JSMARTech
Journal of Smart Bioprospecting and Technology **p-ISSN**: 2686-0805 **e-ISSN**: 2714-7894
October 2024, VOL 05 NO 02 **DOI**:https://doi.org/10.21776/ub.jsmartech.2024.005.02.57

Phylogenetic Study of Several Parasitic Plant Species Based on The *atp*-1 Gene Sequence

Ardo Cahya Pratama[1], Nadaa Karuniasari[1], Agnia Fadillah Rahmadini[1], Achmad Ezra[1], Victor Alvianoes Guterez Hose[3], Turhadi Turhadi[1*], Eko Suyanto[1,2], and Fatchiyah Fatchiyah[1,2]

[1]Department of Biology, Faculty of Mathematics and Natural Sciences, Universitas Brawijaya. Jl Veteran, No 1, Malang, East Java, Indonesia, 65145

[2]Research Center of Smart Molecule of Natural Genetic Resources (SMONAGENES), Universitas Brawijaya. . Jl Veteran, No 1, Malang, East Java, Indonesia, 65145

[3]Department of Biomedical Sciences, Faculty of Medicine, Universitas Brawijaya. Jl Veteran, No 1, Malang, East Java, Indonesia, 65145

Submission: 2 August 2024; **Revised**: 26 September 2024; **Accepted**: 1 October 2024

*__Corresponding author__: Turhadi Turhadi; **e-mail**: turhadibiologi@ub.ac.id

ABSTRACT. *The distinction between parasitic and non-parasitic plants can be determined by analyzing the atp-1 gene, which plays a vital role in respiration and is known for its high mutation rate. This study analyzed the kinship of parasitic plant subclass species through the construction of a phylogenetic tree based on atp-1 gene sequences. The atp-1 gene sequences of parasitic and non-parasitic plants with a total of 29 species were obtained from NCBI. The sequences were then aligned with ClustalW and analyzed for mutation patterns. Sequences that have been aligned, phylogenetic trees were made with MEGA11 software with the Maximum Likelihood method and analyzed using the iTOL website. The sequences were analyzed for similarity and kinship with Matrix Coefficient and Haplotype Construction. The atp-1 gene proved that parasitic plants (hemiparasites) are furthermore related to non-parasitic plants compared to holoparasite parasitic plants. Besides that, the kinship of parasitic plants can be analyzed by several methods, namely matrix coefficient to measure similarity, DnaSP to analyzing haplotype, and haplotype network to find out detailed information on mutations that occur. Matrix coefficients can also be used to measure specific similarities between species. It was found that the same subclass had high similarities, for example the species Santalum album and Heisteria parvifolia with a genetic distance value of 0.00574. Meanwhile, different subclasses have low similarity, such as Cassytha filiformis and Ombrophytum with a genetic distance value of 0.07871. This study shows that the atp-1 gene is effective in analyzing the kinship between parasitic and non-parasitic plants. Hemiparasites are genetically closer to non-parasitic plants than holoparasites, with higher genetic similarity within the same subclass.*

Keywords: *parasitic plant, atp-1, phylogenetic tree, matrix coefficient, haplotype.*

INTRODUCTION

Plants or in phylogenetics the plantae group consists of mostly autotrophic organisms. However, within the angiosperm plant group, there are several groups that are categorized as heterotrophic angiosperm groups. This group consists of mycotrophs and haustorial parasites. Angiosperm parasitic plants only make up about 1% of all angiosperm plants and are only limited to the dicot subclasses Magnoliidae, Rosidae, and Asteridae. These three subclasses are then divided into 12 families, 300 genera and 4500 species [1][2][3].

Parasitic plants are types of plants that are unable to make food independently through photosynthesis. Instead, they rely on host plants or other organisms for the essential nutrients needed for their survival, extracting them through a specialized structure known as the haustorium. The haustorium is a specialized structure that penetrates the host's tissue by extracting water, minerals and other nutrients thereby allowing them to live without photosynthesis. Some parasitic plants can grow without true roots or leaves because they do not need these organs to live. Examples of parasitic plants include dodder (*Cuscuta* spp.) which does not have true leaves and roots and rafflesia (*Rafflesia* spp.) which has large flowers and a foul smell [4].

