Original Research Article

Molecular characterization and Phylogenetic study of *Andrographis echioides* using DNA barcoding Technique

Abstract

With the advancement in various molecular diagnostic tools, DNA Barcoding has emerged as a gold standard molecular diagnostic tool across the globe. Since ancient times, plants with medicinal properties have been widely used in Indian Ayurvedic medicine for treating a variety of ailments. Plants of the genus *Andrographis* have been extensively used for medicinal purposes for treating different types of ailments. In this study, a rarely studied upon medicinal plant species was isolated, sequenced at the genetic level and studied for its evolutionary characteristics using phylogenetic analysis. In the present study, the identity of *A. echioides* was confirmed by targeting different barcoding genes such as ribulose-bisphosphate carboxylase, Internal Transcribed Spacer, RNA polymerase-beta subunit, maturase K, and photosystem II protein D1 genes using phylogenetic approach. After successful isolation and amplification of genomic DNA, specific primers were utilised for sequencing of each barcoding gene, followed by nucleotide BLAST analysis to determine sequence percent identity of each gene with that from other plant species. The best homologs were then utilised for conducting phylogenetic analysis which confirmed the identity of the plant as *Andrographis echioides*.

Keywords: *Andrographis echioides*, DNA barcoding, phylogeny.

1. Background

In the recent advancements in the field of molecular biology, plant species identification is possible by targeting selected gene sets which are conserved for the given plant species. One of the techniques of plant identification by molecular sequencing is known as DNA barcoding. DNA barcoding is an advanced technique which requires an only small part of the plant by which it maintains the intact diversity since one does not need to sample the whole plant. This technique employs the principle of selection of DNA region targeted for amplification which remains conserved within a species. The success of DNA barcoding is now available to plant species ever since it was first proposed to detect animals, especially since the locus known as cytochrome oxidase unit I, a mitochondrial gene was determined as a plant barcode [1].

The technique of DNA barcoding is now extended to identify fungal species also through internal transcribed spacers (ITS) of nuclear ribosomal DNA [2]. In recent study, molecular identification of plant by the number of DNA markers are proposed which are utilized either individually or in combinations [3] [4]. Few previous investigations reported targeting by primers for the two-locus based DNA barcoding and became prominently accepted especially for plastid genes such as ribulose-bisphosphate carboxylase (rbcL) and maturase K (matK) [5]. In many studies for common plant identification at the species level, use of internal transcribed spacer (ITS) region of nuclear ribosomal cistron is advised as a DNA marker [6]. Additionally, plasmid DNA containing trnHsbA spacer region is also one of the choices for plant DNA barcoding. However, targeting of this region is tough since it represents extensive length variations, also the presence of intraspecific micro inversions associated with palindromes and sequencing problems related to mononucleotide repeats.

Even though these markers are termed as standard DNA barcodes but, in some cases, cannot be used directly. For example, animal feeding on plants when screened for their gut plant species sampled it is difficult to target by these barcode markers since sample plant DNA register extensive damage. It is also noted that matK marker did not cover a broad spectrum of plant taxonomic units and hence responded to limited plant species identification. As an alternative now, the involvement of variable region targeting primers are used together and able to detect plant in a concluding analysis. In recent time, use of a plastid intron located in the tRNA Leu UAA gene has been used prominently to study diet analysis. The method also has some drawbacks like that of trnH-psbA such as length variation, and hence the extent of utility for the marker in plant identification is questionable. Still, it is prominently prescribed for plant barcoding in insects mainly for beetles. The trnL intron has also been reported successful for identification of below-ground plant richness (from roots)[7] [8].

Since ancient times, plants with medicinal properties have been extensively used in Indian Ayurvedic medicines for treating a variety of ailments. Out of the numerous plants well known for their medicinal use, approximately 20 out of 40 plants of the genus *Andrographis* occurring in India, have been used on a large scale in traditional Indian medicine for the treatment of dyspepsia, influenza, malaria, respiratory tract infections and as astringent and antidote for poisonous insect stings [9]. Understanding the importance of DNA Barcoding of individual plant species, the current study aims to conduct DNA Barcode profiling of *Andrographis echioides*, which has been previously known to exhibit anti-oxidant and anti-microbial properties against few pathogenic organisms [9] [10].

