TROPICAL GENETICS

Volume 3, No 2, November 2023
https://ojs.genetikawan-muda.com/index.php/tg

Original Research

Analysis of genetic variation of *MatK* gene sequences in *Ammothamnus lehmannii* NCBI Popset 2440747918 using in silico RFLP

Dara Suci Amini , Afifatul Achyar*

Department of Biology, Faculty of Science and Mathematics, Universitas Negeri Padang, 25131, Kota Padang

Corresponding author(s): : afifatul.achyar@fmipa.unp.ac.id

Article Info

Article history:
Received 26 Juni 2023
Accepted 24 Juli 2023
Available online 30 November 2021

Keywords:
Genetic Variation, MatK Genes, in-silico RFLP, HindIII

How to cite:
Amini and Achyar. 2023. Analysis of genetic variation of MatK gene sequences in Ammothamnus lehmannii NCBI Popset 2440747918 using in silico RFLP. *Tropical Genetics,* 3(2): 53-59

Abstract

Genetic diversity or genetic variation is variation that occurs in an organism due to differences in the sequence of nucleotide bases (adenine, thymine, guanine and cytosine) that form DNA in cells. Variation genetics can be studied in silico using available gene sequences in the NCBI genbank database. This study used the *MatK* (Maturase-K) gene sequence with the identity number Popset 2440747918 which was downloaded in fasta format from NCBI . Then screening of restriction enzyme candidates was carried out to determine the restriction enzymes prior to in silico RFLP. The restriction enzyme selected from the screening was the restriction enzyme *Hin*dIII which has the recognition site A'AGC_T. The results obtained from 79 samples of DNA sequences, 76 samples were cut and 3 samples were not. And found three allele variations with the percentage of the presence of fragments A1 (86.07%), A2 (10.12%) and A3 (3.79%). The percentage values and frequencies of these A1, A2 and A3 alleles indicate a low level of genetic variation.

Introduction

Genetics is the study of the ins and outs of genes as the basic biological units that control the inheritance of characteristics. Genetics can be seen from the perspective of individuals and populations. Population genetics is that part of genetics that looks at control a trait in the population (Sulastri et al., 2019). Knowledge of population genetic diversity can be used as information to understand the discussion of micro-level development that occurs at the genetic level (Arisuryanti & Daryono, 2007; Campbell *e*t al., 2003).

Genetic diversity or genetic variation is variation that occurs in an organism due to differences in the sequence of nucleotide bases (adenine, thymine, guanine and cytosine) that form DNA in cells (Harrison et al., 2004). Analysis of genetic variation using specific DNA sequence data is one way to study genetic diversity and genetic kinship of a species. Information about genetic diversity obtained from DNA analysis is also useful for determining kinship between individuals or populations studied. This information can be used as a basis for improving genetic quality (Suryanto, 2003).

Diversity at the population/individual/species level was analyzed with different markers. RAPD (Random Amplified Polymorphic DNA) is a marker that is often used. In addition there are

also Amplified Fragment Length Polymorphism (AFLP), Random Fragment Length Polymorphism (RFLP), Simple Sequence Repeats (SSRs), isozymes and allozymes. At the DNA sequence level, detection of genetic variation can be done by means of RFLP (Random Fragment Length Polymorphism) (Anggraini et al., 2021 ; Zulfahmi, 2013).

RFLP is a band corresponding to a DNA fragment, usually in the 2–10 kb range, that results from digestion of genomic DNA with restriction enzymes. DNA fragments were separated by agarose gel electrophoresis and detected by subsequent Southern blot hybridization to labeled DNA probes . The RFLP method is a restriction enzyme that cuts double-stranded DNA at a certain base sequence, so that fragments of varying lengths are formed (Septiasari et al., 2017; Theodore, 2000).

