TROPICAL GENETICS

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

Original Research

Analysis of genetics variation of the ndhF gene sequence in *Antrophyum* sp. NCBI popset 2496377569 using in silico RFLP

Hafizah Fadhilah, 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 25 Juni 2023
Accepted 10 November 2023
Available online 30 November 2023

Keywords:
Antrophyum sp., ndhF gene, RFLP, in silico, genetics variations

How to cite:
Fadhilah and Achyar. 2023. Analysis of genetics variation of the ndhF gene sequence in Antrophyum sp. NCBI popset 2496377569 using in silico RFLP. *Tropical Genetics,* 3(2): 39-45.

Abstract

Antrophyum is one type of the largest and most diverse of the vittarioid ferns (Pteridaceae) in tropical Asia and the Pacific Islands. The data found in the *Antrophyum* study is still relatively sparse because the diversity of the genus is still uncertain. ndhF gene is an inherited gene from chloroplast DNA (NADH dehydrogenase five). Chloroplast DNA controls the production of ribosomal RNA (rRNA), transfer RNA (tRNA), and most of the proteins contained in the organelles chloroplast. Genes are commonly used in molecular studies such as the analysis of genetic variation in plants.. Genetic variation indicates that the greater the genetic variation that occurs in a population, the better its ability to adapt to the environment.. Besides that, determining genetic variation can also prevent extinction of a plant. This study aims to analyze genetic variation in the ndhF gene sequence of *Antrophyum* sp. NCBI PopSet 2496377569. In this study, the RFLP molecular marker was used which was carried out in silico using two restriction enzymes, PciI and ScaI.. The results of this study showed that there was no genetic variation in digestion with the restriction enzyme PciI, whereas in ScaI there was genetic variation with 3 allele variations from 23 PopSet sequences 2496377569 in NCBI.

Introduction

Antrophyum is one of the largest and most diverse genera of vittarioid ferns (Pteridaceae) in tropical Asia and the Pacific Islands, but also occurs in temperate Asia, Australia, tropical Africa, and the Malagasy region (Chen et al., 2023). To date, more than 80 basionyms have been published for *Antrophyum* (within its current limitations) and this number is more than double the generally accepted number of estimates of around 40 accepted species (Schuettpelz et al, 2016). Recent studies have integrated morphological and molecular data to explore well-focused genus diversity in specific geographic regions (Chen et al., 2017). On a global scale, studies on *Antrophyum* are lacking because the diversity of the genus is still uncertain.

The target gene used for genetic variation analysis in this study was the ndhF (NADH dehydrogenase five) gene. This gene comes from chloroplast DNA which can be used for molecular activities (Virgilio, 2012). Chloroplast DNA is the DNA that controls the production of ribosomal RNA (rRNA), transfer RNA (tRNA) and most of the proteins found in chloroplast organelles. (Tamarins, 2001). Chloroplast DNA is often used in Angiospermae plants such as *Antrophyum* to study their level of kinship. This is because chloroplast DNA has a small genome size, a stable structure, comes from

maternal or maternal genetic descent, a low level of evolution and a low level of genetic recombination, so that when compared to nuclear and mitochondrial DNA, chloroplast DNA is better used in molecular activities (Jing, et al., 2011). However, according to Harsono et al., (2015) chloroplast DNA has a weakness in the form of limitations in observing intraspecific relationships.

According to Kusuma et al. (2016) genetic diversity is a variation that occurs as a result of variation between individuals in a population. Genetic variation can occur when offspring receive a unique combination of chromosomes and genes from their parents through gene recommendations that occur through sexual reproduction, and cause alleles to rearrange randomly to give rise to a variety of different combinations (Afriani al., 2022). The higher the genetic variation found in a population, the better the individual's ability to adapt to environmental changes (Achyar et al., 2021).

Genetic variation analysis can be performed using the RFLP (Restriction Fragment Length Polymorphism) method. RFLP is a molecular marker that uses restriction enzymes to identify DNA sequences (Kanaya, 2023). This method uses the specificity of restriction enzymes, which only cut double-stranded DNA at specific cutting sites. Cutting restriction enzymes will produce DNA fragments of varying lengths and the results of these fragments can be visualized through electrophoresis (Nariska et al., 2022).