2. Methods

2.1. Collection of plant specimen

The plant sample was collected from The Maharaja Sayajirao University of Baroda, Gujarat, India. The collected plant specimen was then immediately flash frozen in liquid nitrogen and then stored at -80 ℃.

2.2. Genomic DNA Isolation and Visual Quantification

For extraction of genomic DNA, preserved plant samples were washed twice using distilled water and small fragments of nearly 0.5mm in length were cut. These fragments were then treated with liquid nitrogen and crushed using mortar and pestle to obtain a fine powder of the plant fragments. The genomic DNA was isolated using the modified Cetyltrimethylammonium bromide (CTAB) protocol. The extracted genomic DNA was then resuspended in 30 µL of TE buffer until further use. The extracted genomic DNA of the plant was then run on 0.8% agarose gel for visual confirmation of successful isolation of genomic DNA. The isolated genomic DNA was quantified using a NanoDrop spectrophotometer, to adjust the DNA concentration of the sample before initiating the Polymerase Chain Reaction (PCR) process in PCR Thermocycler.

2.3. PCR amplification of Barcoding genes from Genomic DNA

The PCR technique can multiply defined region of template genomic DNA by involving marker primers was conducted successfully by using the PCR Thermocycler. In the present study, total five genes were targeted by involving genomic DNA of plant *A. echioides*. The target genes were rbcL, ITS, psbA, rpoB and matK. The details of primers utilized in DNA amplification represented in Table 1. The amplified products obtained from the PCR process were run on 1.5 % agarose gel for visual confirmation of amplified gene sequences.

Table 1: Details of Primers utilized for DNA amplification

Primer Name		Primer sequence 5’ to 3’	Amplicon length in base pairs
rbcL	Forward primer	ATGTCACCACAAACAGAGACTAAAGC	613
	Reverse primer	GTAAAATCAAGTCCACCRCG	
ITS	ITS-1 Forward primer	TCCGTAGGTGAACCTGCGG	700
	ITS-4 Reverse primer	TCCTCCGCTTATTGATATGC	

matk	matk 320 Forward primer	CGATCTATTCATTCAATATTTC	980
	matk 1326 Reverse primer	TCTAGCACACGAAAGTCGAAGT	
psbA-trnHF	psbA- Forward primer	GTTATGCATGAACGTAATGCTC	681
	trnHF-Reverse primer	CGCGCATGGTGGATTCACAATCC	
rpoB - trnCGAR	rpoB - Forward primer	CKACAAAAYCCYTCRAATTG	1083
	trnCGAR Reverse	CACCCRGATTYGAACTGGGG	

2.4. Phylogenetic analysis of *Andrographis echioides*

Initially FASTA sequences of closely related *Andrographis* plants were retrieved and aligned along with the resultant amplified barcode genes gene in MEGA 11 using Multiple Sequence Comparison by Log-Expectation (MUSCLE), a multiple sequence alignment method. Phylogenetic analyses of the *A. echioides* barcoding genes sequences were carried out using the using the distance-based method such Maximum Likelihood (ML) method along with the Bootstrap Phylogeny test with number of bootstrap replications set at 1000 along with the Taimura-Nei model. *Rhinacanthus nasutus* (EU725798.1) and *Thunbergia erecta* (MZ555773.1) from relatively evolutionarily close taxa was set as an outgroup [11].

3. Result

3.1. Genomic DNA isolation and Purification

The plant *A. echioides* was investigated for its genomic DNA by the given protocol and later was analyzed with the help of gel electrophoresis on 0.8% agarose gel. The result obtained showed the presence of genomic DNA above 10 kb. The visualization under gel documentation showcased high quality genomic DNA separation run along with molecular marker as given in Figure 1.

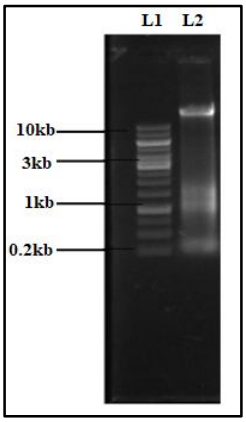

Figure 1: Results for genomic DNA extraction. The genomic DNA extracted from *Andrographis echioides* was analysed on 1.0% Agarose gel stained with Ethidium bromide. L1: DNA ladder of 1kb. L2: gDNA of *Andrographis echioides*

3.2. Amplification of Barcoding genes

The targeting of rbcL, ITS, rpoB, matK, and psbA gene with the genome of *A. echioides* resulted in successful amplification of 200, 400, 900, 700 and 600 base pairs as recorded with 1.5% agarose gel documentation shown in Figure 2.