In this study, the species used to examine its genetic variation was *Ammothamnus lehmannii. Ammothamnus lehmannii* is a type of the genus Ammothamnus. Ammothamnus is a genus of flowering plants in the Fabacea family with a widespread distribution in Central Asia. This plant has the characteristics of a small shrub (50-80 cm high), has many branches, has a taproot system and has cream-colored flowers. This plant is often used as an insecticide against agricultural pests and also as a rodenticide and used as a potential ornamental plant.

The gene sequence used is the *MatK* (maturaseK) gene sequence (Syamsurizal et al., 2021). The *MatK* gene is a gene that encodes the maturase enzyme, which is part of the K subunit found in plant chloroplasts. This region of the nucleotide sequence of the matk gene can be produced around 1500 bp (base pairs) (Soltis et al ., 1998). The *MatK* gene is a maternally inherited chloroplast gene. Currently, the *MatK* gene has been used as an important tool to study genetic diversity within and between species due to its high substitution rate (Hollingsworth et al., 2011).

Currently, advances in science make it easy to identify individual genetic diversity molecularly et al., 2021). Variation genetics can be studied in silico using available gene sequences in the NCBI genbank database. In silico study is the term for experiments or tests carried out using computer simulation methods. In silico can be used as a condition assessment method for computer simulations with certain application programs or software (Khaira et al., 2022). In silico studies are preliminary studies before being assisted by other methods such as in vivo and in vitro which are useful for predicting and hypothesizing while reducing the need for expensive laboratory work (Achyar et al., 2021).

Based on the description above, this study aims to analyze the genetic variation of the MatK gene sequence in *Ammothamnus lehmannii* NCBI Popset 2440747918 using *In Silico* RFLP.

Material and Methods

1. Material

In this study, the *MatK* (Maturase-K) gene sequence was used with the identity number Popset 2440747918 which was downloaded in fasta format from NCBI . The name of the organism from this Popset number is *Ammothamnus lehmannii*, with an accession number ON996768 published by Liao,M., Shepherd,LD, Zhang,JY, Feng,Y., Mattapha,S., and Zhang,LB, Gao,XF and Xu,B. This popset has 79 gene sequences with accession numbers (Table 1.).

Table 1 . Sample Sequences of the *MatK* Gene *Ammothamnus lehmannii*

No.	Acc number	No.	Acc Number	No.	Acc Number
1.	ON996846.1	28.	ON996818.1	53.	ON996791.1

2.	ON996845.1	29.	ON996817.1	54.	ON996790.1
3.	ON996844.1	30.	ON996816.1	55.	ON996789.1
4.	ON996843.1	31.	ON996815.1	56.	ON996788.1
5.	ON996842.1	32.	ON996814.1	57.	ON996787.1
6.	ON996841.1	33.	ON996813.1	58.	ON996786.1
7.	ON996840.1	34.	ON996812.1	59.	ON996785.1
8.	ON996839.1	35.	ON996811.1	60.	ON996784.1
9.	ON996838.1	36.	ON996810.1	61.	ON996783.1
10.	ON996837.1	37.	ON996809.1	62.	ON996782.1
11.	ON996836.1	38.	ON996808.1	63.	ON996781.1
12.	ON996835.1	39.	ON996807.1	64.	ON996780.1
13.	ON996834.1	40.	ON996806.1	65.	ON996779.1
14.	ON996833.1	41.	ON996805.1	66.	ON996778.1
15.	ON996832.1	42.	ON996804.1	67.	ON996777.1
16.	ON996831.1	43.	ON996803.1	68.	ON996776.1
17.	ON996830.1	44.	ON996802.1	69.	ON996775.1
18.	ON996829.1	45.	ON996801.1	70.	ON996774.1
19.	ON996828.1	46.	ON996800.1	71.	ON996773.1
20.	ON996827.1	47.	ON996799.1	72.	ON996772.1
21.	ON996826.1	48.	ON996798.1	73.	ON996771.1
22.	ON996825.1	49.	ON996797.1	74.	ON996770.1
23.	ON996824.1	50.	ON996796.1	75.	ON996769.1
24.	ON996823.1	51.	ON996795.1	76.	ON996768.1
25.	ON996822.1	52.	ON996793.1	77.	ON996791.1
26.	ON996821.1	53.	ON996792.1	78.	ON996790.1
27.	ON996820.1	54.	ON996818.1	79.	ON996789.1