Restriction and visualization of restriction fragments with the aim of predicting genotyping results before conducting RFLP in the laboratory can be done in silico (Kanaya et al., 2023). In silico test is an experimental term or test carried out using a computer simulation method (Yeriska et al., 2021).

Based on the description above and armed with gene sequence data from the NCBI bioinformatic database, the purpose of this study was to analyze genetic variation in the ndhF gene sequence in *Antrophyum* sp. NCBI PopSet 2496377569 using in silico RFLP.

Material and Methods

1. Material
   The *Antrophyum* sequence used was the Vittarioideae NADH dehydrogenase subunit F (*ndhF*) gene with PopSet identity number 2496377569 which was downloaded in FASTA format from NCBI https://www.ncbi.nlm.nih.gov/popset/?term=2496377569. This PopSet originally submitted by Chen,C.W., Lindsay,S., Nitta,J., Rouhan,G., Sundue,M., Perrie,L.R., Huang,Y.M., Chiou,W.L. and Chung,K.F. in their study entitled "Systematics and biogeography of the Old World fern genus Antrophyum". In the PopSet there are 64 sequences, but after alignment, only 23 sequences will be used in this study.
2. Methods
   a. Multiple Sequence Alignment Method
      MEGA 11 is a genetic analysis application in sequence data processing that implements certain statistical analysis and algorithms. One of its uses is to align sequences. Software can be downloaded at the link https://www.megasoftware.net/. The FASTA file obtained from the reference search results using PopSet 2496377569 on NCBI is the main data source. After getting the FASTA format then input the sequence file. The next step is to select the sequences to align then click the alignment tools, select Align by clusterW for multiple alignment. Next, the Settings Options display for alignment will appear. Click OK to run. Trim the left and right then discard short sequences and have "N" repeated for a good alignment. Finally save in FASTA form for the next stage.
   b. Screening Candidate Restriction Enzyme Method
      Screening for restriction enzyme candidates can be done using In Silico with the website http://insilico.ehu.es/restriction/ In this method the selected tool is to compare restriction patterns of many sequences. The next step is to enter the sequence file that has been aligned in multiple alignments in FASTA form on the NCBI website. Then click

Go to the next step, then Only restriction enzymes with known bases (no N, R, Y...) are selected to get restriction enzymes whose bases are absolutely clear, there are no N, R or Y bases. The next step is selecting a list of restriction enzymes so that candidate restriction enzymes are obtained which will be used in the next stage (Aulia et al., 2021).

c. RFLP Method by In silico
RFLP is done virtually using the https://www.benchling.com/ site. This site is a free site with the first step is to register first. If you have already registered, proceed with importing the aligned RNA sequences into the project folder on the Benchling site. Then make the cut by clicking on the scissor (digest) sign in the right corner. Then the "find enzyme" tool will appear and select the names of the restriction enzymes that have been previously determined during screening. The next step is to click the "run digest" menu for the restriction. To view the electropherogram, you can use the "Virtual digest" tool (Kanaya et al., 2023).

Results and Discussion

1. Multiple Sequence Alignments
Sequence alignment method is a way of arranging deoxyribonucleic acid (DNA), ribonucleic acid (RNA), or protein sequences to identify areas of similarity that may be a consequence of functional, structural, or evolutionary relationships between sequences (Sing et al, 2011) .
In this study, the sequences downloaded from the NCBI website required multiple sequence alignment using MEGA11 because the initial size of each sequence was different. Sequence alignment in *Antrophyum* using the ndhF gene found a mismatch in the alignment between species which was marked by a gap (–) sign associated with the insertion or deletion process (de Groot et al., 2011). Therefore, sequences that have been aligned multiple still need to be tidied up by removing the initial part until the same nucleotide base is found in each sequence. This is also done at the end by removing different parts until it encounters the same nucleotide base in each sequence. Sequences with short sizes and repeated "N" were removed, after multiple alignment and trimming, 23 sequences were used with acc numbers and species names as shown in Table 1.