Figure 2: Result for PCR amplified genes from genomic DNA. (A) L1: DNA ladder of 1.5 kb, L2: amplified rbcL gene, L3: amplified ITS gene from *Andrographis echioides* (B) L1: DNA ladder of 1.5 kb, L2: amplified rpoB gene, L3: amplified matK gene, L4: amplified psbA gene from *Andrographis echioides*

3.3. Sequencing of Marker genes

Once the amplicon obtained from the PCR for gene-targeted from *A. echioides plant*, it was successfully sequenced by Sanger Sequencing. The result showed, psbA gene partially sequenced for 351 nucleotides. Similarly, the gene such as rbcL, matK, ITS and rpoB were successfully sequenced for partial 360, 729, 510, and 455 base pairs respectively as shown in Figure 3.

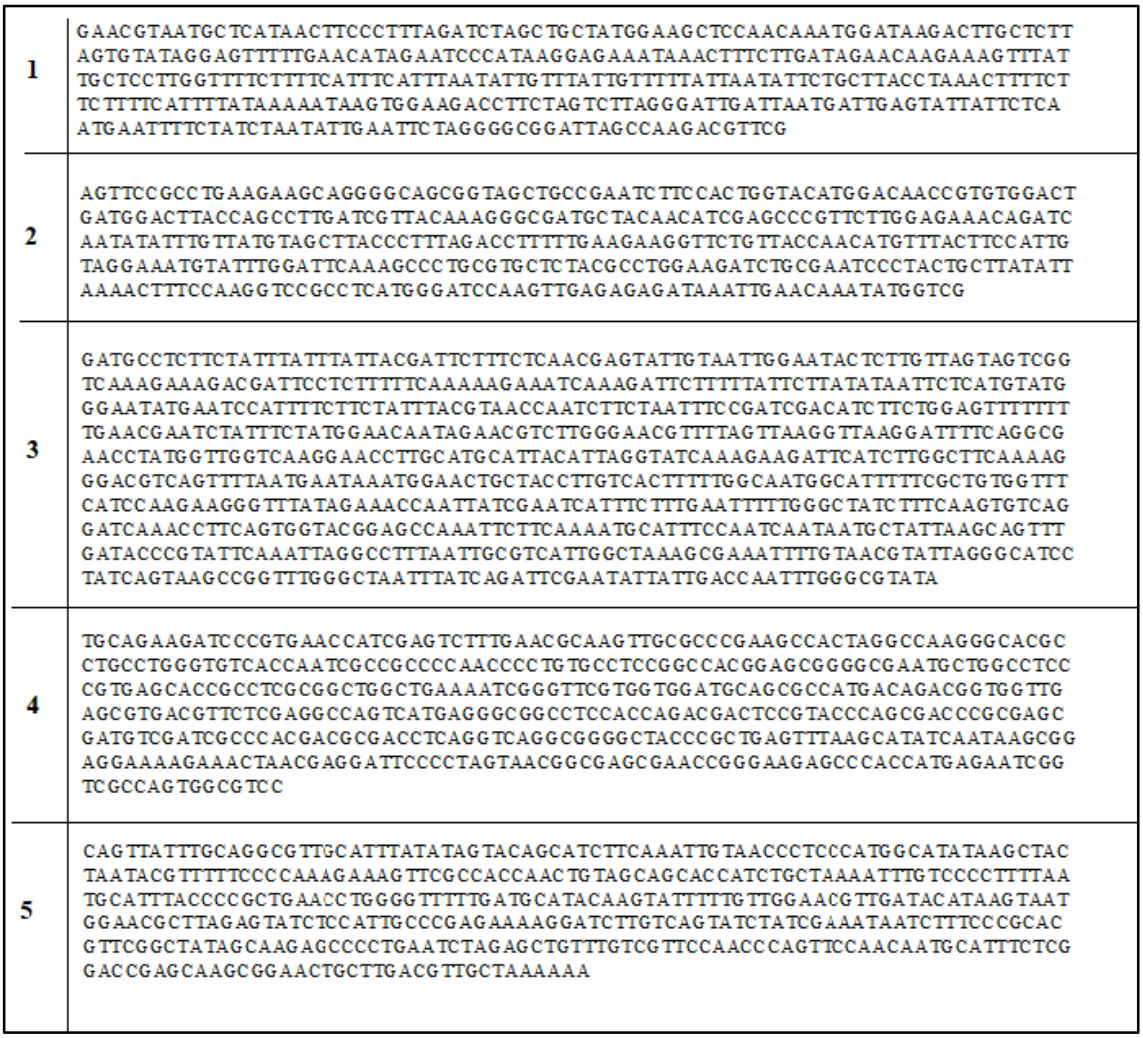

Figure 3: (1) Partial gene sequence of psbA from *Andrographis echioides* (2) Partial gene sequence of rbcL from *Andrographis echioides* (3) Partial gene sequence of matK from *Andrographis echioides* (4) Partial gene sequence of ITS from *Andrographis echioides* (5) Partial gene sequence of rpoB from *Andrographis echioides*

3.4. Phylogenetic Analysis

The evolutionary history was inferred using the Maximum Likelihood method [12]. The evolutionary distances were computed using the Taimura-Nei method [13][14] and are in the units of the number of base substitutions per site. All ambiguous positions were removed for each sequence pair (pairwise deletion option). The proportion of sites where at least 1 unambiguous