Parasitic plants could share genetics with their hosts through a horizontal gene transfer mechanism. Therefore, the ability of parasitic plants depends on the host plant, so that the characteristics and ability of parasitic plants to adapt differ in each environment. The adaptation of parasitic plants is strongly influenced by genes that regulate various important aspects of their life

cycle and interactions with the host. The *atp*-1 gene is a gene that plays an important role in respiration by encoding a proton pump protein located in the cell membrane of parasitic plants. Parasitic plants have a high gene mutation rate, especially in the *atp*-1 gene, which is due to the important role of this gene in the adaptation of parasitic plants to the host [4][5].

The aim of this study was to analyze the kinship of parasitic plant subclass species through the construction of a phylogenetic tree based on *atp*-1 gene sequences. Through bioinformatics, certain gene sequences in parasitic plants can be compared with non-parasitic plants thus mutations can be detected. Mutations can influence the phenotypic characteristics of plant organisms, making them valuable for studying plant evolution. Understanding these mutations also provides insight into how plants respond to their environment, which can inform effective conservation strategies.

RESEARCH METHODS

Identification of *atp*-1 Gene in Parasitic Plant

The tool used in this research is NCBI (www.ncbi.nlm.nih.gov) for data mining of nucleotide sequences of parasitic plants. The data used is the *atp*-1 gene sequence from 11 different species, consisting of several parasitic plant species representing several families including *S. schreberi* (GU351112.1), *C. kilimanjari* (OM749771.1), *O. fasciculata* (AY741831.1), *C. filliformis* (EU281002.1), *D. pentandra* (EU280999.1), *P. arenarium* (EU281008.1), and non-parasitic plant species namely *N. tabacum* (AY596704.1) as out-group species. Then gene sequence alignment was performed to confirm the conserved region of the predicted *atp*-1 gene sequence [1].

Multiple Sequence Alignment (MSA) and Phylogenetic Tree Construction

The *atp*-1 gene sequence of the parasitic plants *S. schreberi* (GU351112.1), *C. kilimanjari* (OM749771.1), *O. fasciculata* (AY741831.1), *C. filliformis* (EU281002.1), *D. pentandra* (EU280999.1), *P. arenarium* (EU281008.1), were compared to, non-parasitic plant *N. tabacum* (AY596704.1) by alignment with MEGA11 software. The phylogenetic tree was constructed with Maximum Likelihood method, 1000 replications, and grouped based on genetic distance. The phylogenetic tree was converted to the iTOL website.

Analysis of Parasitic Plant Similarity and Relatedness

Each parasitic plant based on the *atp*-1 gene can be analyzed for similarity between individuals. This analysis can be done using the Matrix Coefficient Analysis and Haplotype Construction methods. The coefficient matrix for genetic distance analysis can also be done in MEGA11 software. The way to do this is by selecting on the distance menu option then select compute pairwise distances. The Data results of the genetic distance coefficient matrix can be downloaded in Excel or CSV format.

Another method that can be used is haplotype construction using DnaSP software. The method is to use sequences that have been previously aligned in the MEGA software. Besides DnaSP, the haplotype construction need Network software. The analysis is using Calculate Network option and needed the RDF File previously obtained from DnaSP. After obtaining the results, therefore customize the placement, adjust the circle diameter, and modify the color of each haplotype based on the reference settings to suit the analysis.

RESULTS AND DISCUSSIONS

Data mining on the NCBI website identified 29 species sequences from parasitic plant subclasses, including Asteridae, Magnoliidae, and Rosidae. These sequences were filtered using four specific criteria outlined in Table 1.

Table 1. Criteria for species used.

Criteria	Number The Fulfill
Parasitic plants and availability of sequences on NCBI	29/29
Taxa status is species not sp.	29/29
Sequences are partial cds	29/29
Parasitic plant group	10/29

Data were analyzed according to the kinship of parasitic plant species from several subclasses and identification of parasitic plant genes that possibly used as a markers in phylogenetic analysis. Based on those criteria, 11 species were obtained in accordance with the research objectives (Table 2).