Table 1. Samples of the *ndhF* gene sequence in *Antrophyum* sp. NCBI

Acc Number	Species Name	Acc Number	Species Name
OQ440596.1	*Antrophyum ledermannii*	OQ440609.1	*Antrophyum callifolium*
OQ440597.1	*Antrophyum ledermannii*	OQ440656.1	*Antrophyum brookei*
OQ440616.1	*Antrophyum nanum*	OQ440620.1	*Antrophyum sessilifolium*
OQ440618.1	*Antrophyum nanum*	OQ440657.1	*Antrophyum hovenkampii*
OQ440621.1	*Antrophyum nanum*	OQ440611.1	*Antrophyum annamense*
OQ440608.1	*Antrophyum immersum*	OQ440655.1	*Antrophyum nambanense*
OQ440598.1	*Antrophyum castaneum*	OQ440610.1	*Antrophyum henryi*
OQ440624. 1	*Antrophyum novae-caledoniae*	OQ440622.1	*Antrophyum semicostatum*
OQ 440606.1	*Anthrophyopsis boryana*	OQ440631.1	*Anthrophyopsis manniana*
OQ440637.1	*Anthrophyopsis bivittate*	OQ440612.1	*Antrophyum vittarioides*
OQ440595.1	*Antrophyum aff. callifolium*		
OQ440601.1	*Antrophyum aff. callifolium*		
OQ440602.1	*Antrophyum callifolium*		

2. Screening Candidate Restriction Enzyme
Based on the results of In silico in screening restriction enzyme candidates, there are 104 restriction enzyme recognition sites. From these enzymes, *Pci*I restriction enzymes with A'CATG_T recognition sites and *Sca*I restriction enzymes with AGT'ACT recognition sites were

selected. These two restriction enzymes were chosen because they have variations in the cutting sites in each sequence.

The *Pci*I enzyme is a digestive restriction enzyme derived from an *E. coli* strain carrying the *Pci*I gene cloned from *Planococcus citreus* SE-F45 (S.K. Degtyarev) (New England Biolabs, 2023). The *Sca*I enzyme comes from *Streptomyces caespitosus*, inactivates at 80°C and incubates at 37°C (Kanaya et al., 2023).

3. RFLP by In Silico

RFLP (Restriction Fragment Length Polymorphism) is a technique that functions as an estimator of DNA variation. Variation is detected in the form of polymorphic (double) long sequence cleavage where the assessment of the variation sequence is possible from the data of the fragment itself, the long series variation in a section can be assessed from nucleotide substitutions (Fatchiyah et al., 2011).

The results of the analysis using the In silico RFLP technique with 23 *ndhF* gene sequences in *Antrophyum* sp. using the Benchling application with the restriction enzyme that has been selected, namely *Pci*I can be seen in Figure 1 and the restriction enzyme *Sca*I can be seen in Figure 2 which has been visualized by virtual gel electrophoresis. .

Figure 1. In silico restriction electropherogram with *Pci*I enzyme. note: Left (Ladder Life 1 kb Plus), (1) OQ440596.1, (2) OQ440597.1, (3) OQ440616.1, (4) OQ440618.1, (5) OQ440621.1, (6) OQ440608.1, (7) OQ440598.1, (8) OQ440624.1, (9) OQ440606.1, (10) OQ440637.1, (11) OQ440595.1, (12) OQ440601.1, (13) OQ440602.1, (14) OQ440609.1, (15) OQ440656.1, (16) OQ440620.1, (17) OQ440657.1, (18) OQ440611.1, (19) OQ440655.1, (20) OQ440610.1, (21) OQ440622.1, (22) OQ440631.1, (23) OQ440612.1.

Figure 2. In silico restriction electropherogram with *Sca*I enzyme. note: Left (Ladder Life 1 kb Plus), (1) OQ440596.1, (2) OQ440597.1, (3) OQ440616.1, (4) OQ440618.1, (5) OQ440621.1, (6) OQ440608.1, (7) OQ440598.1, (8) OQ440624.1, (9) OQ440606.1, (10) OQ440637.1, (11) OQ440595.1, (12) OQ440601.1, (13) OQ440602.1, (14) OQ440609.1, (15) OQ440656.1, (16) OQ440620.1, (17) OQ440657.1, (18) OQ440611.1, (19) OQ440655.1, (20) OQ440610.1, (21) OQ440622.1, (22) OQ440631.1, (23) OQ440612.1.

Based on the results of the restriction electropherogram using the *Pci*I enzyme from 23 sequences (Figure 1) it showed that there was no allelic variation that occurred, so there was only one allele. The bands produced by this allele are 2 bands with band lengths of 104 bp and 991 bp.