base is present in at least 1 sequence for each descendent clade is shown next to each internal node in the tree. This analysis involved 6 nucleotide sequences. Codon positions included were 1st+2nd+3rd+Noncoding. Evolutionary analyses were conducted in MEGA 11.

3.4.1. Phylogenetic Analysis of psbA gene

The analysis indicated that the psbA amplicon had 72% percent similarity with psbA gene of *Andrographis glandulosa* (Figure 4). There was a total of 410 positions in the final dataset.

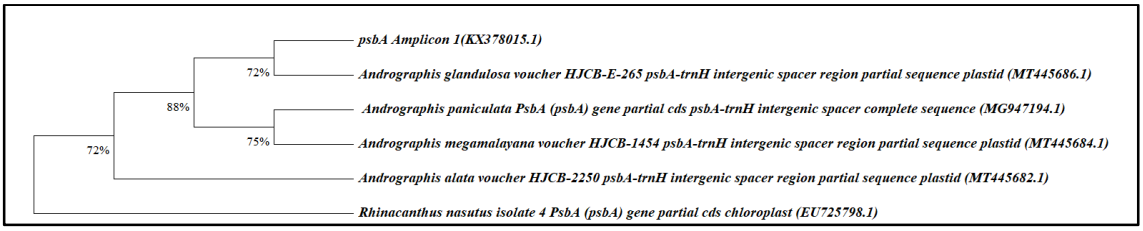

Figure 4: Phylogenetic tree of *Andrographis echioides* psbA gene sequence obtained by Maximum Likelihood (ML) method using MEGA 11 software. The accession numbers of the organisms are included in parentheses

3.4.2. Phylogenetic Analysis of rbcL gene

The analysis indicated that the rbcL amplicon had 51% percent similarity with rbcL gene of *Andrographis serpyllifolia* (Figure 5). There was a total of 703 positions in the final dataset.

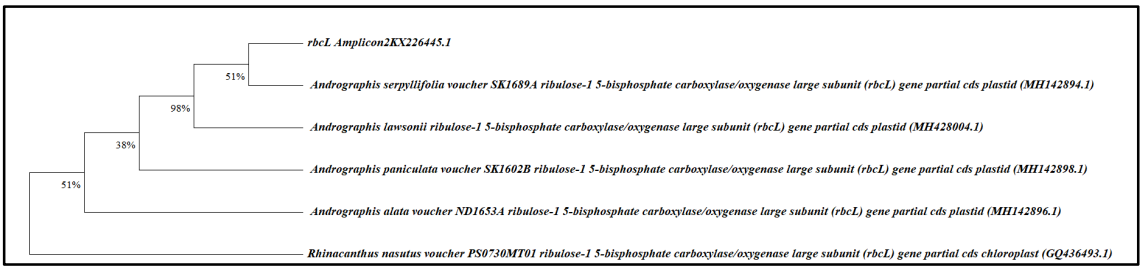

Figure 5: Phylogenetic tree of *Andrographis echioides* rbcL gene sequence obtained by Maximum Likelihood (ML) method using MEGA 11 software. The accession numbers of the organisms are included in parentheses

3.4.3. Phylogenetic Analysis of matK gene

The analysis indicated that the matK amplicon had 79% percent similarity with matK gene of *Andrographis nallamalayana* (Figure 6). There was a total of 915 positions in the final dataset.