Phylogenetic Construction

The phylogenetic tree of the *atp*-1 gene of parasitic plants from three subclasses, namely Rosidae, Asteridae, and Magnoliidae was constructed using MEGA 11 and visualized using iTOL (Interactive Tree of Life) software (Figure 1).

The phylogenetic tree was visualized using iTOL software. iTOL is a visualization tool used to introduce tree annotations with nodes and interpret evolutionary relationships between organisms. The position and distance of the branches on the phylogenetic tree illustrate the evolutionary relationship, where the closer the branches between organisms mean that the species are closely related while the farther the branches of organisms indicate that the kinship between species is not close. Colors on the phylogenetic tree visualize data related to the organ, such as genomic, phenotypic, or other metadata information. Then the labels on the phylogenetic tree can provide information about the species name or geographical origin of the species [6].

The phylogenetic tree visualization shown in Figure 1 is based on species (Table 2) that have been screened. The visualization shows that there is a kinship between species based on the angiosperm parasitic plant group. *S. album*, *S. shcreberi*, and *H. parvifolia* are in the same clade, which shows that the species are in the Rosidae subclass. Then, *O. fasciculata* and *C. kilimanjari* show the same clade with the Rosidae subclass. The result represents a close kinship of each species because they are in the same subsclass.

The spesies used in this research were parasitic plants as an in-group consisting of *S. shcreberi, O. subterraneum, H. parvifolia, S. album, P. thurberi, C. kilimanjari, O. fasciculata, C. filliformis, D. pentandra, P. arenarium,* and non-parasitic plant species namely *N. tabacum* as out-group species. Based on Figure 1. Shows the presence of color on each branch. The red color indicates that the parasitic plants share the same type of parasite, specifically holoparasites. While the green color indicates the species is classified as hemiparasitic and the black color indicates thath have cholorophyll and are able to photosynthesize but obtain nutrients through the haustorium while holoparasites in red color are types of parasites that do not have chlorophyll so they need a host to obtain water and nutrients [7].

Table 2. Screened species results.

No.	Species	Code	Subclass	Type of parasites
1.	*Schoepfia schreberi*	Ss	Rosidae	Hemiparasite
2.	*Ombrophytum subterraneum*	Os	Rosidae	Hemiparasite
3.	*Heisteria parvifolia*	Hp	Rosidae	Hemiparasite
4.	*Santalum album*	Sa	Rosidae	Hemiparasite
5.	*Pilostyles thurberi*	Pt	Asteridae	Holoparasite
6.	*Cuscuta kilimanjari*	Ck	Asteridae	Holoparasite
7.	*Orobanche fasciculata*	Of	Asteridae	Holoparasite
8.	*Cassytha filiformis*	Cf	Magnoliidae	Holoparasite
9.	*Dendrophthoe pentandra*	Dp	Magnoliidae	Holoparasite
10.	*Pholisma arenarium*	Pa	Magnoliidae	Holoparasite
11.	*Nicotiana tabacum*	Nt	Asteridae	Non-Parasitic