The results of the restriction on the *Sca*I enzyme Figure 2 shows that there are 3 allele variations, namely the B1 allele produces one band with a size of 1095 bp. This happens because this allele does not cut the allele by the *Sca*I enzyme so that it only produces one DNA strand.The B2

allele produced 2 bands with lengths of 120 bp and 975 bp, in this allele the *Sca*I enzyme was cut once. Whereas in B3 the *Sca*I enzyme was cut 2 times so that it showed that there were 3 DNA bands with lengths of 120 bp, 453 bp and 522 bp.

Table 2. Allele Frequency of the *ndhF Antrophyum* sp. NCBI Based on In Silico RFLP Results

Enzyme Restriction	Recognition Site	Size (bp)	Alleles	Amount Fragment Presence (N=23)	Percentage Fragment Presence (%)	Alleles Frequency
*Pci*I	AGT'ACT	104 991	A1	23	100	1
*Sca*I	AGT'ACT	1095	B1	13	56.52	0.5652
		120 975	B2	8	34.79	0.3479
		120 453 522	B3	2	8.69	0.0869

Based on Table 2, it is known that the *Pci*I restriction enzyme with the AGT'ACT recognition site has one allele frequency (A1) with a band length of 104 and 991 with the number of fragments present being 23 because they are present in all sequences, so that the percentage of presence of fragments is 100%. Whereas in the restriction enzyme *Sca*I with the AGT'ACT recognition site, there are 3 variations of the B1 allele with a band length of 1095, in this allele there is no cutting of the *Sca*I enzyme so there is only one DNA band with the highest number of fragments present as many as 13 sequences, the percentage of presence of fragments is 56.52%. and allele frequency 0.5652. In the B2 allele, there are two DNA bands with a length of 120 bp and 975 bp with the number of fragments present in 8 sequences, the percentage of presence is 34.79% and the allele frequency is 0.3479. Furthermore, the B3 allele has 3 bands with lengths of 120 bp, 453 bp, and 522 bp. The number of fragments present in 2 sequences with a percentage of 8.69% and an allele frequency of 0.0869 (Syamsurizal, 2019).

Conclusion

In silico RFLP results with restriction enzymes *Pci*I and *Sca*I showed that in the restriction enzyme *Pci*I there was no variation because only one allele appeared, whereas in the restriction enzyme *Sca*I there was genetic variation with the appearance of 3 alleles in 23 *ndhF* gene sequences in *Antrophyum* sp. with PopSet number 2496377569 from NCBI.

References

Achyar, A., Hindayageni, A., Humaira, F., Wijaya, N.N., Aqsha, N., Zultsatunni'mah. 2021. Analysis of Genetic Variations in POLY Gene Sequences in Dengue Virus 2 Using In-Silico RFLP. *Jurnal Bioscience*. Vol. 05. No. 1: 80-86.

Afriani, T., Purwati, E., Hellyward, J., Jaswandi, J., Yurnalis, Y., Mundana, M., & Rastosari, A. (2022). Identifikasi Keragaman Gen FSH Bagian Ekson 2 Menggunakan Enzim Restriksi TasI pada Sapi Pesisir. *Jurnal Peternakan*, *19*(2), 86-95.

Aryani, N. L. T., Suratmi, S., & Saifuddin, S. (2017). TEKNIK Restriction Fragment Length Polymorphism (RFLP) PADA IKAN NAPOLEON (Cheilinus undulatus). *Buletin Teknik Litkayasa Akuakultur*, *8*(2), 181-184.

Aulia, N., Putri, C. K., Khairiyah, L., Khairunnisa, F. A., & Achyar, A. (2022, May). Analysis of Genetic Variation of Measles Virus Genotype and Isolate MF Intergenic Spacer Sequences PopSet: 2105287799 using Restriction-Fragment Length Polymorphism (RFLP) In Silico. *InProsiding Seminar Nasional Biologi* (Vol. 1, No. 2, pp. 1139-1146).

Chen, C. W., Lindsay, S., Nitta, J., Rouhan, G., Sundue, M., Perrie, L. R., ... & Chung, K. F. (2023). Systematics and biogeography of the Old World fern genus Antrophyum. *Cladistics*.