2. Methods
   a. Restriction Enzyme Candidate Screening Method
      The site used for screening restriction enzyme candidates is http://insilico.ehu.es/restriction . This site will feature various tools. In this method the tools used compare restriction patterns of many sequences. Then enter the fasta file from the gene sequence that we downloaded earlier then click go to next step. The alignment results will be visible, before clicking get lint of restriction enzymes, first click only restriction enzymes with known bases (no N ,R,Y). The next step is to choose the type of restriction enzyme that will be used in the next step.
   b. *In-silico* RFLP method
      The site used for in-silico RFLP is https://www.benchling.com/ . The first step is to create an account first. After that import the DNA sequences that we have downloaded from NCBI in fasta form. After the sequences appear, click on the first sequence, then restrict it by clicking on the scissor sign in the right corner. Find enzymes according to the name of the restriction enzyme that has been selected at the time of screening. Next, click "run digest" for the restriction. To see an image of the gel electrophoresis, click virtual digest. Do the same steps for the next sequence.

Results and Discussion

1. Screening of restriction enzyme candidates
   The results obtained after carrying out the enzyme screening were the *Hin*dIII restriction enzyme selected on the 5'-A'AGCTT-3' recognition site which inactivated at 80 $^{\circ}C$ and incubated at 37 $^{\circ}C$. This restriction enzyme was chosen because it only cuts once and does not have variations in cutting, making it easier for the next step. *Hin*dIII enzyme is a type II restriction enzyme isolated from *Haemophilus influenza.* This enzyme is found in prokaryotic organisms such as bacteria. Bacteria use this enzyme as a defense mechanism against viruses such as bacteriophages.
   Restriction enzymes are enzymes that can cut DNA genomes to obtain DNA sequences by breaking them into fragments by enzymes that can break the DNA phosphodiester chains . Restriction enzymes produced by bacteria are also called endonucleases. This enzyme can recognize specific nucleotide sequences of 4-8 bp. Restriction enzymes require optimal conditions to cut completely and produce well-fitting restriction fragments . These optimal conditions are temperature, pH buffer, digestion time, and salt concentration of the buffer used (Putri et al., 2017; Gerstein, 2001). This specific nucleotide sequence is called a restriction site, which is usually a short palindromic sequence and the same sequence pattern when read in the 5'→3' direction (Muthiadin , 2014).
2. RFLP *in-silico*
   RFLP is the oldest DNA analysis and produces black and white images. In simple terms, this technique cuts DNA with restriction enzymes (enzymes that break base bonds in the DNA chain) by targeting certain sequences into strands of different lengths. And the length is separated by a special gene (usually the gel matrix comes from agarose) (Winaya, 2017). RFLP is useful for detecting variations at the level of DNA sequences that use the ability of restriction enzymes to DNA. With the RFPL method, the similarity or variability of genes can be known (Wavedi, 2015). The development of science and technology is currently very

profitable in scientific fields such as genetics. The presence of various kinds of bioinformatics *tools* makes the screening process of restriction enzymes and visualization of restriction fragments can be carried out in *silico.*

In-silico RFLP on 79 *MatK* gene sequences in *Ammothamnus lehmannii* NCBI Popset 2440747918 was carried out in the Benchling application using a restriction enzyme that was already at the restriction enzyme candidate screening stage, namely the *Hin*dIII enzyme. *In silico* RFLP results were visualized by virtual gel electrophoresis as shown in Figure 1 for restriction with HindIII enzymes.

