of the amplification test showed that the amplicon was only for *Theobroma cacao* and *Carica papaya*. Meanwhile, each candidate of Liliopsida primer could not amplify all of the samples.

Based on the Magnoliopsida consensus, 124 candidates of primers were obtained, including 64 forward and 60 reverse primers. Meanwhile, based on the Liliopsida consensus, 100 candidates of primers were obtained, including 53 forward and 47 reverse primers. Furthermore, the FastPCR program has another feature that can be useful to take in silico PCR tests for each primer of Magnoliopsida and Liliopsida. The results of the in silico PCR showed in the Result Report

Jurnal Biodjati 6(2):295–302, November 2021

JURNAL BIODJATI

http://journal.uinsgd.ac.id/index.php/biodjati

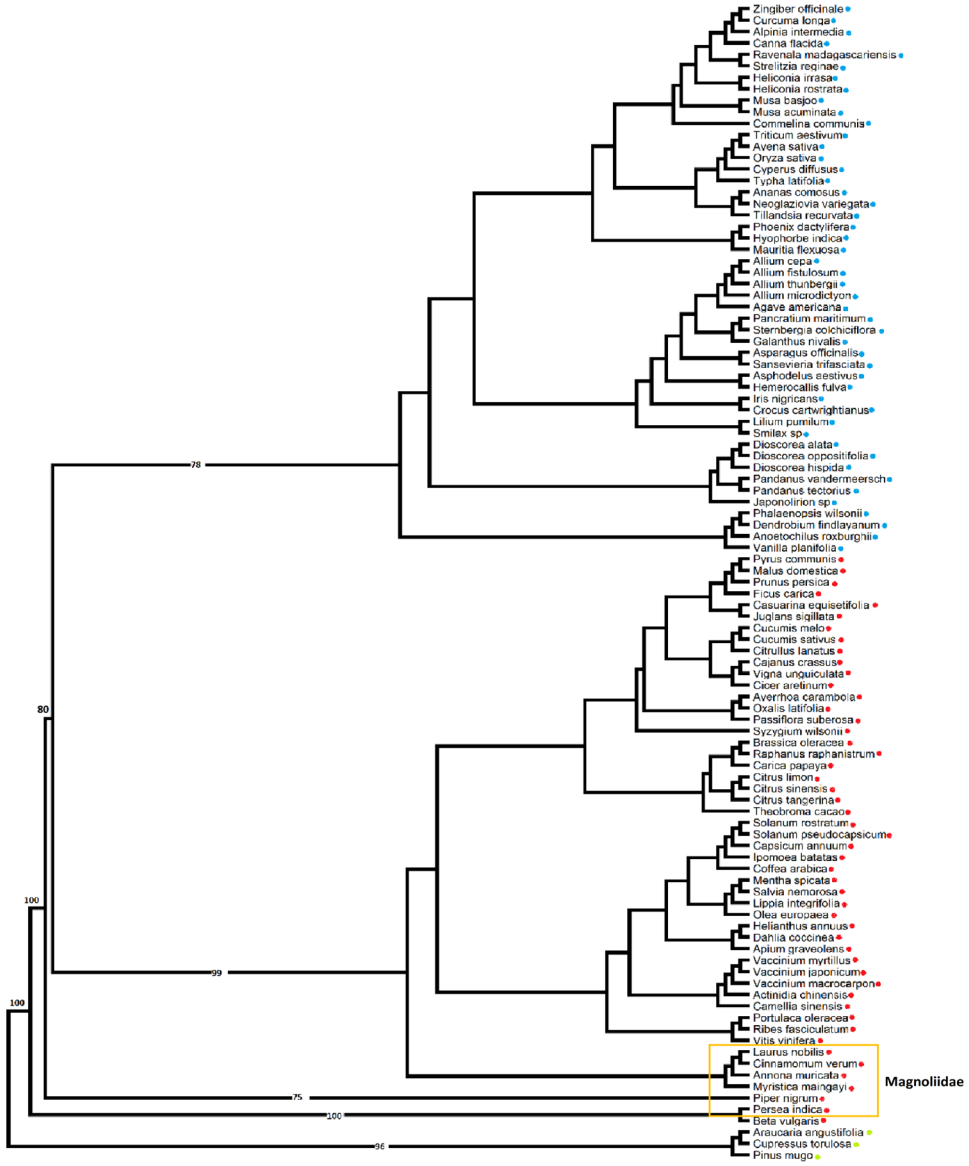

Figure 1. Phylogenetic tree reconstruction; Magnoliopsida, Liliopsida, Outgroup

Figure 2. Consensus sequences

column. A successful amplification process using each primer was indicated based on the amplicon that shows up in the Result Report column.

To test the specificity of the primer candidates in amplifying the target DNA according to its class, then the forward and reverse Magnoliopsida primer candidates were used to amplify the Liliopsida sample sequences and vice versa. If there is an amplicon that appears in the Result Report column, then the primers should be excluded from the candidates of forward and reverse primer pairs. Specific candidates of the forward and reverse primer were paired one by one. The next step is conducting in silico PCR process. Based on the results of in silico PCR, the candidate of forward and reverse Magnoliopsida primer were able to amplify 43 out of 50 species samples (86%). Meanwhile, the candidate of forward and reverse Liliopsida primer can amplify 40 of 48 species samples (83%). The following table shows the primer pairs of Magnoliopsida and Liliopsida along with information related to criteria of the primer:

From this study, it can be seen that phylogenetic reconstruction based on the mat-K gene was related to previous studies, Magnoliopsida is the ancestor of Liliopsida. Then, two primer pairs were obtained based on the mat-K gene, each for Magnoliopsida and Liliopsida. These results can contribute to identifying plants and grouping them into classes Magnoliopsida or Liliopsida. Trying to use the multiple data sets or combinations of DNA markers can potentially provide more accurate results.

Table 3. Primer pair of Magnoliopsida

Primer ID	Sequence (5'-3')	Length	Tm (in silico PCR)	Tm (consensus)	CG	Primer Quality	Amplicon Size
1F45_1_941-962	cttcagtggtacggagtcaaat	22	50.1 - 56.3°C	54.2°C	45.5%	90%	303 bp
1R5_1_1221-1243	gagccaaagttttagcacaagaa	23	50.7 - 54°C	54°C	39.1%	75%	

Table 4. Primer pair of Liliopsida

Primer ID	Sequence (5'-3')	Length	Tm (in silico PCR)	Tm(consensus)	CG	Primer Quality	Amplicon Size
1F16_1_424-446	cccatccatatggaaatcttggt	23	54.1-54.4	54.1	43.5%	75%	400 bp
1R24_1_804-825	ttgaagccagaattgcttttcc	22	45.7 - 54	54	40.9%	79%	

ACKNOWLEDGMENTS

The authors would like to acknowledge two anonymous reviewers for their valuable comments on the manuscript.

REFERENCES

Anonym. (2018). Indonesia, Negara Megabiodiversitas. Retrieved from https://indonesia.go.id/ragam/keanekaragaman-hayati/ekonomi/indonesia-negara-megabiodiversitas

Barthet, M. M. & Hilu, K. W. (2007).

Expression of matK: Functional and Evolutionary Implications. *American Journal of Botany*, *94*(8), 1402–1412.

Campbell, N. A., Reece, J. B., Urry, L. A., Cain, M. L., Wasserman, S. A., Minorsky, P. V, & Jackson, R. (2008). *Biology*. San Francisco: Pearson, Benjamin Cummings.

Cristescu, M. E. (2014). From Barcoding Single Individuals to Metabarcoding Biological Communities: Towards an Integrative Approach to the Study of Global biodiversity. *Trends in Ecology and Evolution*, *29*(10), 566–571.