Chen, C.-W., Lindsay, S., Kuo, L.-Y., Fraser-Jenkins, C.R., Ebihara, A., Luu, H.T., Park, C.W., Chao, Y.-S., Huang, Y.-M. and Chiou, W.-L. (2017). *A systematic study of east Asian vittarioid ferns (Pteridaceae, Vittarioideae)*. Bot. J. Linn. Soc. 183, 545–560.

de Groot, G. A., During, H. J., Maas, J. W., Schneider, H., Vogel, J. C., & Erkens, R. H. J. 2011. Use of rbcL and trnL-F as a two-locus DNA barcode for identification of NW-European ferns: An ecological perspective. *PLoS ONE*. vol. 6(1): 1- 10.

Dunham, RA. (2004). Aquaculture and Fisheries Biotechnology Genetic Approaches. USA: Department of Fisheries and Allied Aquacultures Auburn University Alabama.

Harsono, T., Pasaribu, N., Sobir, Firmawati, & Prasetya, E. (2015). Variasi intraspesifik berdasarkan DNA kloroplas (cpDNA) pada Bouea macrophylla Griffith. *Prosiding Seminar Nasional Biodiversitas Dan Ekologi Tropika Indonesia*.

Jing, Y. U., Jiang-Hua, X. U. E., & Liang, Z. H. O. U. (2011). New universal matK primer for DNA barcoding angiosperm. *Journal of Systematic and Evolution*, 49(3), 176-181.

Kanaya, O. N. K. N., & Achyar, A. (2023). Analysis of Genetik Variations in PHT1 Gene Sequences in Rice (Oryza sativa) NCBI Popset 240028097 Using In-Silico RFLP. *Jurnal Serambi Biologi*, *8*(1), 5-9.

Kusuma, R., N. Sa'diyah, & Y. Nurmiaty. 2016. Keragaman fenotipe dan heritabilitas kedelai (Glycine max [L.] Merrill) generasi F5 hasil persilangan Wilis x B3570. *Jurnal Penelitian Pertanian Terapan*.15 (3): 200-207.

Nariska, H., Lestari, P., & Karima, R. (2022). Optimasi Metode Polymerase Chain Reaction-Restriction Fragment Length Polymorphism (PCR-RFLP) untuk Deteksi Polimorfisme Gen MTHFR (C677T) pada Pasien Leukemia Limfoblastik Akut (LLA) Anak. *Prosiding Asosiasi Institusi Pendidikan Tinggi Teknologi Laboratorium Medik Indonesia*, 1-13.

New England Biolabs. (2023). *Pci*I. URL https://international.neb.com/products/r0655-pcii#Product%20Information , diakses pada 24 Juni 2023.

Schuettpelz, E., Chen, C. W., Kessler, M., Pinson, J. B., Johnson, G., Davila, A., ... & Pryer, K. M. (2016). A revised generic classification of vittarioid ferns (Pteridaceae) based on molecular, micromorphological, and geographic data. *Taxon*, *65*(4), 708-722.

Sing, et al. 2011. *Role Of Bioinformatics In Agriculture And Sustainable Development*. Banaras Hindu University, India.

Syamsurizal, S., Handayani, D., Kadri, H., & Badriyya, E. (2019). Genotyping SNP rs7903146 TCF7L2 gene for detection T2DM in Indonesian melayu ethnic. *Journal of Physics: Conference Series*, *1317*(1), 12090. IOP Publishing.

Tamarin, R. H. (2001). *Principles of genetics (seventh edition).* Massachusetts: The McGraw-Hill Companies.

Virgilio, M., Jordaenis, K., Breman, Barr, N., Backeljau, T., & Meyer, D. M. (2012). Turning DNA barcodes into an alternative tool for identification: african fruit flies as model. *Consortium for the Barcode of Life (CBOL).*

Yeriska, F., Umar, M. Z., Hijriah, N. W., Fadhlurrohman, R., & Achyar, A. (2021). Analisis Variasi Genetik Sekuen Gen Pengkode Protein Spike Virus MERS-CoV (PopSet: 1843801421) Menggunakan RFLP Secara In Silico. In *Prosiding Seminar Nasional Biologi* (Vol. 1, No. 2, pp. 898-906).