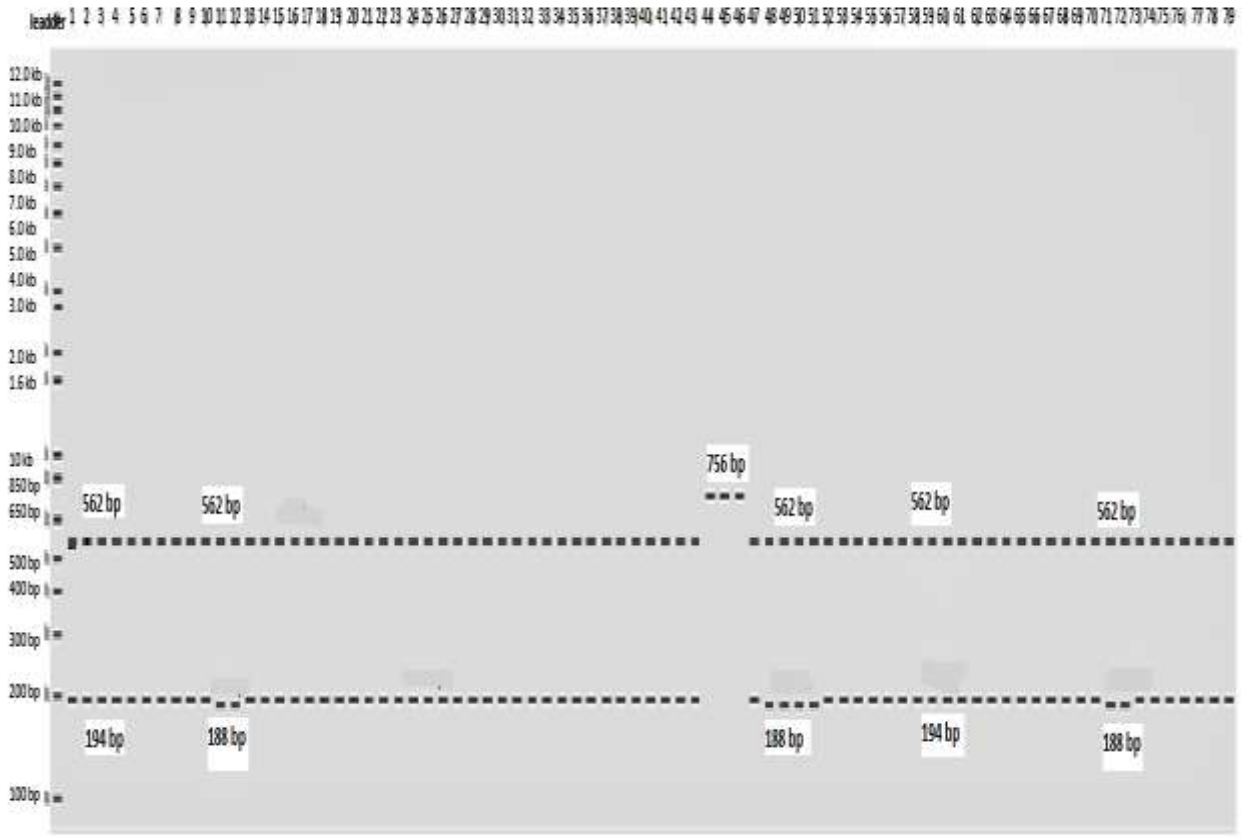

Figure 1 . An electrophorogram of the restricted *MatK* gene sequence using the in-silico restriction enzyme *Hin*dIII. Note : left (Leader Bioline 1 kb), (1) ON996846.1 (2) ON996845.1 (3) ON996844.1 (4) ON996843.1 (5) ON996842.1 (6) ON996841.1 (7) ON996840.1 (8) ON996839.1 (9) ON996838.1 (10) ON996837.1 (11) ON996836.1 (12) ON996835.1 (13) ON996834.1 (14) ON996833.1 (15) ON996832.1 (16) ON996831.1 (17) ON996830.1 (18) ON996829.1 (19) ON996828.1 (20) ON996827.1 (21) ON996826.1 (22) ON996825.1 (23) ON996824.1 (24) ON996823.1 (25) ON996822.1 (26) ON996821.1 (27) ON996820.1 (28) ON996818.1 (29) ON996817.1 (30) ON996816.1 (31) ON996815.1 (32) ON996814.1 (33) ON996813.1 (34) ON996812.1 (35) ON996811.1 (36) ON996810.1 (37) ON996809.1 (38) ON996808.1 (39) ON996807.1 (40) ON996806.1 (41) ON996805.1 (42) ON996804.1 (43) ON996803.1 (44) ON996802.1 (45) ON996801.1 (46) ON996800.1 (47) ON996799.1 (48) ON996798.1 (49) ON996797.1 (50) ON996796.1 (51) ON996795.1 (52) ON996793.1 (53) ON996792.1 (54) ON996818.1 (55) ON996791.1 (56) ON996790.1 (57) ON996787.1 (58) ON996786.1 (59) ON996785.1 (60) ON996784.1 (61) ON996783.1 (62) ON996782.1 (63) ON996781.1 (64) ON996780.1 (65) ON996779.1 (66) ON996778.1 (67) ON996777.1 (68) ON996776.1 (69) ON996775.1 (70)

ON996774.1 (71) ON996773.1 (72) ON996772.1 (73) ON996771.1 (74) ON996770.1 (75) ON996769.1 (76) ON996768.1 (77) ON996791.1 (78) ON996790.1 (79) ON996789.1

Restriction with the restriction enzyme *Hin*dIII on 79 *MatK* gene sequences in *Ammothamnus lehmannii* produced three allele variations, namely the A1 allele which produced two DNA strands of 562 bp and 194 bp in size. The A2 allele produces two DNA strands of 562 bp and 188 bp. Meanwhile, the A3 allele only produces one DNA band with a size of 756 bp. This is because the A3 allele does not have a *Hin*dIII restriction recognition site so no cutting occurs by the