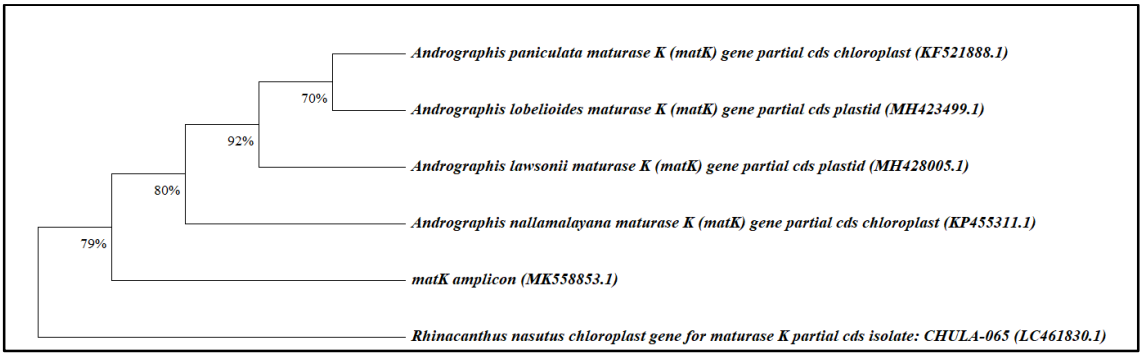

Figure 6: Phylogenetic tree of *Andrographis echioides* matK gene sequence obtained by Maximum Likelihood (ML) method using MEGA 11 software. The accession numbers of the organisms are included in parentheses

3.4.4. Phylogenetic Analysis of ITS gene

The analysis indicated that the ITS amplicon had 69% percent similarity with ITS gene of *Andrographis paniculate* (Figure 7). There were a total of 738 positions in the final dataset.