Darajati, W., Pratiwi, S., Herwinda, E., Radiansyah, A. D., Nalang, V. S., Nooryanto, B. & Hakim, F. (2016). *Indonesia Biodiversity Startegy and Action Plan (IBSAP), 2012-2020*. Jakarta: BAPPENAS.

Datta, S. (2018). How Do I Edit Away These Strage Nucleotide Letters (N, K, Y, B, etc) from a Sanger Sequence DNA? Retrieved from https://www.researchgate.net/post/How_do_i_edit_away_these_strange_nucleotide_letters_N_K_Y_B_etc_from_a_Sanger_sequence_DNA_read/5af7213a35e538013a5ae352/citation/download

Hebert, P. D. N., Cywinska, A., Ball, S. L., DeWaard, J. R., Hebert, P. D. N. & Cywinska, A. (2003). Biological Identifications Through DNA Barcodes. *Proceedings of the Royal Society B: Biological Sciences*, *270*(1512), 313–321.

Hidayat, T., Saputro, N. W., Khamid, M. B. R., Bayfurqon, F. M., Hidayat, T., & Saputro, N. W. (2021). First Phylogenetic Treatment of Apple Cucumber (Family cucurbitaceae) from Indonesia Utilizing DNA Variation of Internal Transcibed Spacer Region. *HAYATI Journal of Biosciences*, *28*(1), 48–53.

Hussain, M. A., Verma, V. & Abdin, M. Z. (2008). Molecular Analysis of Dicot-Monocot Split and Relationship Among Major Angiosperm Groups. *African Journal of Plant Science*, *2*, 1–4.

Kalendar, R., Lee, D., & Schulman, A. H. (2014). FastPC Software for PCR, in Silico PCR, and Oligonucleotide assembly and analysis. *Methods in Molecular Biology (Clifton, N.J.)*, *1116*, 271–302.

Kusmana, C. & Hikmat, A. (2015). The Biodiversity of Flora in Indonesia. *Journal of Natural Resources and Environmental Management*, *5*(2), 187–198.

Retnaningati, D. (2018). Hubungan Filogenetik Intraspesies Cucumis Melo L. Berdasarkan DNA Barcode Gen matK. *Journal of Biota*, *2*(2), 62–67.

Retnowati, A., Rugayah, Rahajoe, J. S., Arifiani, D., Retnowati, A. & Rugayah. (2019). *Status Keanekaragaman Hayati Indonesia: Kekayaan Jenis Tumbuhan dan Jamur Indonesia*. Jakarta: LIPI Press.

Samuel, R., Gutermann, W., Stuessy, T. F., Ruas, C. F., Lack, H. W., Tremetsberger, K. & Ehrendorfer, F. (2006). Molecular Phylogenetics Reveals Leontodon (Asteraceae, Lactuceae) to be Diphyletic. *American Journal of Botany*, *93*(8), 1193–1205.

Simpson, M. G. (2006). *Plant Systematics*. London: Elsevier Academic Press.

Suparman. (2012). Markah molekuler dalam Identifikasi dan Analisis Kekerabatan Tumbuhan Serta Implikasinya Bagi Mata Kuliah Genetika. *Jurnal Bioedukasi*, *1*(1), 59–68. Sutarno, & Setyawan, A. D. (2015). Biodiversitas Indonesia:

Penurunan dan Upaya Pengelolaan untuk Menjamin Kemandirian Bangsa, *1*, 1–13.

Witarto, A. B. & Sajidan. (2010). *Bioinformatika: Trend dan Prospek dalam Pengembangan Keilmuan Biologi*. *Prosiding Seminar Biologi*, *7*(1), 15–16.

Zhang, D., & Jiang, B. (2020). Species identification in complex groups of medicinal plants based on DNA barcoding: a case study on Astragalus spp. (Fabaceae) from southwest China. *Conservation Genetics Resources*, *12*(3), 469–478.