restriction enzyme so that the length of the fragment remains intact, namely 756 bp, while the A2 and A3 alleles have a *Hin*dIII restriction recognition site (A'AGC_T) and a cutting site at the 562nd base. so that it will show 2 strands of DNA. The A1 allele has a higher allele frequency than the A2 and A3 alleles because it has an attendance percentage of 86.07% of the 79 sequences in NCBI Popset 2440747918. The allele frequencies in the *MatK* gene sequence with the restriction enzyme *Hin*dIII can be seen in Table 2.

Table 2. Allele Frequency of the *MatK* Gene in *Ammothamnus lehmannii* NCBI Based on In Silico RFLP Results

Restriction Enzyme	Restriction recognition site	Fragment size (bp)	alleles	Number of fragments present (N=79)	Percentage of presence of fragments (%)	Allele frequency
*Hin*dIII	A'AGC_T	562 and 194	A1	68	86.07%	0.8607
		562 and 188	A2	8	10.12%	0.1012
		756	A3	3	3.79%	0.0379

We can see from table 2 that the A1 allele dominates the population in Popset 2440747918. Of the 79 samples, the amount of DNA that was cut was 76 samples while the other 3 samples were not cut. The A1 allele had the highest number of fragments present, namely 68 times with a percentage of 86.07%, followed by the A2 allele 8 times with a percentage of 10.12% and the A3 allele 3 times with a percentage of 3.79%. The percentage values and frequencies of these A1, A2 and A3 alleles indicate a low level of genetic variation.

Conclusion

The in silico RFLP results in this study indicate that the MatK gene sequence in Ammothamnus lehmannii NCBI Popset 2440747918 has three genetic variations on the recognition side of the HindIII restriction enzyme, with low genetic variation.

References

Achyar, A., Alvenaya H., Fadhila H., et al. 2021. Analysis of Genetic Variations in POLY Gene Sequences in Dengue Virus 2 Using In-Silico RFLP. *Bioscience*, 5 (1), pp. 80-86 ISSN: Print 1412-9760 – Online 2541-5948.