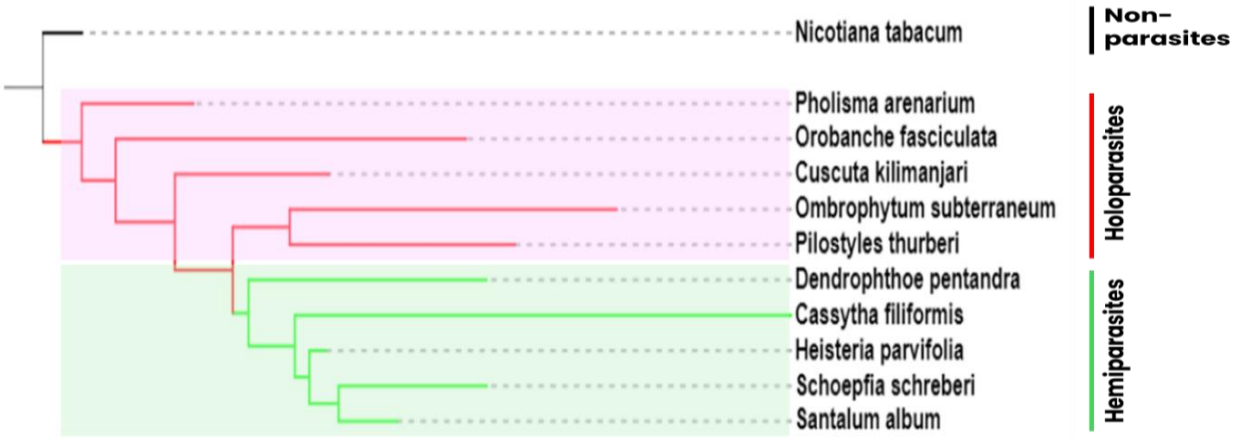

Figure 1. Phylogenetic tree of plants parasitizing the *atp*-1 gene. (Black line: Non-parasites, Red line: Holoparasites, Green line: Hemiparasites.

Matrix Coefficient

Matrix coefficient or correlation matrix is a measurement that shows the correlation of similarities between species and other species. If we look at the results obtained (Table 3), the closest genetic distance was obtained in the species *S. album* and *H. parvifolia* with a value of 0.00574. Meanwhile, the farthest genetic distance was obtained in the species *C. filliformis* and *O. subterraneum* with a value of 0.07871. If confirmed from biosystematics, *S. album* and *H. parvifolia* have high similarity because they are in the same subclass, while *C. filiformis* and *O. subterraneum* have low similarity because they are in different subclasses [8].

Conserve Region and Mutations

Conserve regions and the presence of mutations was shown at the time of alignment using the MEGA11 software (Figure 2). The aligned sequences have a conserve region of 907/1047 bases, while the mutation region is 142/1047 bases. Conserve region refers to the same part of the sequence from several species in the alignment, for example is the yellow area on Figure 2.

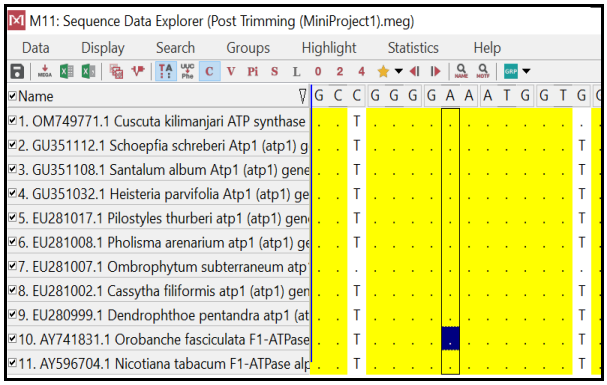

Figure 2. Alignment of conserve (yellow) and mutation region (white) in MEGA11 software.

These sequences have crucial functions that are maintained as evolution progresses. Meanwhile, mutation regions are regions that have different and unequal nucleotide bases between all species in the alignment, for example sites 3 and sites 15 on Figure 2 [12].

Mutations indicate the presence of genetic variation among the organisms or samples analyzed. Table 4 shows the percentage of each mutation that occurred. The symbol "*" indicates the percentage of transitions in the nucleotide base, while the symbol "**" interprets the percentage of transversions in the nucleotide base. Transitions are changes in nucleotides from one type of purine to another (A > G, G > A), as well as in pyrimidines (C > T, T > C). Meanwhile, transversion is the change of purine base to pyrimidine or vice versa (A > C, A > T, G > C, G > T, C > A, C > G, T > A, T > G) [11]. The transition that occurs from base T to A is 0.09%, A to T is 0.19%. Other transitions also occurred in C to G by 0.28% and G to C by 0.19%. Then transversions occurred in G to A by 0.38%, C to A by 0.28%, G to T by 0.19%, C to T by 0.47%. Other transversions are A to G by 0.28%, T to G by 0.28%, A to C by 0.28% as well, then T to C by 0.47%.