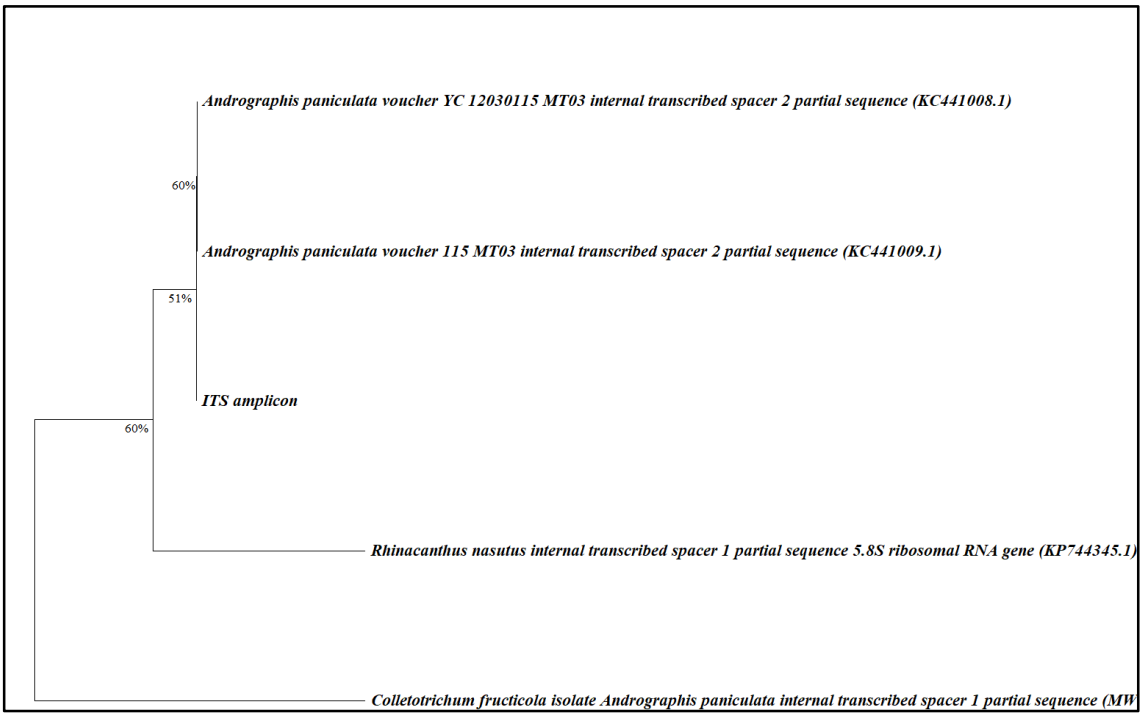

Figure 7: Phylogenetic tree of *Andrographis echioides* ITS gene sequence obtained by Maximum Likelihood (ML) method using MEGA 11 software. The accession numbers of the organisms are included in parentheses

3.4.5. Phylogenetic Analysis of rpoB gene

The analysis indicated that the rpoB amplicon had 94% percent similarity with rpoB gene of *Andrographis paniculate* (Figure 8). There were a total of 672 positions in the final dataset.

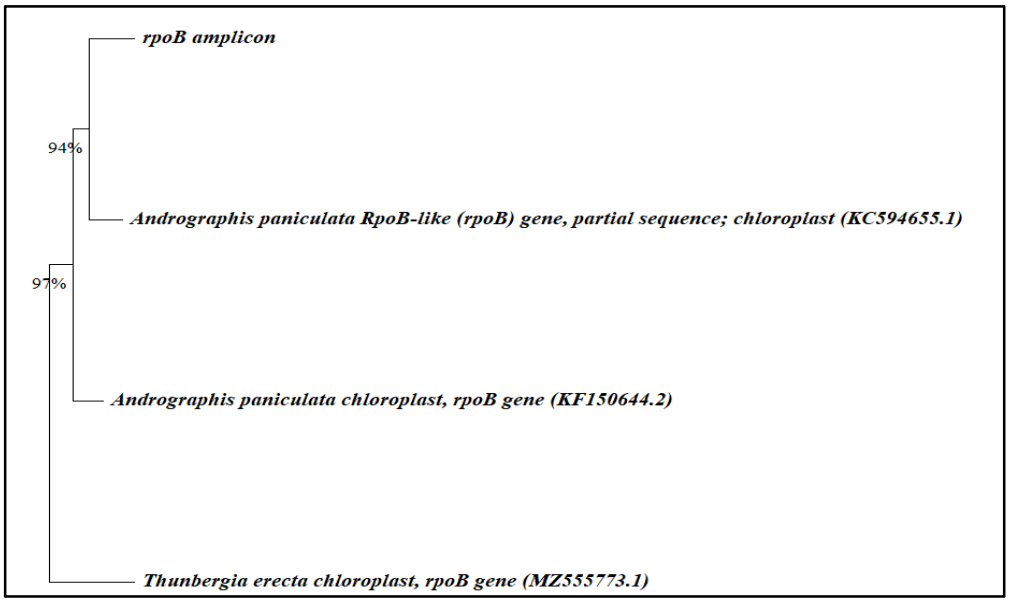

Figure 8: Phylogenetic tree of *Andrographis echioides* rpoB gene sequence obtained by Maximum Likelihood (ML) method using MEGA 11 software. The accession numbers of the organisms are included in parentheses

4. Discussion

In the present study, the identity of *A. echioides* was confirmed by targeting the ITS region. In a similar approach, a potential DNA barcode nuclear ribosomal ITS region was used previously to extensively classify plants on the phylogenetic basis [15] [16] [3]. According to few authors ITS region being localized in the nuclear region represented some drawbacks [17][18]. Taking concern of it, the use of highly conserved chloroplast-based DNA barcodes are preferred along with nuclear ribosomal ITS [15][3]. As we also know that non-coding region is more vulnerable to the genetic variation once compared with coding regions since they are responsible for protein-based functional constraints [19][20]. Taking into consideration above points, the present study investigated the plastid-encoded psbA, matK and rpoB genes, which give more precise identification of *A. echioides*. Similar attempt earlier reported by [3] indicated that the use of plastid trnH-psbA spacer is a suitable locus for species identification.

5. Conclusion

The use of DNA barcoding for plant species identification by involving several nuclear and plastid encoded genes able to identify *A. echioides* plant under investigation up to species level. The study showcased the more than 50% homology with every sequenced data obtained from psbA, rbcL, matk, ITS and rpoB gene which confirmed the experimental plant as *A. echioides*.