Anggraini, Serli Febri, Ayunda, R.K., Deratih Bunga, Marisa Pela., Vira Fill Jannah, Wulan, P.P., dan Afifatul Achyar. Analysis of Genetic Variant Of NADH Dehydrogenase Subunit 4-Like Gene

Sequences of Aedes Aegypti Popset: 1831566147 Using In Silico RFLP. *Prosiding Semnas Bio*, Vol 1 : 102-110.

Arisuryanti.,T dan Daryono.,B.S. (2007). *Genetika Populasi.* Yogyakarta: Fakultas Biologi Universitas Gadjah Mada.

Harrison, I., M. Laverty and E. Sterling. 2004. Genetic Diversity. *Connexions modul*e: m12158.

Hollingsworth, P. M., Graham, S. W., and Little, D.P. 2011. Choosing and Using a Plant DNA Barcode. *PLoS one*, 6: e19254.

Khaira, Annisa., Livia Juniarti, Nafisa Arini, Rinti Mutiara Sari, dan Afifatul Achyar. 2022. Analisis variasi genetik Sapi (Bos Taurus) pada sekuen gen Cytochrome Oxidase Subunit 1(COI) menggunakan RFLP in silico. *Prosiding Semnas Bio*, 2809-8447.

Muthiadin, Cut. 2014. *Pengantar Rekayasa Genetika*. Universitas Islam Negeri Alauddin. Makassar.

Putri, I.A.R.D.N., Sagung, C.Y., dan I Nengah Wirajana. 2017. Optimasi Digesti Enzim Restriksi SacII Pada Isolat *Mycobacterium tuberculosis* H37Rv Untuk Deteksi Mutasi Promoter inhA Pada Kasus MDR-TB Dengan Metode PCR-RFLP. *Jurnal Metamorfosa*, IV (1), 87-93.

Putri, K. H. A., Kartika Puspita Sari, Khairunnisa, Monicha Yhuyhen Safitri, Afif Putra, Katon Agusdi, Fahrul Rozi Oktavian, dan Afifatul Achyar. (2021). Analisis Variasi Genetik Sekuen Gen Surface Glycoprotein (S) Pada SARS –CoV-2 Popset : 1843471817 Menggunakan RFLP Secara In-Sillico. *Prosiding Seminar Nasional Biologi*, 1(1), 44–52.

Septiasari, Ni Putu Senshi., I Ketut, Junitha, dan Ni Nyoman, Wirasiti,. 2017. Ragam Alel Dna Mitokondria Masyarakat Soroh Pande di Bali Dengan Metode PCR-RFLP. *Jurnal Metamorfosa*, Vol IV (2) : 210-217.

Sulastri., Hamdani, M.D.I dan Akhmad, Dakhlan. *Dasar Pemuliaan Ternak*. Bandar Lampung : CV. Anugrah Utama Raharja.

Soltis, P. S., D. E. Soltis, and J. J. Doyle. 1998. *Molecular Systematics of Plants*. International Thomson Publishing, New York.

Syamsurizal, S., Ardi, A., Des, M., Fevria, R., Atifah, Y., Badriyya, E., & Achyar, A. (2021). Design of primer Ipomoea batatas chloroplast gene matK. *Tropical Genetics*, *1*(1), 12–16.

Wahyudi, Ivan Arie. 2015. Resensi Biologi Molekular Adalah Ilmu yang Menyenangkan dan Mudah. *Jurnal Teknosains,* Vol 4 : 101-198.

Winaya, Aris. 2017. Teknologi Analisis DNA Dalam Penguatan Strategi Konservasi Genetik Sapi Lokal Indonesia. *Prosiding Seminar Teknologi dan Agribisnis Perternakan V : Teknologi dan Agribisnis Perternakan Untuk Mendukung Ketahanan Pangan*, Fakultas Perternakan Universitas Jenderal Soedirman, Purwokerto: 18 November 2017.

Yuliani, Yani., Alice Yuniaty, dan Agus Hery Susanto. 2017. Variasi Sekuen DNA yang Diamplifikasi Menggunakan Primer ATPB-RBCL Pada Beberapa Kultivar Kacang Tanah. *Scripta Biologica*, Vol 4, No. 1.