Haplotype Analysis Using DnaSP Software

DnaSP (DNA Sequence Polymorphism) software is a tool used with the aim of analyzing haplotypes to see how closely related or different certain haplotype groups are. Based on the results obtained in the form of nucleotide sequence information data and the presence of haplotypes in the sequences, the haplotype diversity (Hd) value was obtained at 1.0000, which indicates that each sequence used has a unique haplotype and there are no duplications between the sequences analyzed. According to [9], the haplotype diversity value has a range of 0.0000-1.0000, the

higher the value obtained, the individual whose sequence is used has a unique haplotype. The level of polymorphism shown in the data obtained is that there are 137 variable sites out of the total number of sites, namely 1049, this indicates that the genetic diversity in this population is not too large.

Genetic Relationships and Haplotype Variation in Parasitic Plants

Haplotype network construction is a widely used approach for analyzing and visualizing relationships among DNA sequences within a population or species. Analyzing DNA sequences from sampled individuals from one or multiple populations allows for answering various questions about micro-evolutionary processes, such as population structure, gene flow, past demographic expansions, and geographic extinctions. A crucial aspect of this analysis is inferring ancestral relationships between haplotypes. In haplotype analysis, some observed sequences may be ancestors of others [10].

Network methods can be classified according to different criteria. For instance, they can be defined into two categories depending on their main objectives: explicit networks, which describe reticulate evolutionary processes such as hybridization, horizontal gene transfer, or admixture, and implicit networks, which represent some form of uncertainty over non-reticulate tree representations. The median joining network (MJN) method is one where alternative branches are found by examining potential ancestral sequences for each triplet of sequences. MJN includes various variants such as reduced median network or quasi-median network [10].

The median joining network describes the variation in parasitic plant species into 11 haplotypes. The haplotype network shown in Figure 2 illustrates the genetic relationships between different haplotypes within a population. Each labeled colored circle, such as H_1, H_2, H_3, and so on, represents individual haplotypes, while small red circles labeled mv2, mv3, mv4, etc., represent mutations or differences between haplotypes. The lines connecting the haplotypes indicate evolutionary paths or relationships between the haplotypes. The more mutations present between haplotypes, the longer the connecting path. The position of haplotypes in this network can provide clues about which haplotypes might be older or ancestral and which are more recent.

The length and complexity of the lines indicate the genetic distance or closeness between haplotypes. Haplotype H_1 appears to have several long branches, indicating many mutations occurring from H_1 to other haplotypes. Several clusters of haplotypes connected through a series of mutations can be seen, such as haplotypes H_8 and H_11, which are connected through two mutation steps (mv1 and mv3), indicating a close genetic relationship. Haplotype H_10 is located at the center of a more complex network, connected to several other haplotypes through various mutation steps, suggesting that H_10 may be an ancestral haplotype or a common haplotype from which many others have emerged. Additionally, some haplotypes appear more isolated, such as H_1, H_5, and H_6, which are connected to other haplotypes only through a series of long mutations, indicating that they may have undergone more separate or independent evolution. Furthermore, the colors within each haplotype circle indicate the regional origin of the respective parasitic plant species, as shown in Figure 3 below.

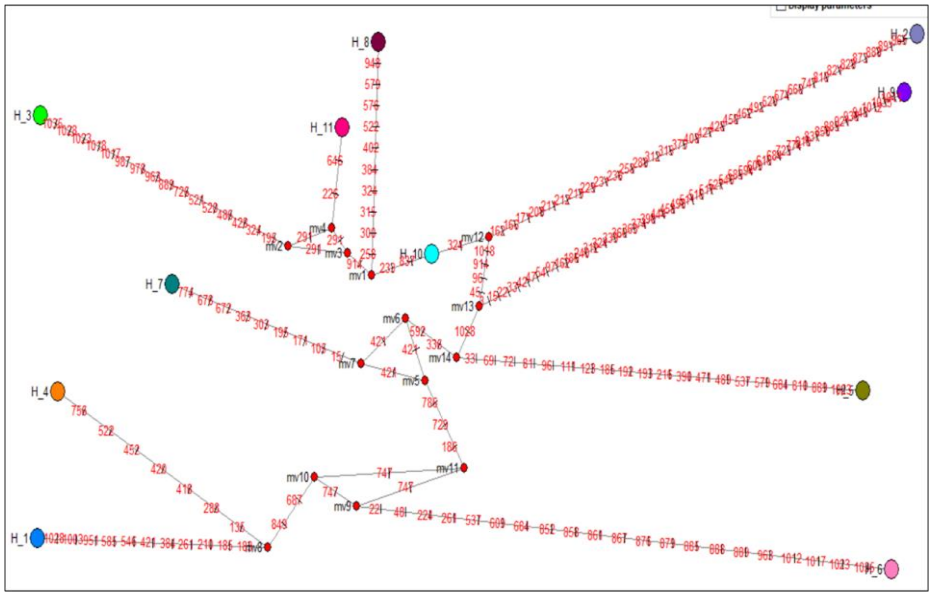

Figure 3. Haplotype network among parasitic plant species with specific mutation location on the DNA sequence.

Table 3. Genetic distance among parasitic plants.

	Os	*Hp*	*Sa*	*Pt*	*Of*	*Ck*	*Cf*	*Dp*	*Pa*	*Nt*	*Ss*
Os											
Hp	0,048										
Sa	0,050	0,006									
Pt	0,057	0,023	0,027								
Of	0,063	0,034	0,038	0,040							
Ck	0,054	0,018	0,022	0,032	0,036						
Cf	0,079	0,033	0,039	0,053	0,065	0,049					
Dp	0,057	0,020	0,019	0,032	0,035	0,027	0,048				
Pa	0,061	0,021	0,025	0,035	0,031	0,023	0,049	0,033			
Nt	0,060	0,022	0,026	0,036	0,034	0,025	0,052	0,036	0,017		
Ss	0,056	0,012	0,013	0,032	0,045	0,029	0,041	0,026	0,031	0,031	

Table 4. Type of mutation in *atp*-1 gene among parasitic plant species (* = transition, ** = transversion).

	A	**T**	**G**	**C**
A	**26,5**	0,09*	0,38**	0,28**
T	0,19*	**25,39**	0,19**	0,47**
G	0,28**	0,28**	**24,8**	0,28*
C	0,28**	0,47**	0,19*	**19,6**

Table 5. Information on the presence of haplotypes in the nucleotide sequence among parasitic plant species.

H_1	Amerika Tengah (Guatemala, el Salvador, Nicaragua, Puerto rico, republic Dominika)
H_2	Amerika Selatan (Bolivia)
H_3	Afrika barat & tengah (senegal, guniea, sierra leone, ghana, nigeria, niger, cameroon, republik afrika tengah, gabon, kongo)
H_4	Asia Selatan (India)
H_5	Amerika Utara (Mexico)
H_6	Afrika Timur (Tanzania)
H_7	Amerika
H_8	Amerika (Honduras, Brazil, Guyana, Mexico) Asia (Indonesia, Jepang) Afrika (Ghana) Oseania (Papua New Guinea) Australia
H_9	Asia Tenggara (Indonesia, Thailand, Filipina)
H_10	Amerika Utara (Mexico)
H_11	Eropa (Jerman, Austria, Czech Republic, Polandia, Serbia, Romania, Hungary)

Note:
H1 (*N. tabacum*); H2 (*C. filiformis*); H3 (*D. pentandra*); H4 (*P. arenarium*); H5 (*P. thurberi*); H6 (*O. fasciculata*); H7 (*C. kilimanjari*); H8 (*S. schreberi*); H9 (*O. subterraneum*); H10 (*H. parvifolia*); H11 (*S. album*).

CONCLUSION

The phylogenetic study of various parasitic plant species based on the *atp*-1 gene sequence aims to analyze the kinship between parasitic and non-parasitic plants. The *atp*-1 gene, which plays a crucial role in respiration, has a high mutation rate, making it a marker for studying the evolution of parasitic plants. In this study, a phylogenetic tree was constructed using the Maximum Likelihood method based on the *atp*-1 gene sequences of several species. The results indicated that parasitic plants, particularly hemiparasites, are more distantly related to non-parasitic plants compared to holoparasites. There are specific regions of *atp*-1 that differentiate between parasitic and non-parasitic plants, for example in non-parasitic plants at site 183 which is the Thymine base, site 210 Guanine base, site 951 Guanine base, and site 1003 Adenine base which are not found in parasitic plants. Further analysis using matrix coefficient and haplotype network reinforced these findings, showing that species from the same subclass had closer genetic distances than those from different subclasses. Understanding these mutations and genetic relationships contributes to the conservation of parasitic plants and enhances our knowledge of their evolution and adaptation to their environment.

ACKNOWLEDGEMENT

The authors would like to thank to Department of Biology, Faculty of Mathematics and Natural Sciences, Universitas Brawijaya for supporting this research.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

REFERENCES

[1] D. L. Nickrent., R. J. Duff, A. E. Colwell, A. D. Wolfe, N. D. Young, K. E. Steiner, and C. W. dePamphilis. Molecular Phylogenetic and Evolutionary Studies of Parasitic Plants. *Molecular Systematics of Plants II*. **1998**.

[2] A. Zervas., G. Petersen, and O. Seberg. Mitochondrial Genome Evolution in Parasitic Plants. *BMC Evolutionary Biology*. **2019**. Vol. 19. Pp. 1-14.

[3] Al-Robai, S. Asir. Ecology and Diversity of Angiosperm Parasites and Their Host Plants along Elevation Gradient in Al-Baha Region, Saudi Arabia. *Diversity*. **2023**. Vol. 15. No. 10.

[4] H. Bouwmeester., N. Sinha, and J. Scholes. Parasitic plants: Physiology, Development, Signaling, and Ecosystem Interactions. *Plant Physiology*. **2021**. Vol.185. No. 4. Pp. 1267-1269.

[5] B. Moncalvillo and D. Matthies. Host Age Affects the Performance of the Root Hemiparasitic Plant *Rhinanthus alectorolophus. Ecology and Evolution.* **2023**. Vol. 13. No. 6.

[6] I. Letunic and P. Bork. Interactive Tree of Life (iTOL) v5: An Online Tool for Phylogenetic Tree Display and AAnnotation. *Nucleic Acids Research*. **2021**. Vol. 49. Pp. 293-296.

[7] M. R. Tambunan., and R. Raihandhany. Jenis-Jenis Tumbuhan Parasit dan Persebarannya di Institut Teknologi Bandung (ITB) Kampus Ganesha. Jurnal Sumberdaya HAYATI. **2020**. Vol. 6. No. 2. Pp. 47-55.

[8] D. M. Chukwuma., and E. A. Ayodele. Taxonomic Studies of the Genus *Tephrosia* Pers. (Papilionaceae) in Nigeria. *Plant Science Today.* **2016**. Vol. 3. No. 1. Pp. 9-18.

[9] O. Fan., J. Fjeldsa, X. Liu, Y. Dong, Y. Chang, Y. Qu, G. Song, and F. Lei. An Approach for Estimating Haplotype Diversity from Sequences with Unequal Lengths. *Methods in Ecology and Evolution*. **2021**. Vol. 12. No. 9. Pp. 1658-1667

[10] E. Paradis. Analysis of haplotype networks: The Randomized Minimum Spanning Tree Method. *Methods in Ecology and Evolution*. **2018**. Vol 9. No. 5. Pp.1308-1317.

[11] R. V. S. A. Morihito., S. E. Chungdinata, T. A. Nazareth, M. I. Pulukadang, R. A. M. Makalew, and B. Pinontoan. Identifikasi Perubahan Struktur DNA terhadap Pembentukan Sel Kanker menggunakan Dekomposisi Graf. *Jurnal Ilmiah Sains*. **2017**. Vol. 17. No. 2.

[12] B. Song., E. S. Buckler, H. Wang, Y. Wu, E. Rees, E. A. Kellogg, D. J. Gates, M. K. Burch, P. J. Bradbury, J. R. Ibarra, M. B. Hufford, and M. C. Romay. Conserved Noncoding Sequences Provide Insights into Regulatory Sequence and Loss of Gene Expression in Maize. *Genome Res.* **2021**. Vol. 31. No. 17. Pp. 1245-1257.