Abbreviations:

ITS, Internal Transcribed Spacers; rbcL, ribulose-bisphosphate carboxylase; matK, maturase K; psbA, photosystem II protein D1; rpoB, RNA polymerase-beta subunit; CTAB, Cetyltrimethylammonium bromide; PCR, Polymerase Chain Reaction; MUSCLE, Multiple Sequence Comparison by Log-Expectation; ML, Maximum Likelihood

Data Availability Statement: All data generated or analysed during this study are included in this published article (and its supplementary information files).

Ethical statement: No animals were harmed during this study.

6. References

[1] P. D. N. Hebert, S. Ratnasingham, and J. R. DeWaard, "Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species.," *Proc. R. Soc. B Biol. Sci.*, vol. 270, no. Suppl 1, p. S96, 2003, doi: 10.1098/RSBL.2003.0025.

[2] K. A. Seifert, "Progress towards DNA barcoding of fungi," *Mol. Ecol. Resour.*, vol. 9 Suppl s1, no. SUPPL. 1, pp. 83–89, May 2009, doi: 10.1111/J.1755-0998.2009.02635.X.

[3] W. J. Kress and D. L. Erickson, "A Two-Locus Global DNA Barcode for Land Plants: The Coding rbcL Gene Complements the Non-Coding trnH-psbA Spacer Region," *PLoS One*, vol. 2, no. 6, p. e508, 2007, doi: 10.1371/JOURNAL.PONE.0000508.

[4] "A Proposal for a Standardised Protocol to Barcode All Land Plants on JSTOR." https://www.jstor.org/stable/25065788 (accessed May 11, 2022).

[5] A. J. Fazekas *et al.*, "Multiple Multilocus DNA Barcodes from the Plastid Genome Discriminate Plant Species Equally Well," *PLoS One*, vol. 3, no. 7, p. e2802, Jul. 2008, doi: 10.1371/JOURNAL.PONE.0002802.

[6] I. Álvarez and J. F. Wendel, “Ribosomal ITS sequences and plant phylogenetic inference,” *Mol. Phylogenet. Evol.*, vol. 29, no. 3, pp. 417–434, 2003, doi: 10.1016/S1055-7903(03)00208-2.

[7] D. P. Little, “A DNA mini-barcode for land plants,” *Mol. Ecol. Resour.*, vol. 14, no. 3, pp. 437–446, May 2014, doi: 10.1111/1755-0998.12194.

[8] M. L. Hollingsworth *et al.*, “Selecting barcoding loci for plants: evaluation of seven candidate loci with species-level sampling in three divergent groups of land plants,” *Mol. Ecol. Resour.*, vol. 9, no. 2, pp. 439–457, Mar. 2009, doi: 10.1111/J.1755-0998.2008.02439.X.

[9] D. Y. Shen *et al.*, “Chemical Constituents from Andrographis echioides and Their Anti-Inflammatory Activity,” *Int. J. Mol. Sci.*, vol. 14, no. 1, p. 496, Jan. 2013, doi: 10.3390/IJMS14010496.

[10] K. Jeevanantham, A. Zahir Hussain, and A. Z. Hussain, “ANTIMICROBIAL AND ANTIOXIDANT ACTIVITY OF Andrographis echioides (L.) Nees LEAVES EXTRACTS,” 2021, doi: 10.31032/IJBPAS/2021/10.11.1001.

[11] R. Solgi *et al.*, “Morphological and molecular description of parasitic leeches (Annelida: Hirudinea) isolated from rice field of Bandar Anzali, North of Iran,” *Gene Reports*, vol. 23, no. January, p. 101162, 2021, doi: 10.1016/j.genrep.2021.101162.

[12] N. Saitou and M. Nei, “ESCALA CIWA-AR Escala CIWA-Ar(Clinical Institute Withdrawal Assesment for Alcohol) Evaluación del Síndrome de Abstinencia Alcohólica,” *Mol. Biol. Evol*, vol. 4, no. 4, pp. 406–425, 1987.

[13] felse, “Confidence Limits on Phylogenies : An Approach Using the Bootstrap Author (s): Joseph Felsenstein Published by : Society for the Study of Evolution Stable URL : http://www.jstor.org/stable/2408678 Accessed : 26-05-2016 15 : 14 UTC Your use of the JSTOR,” *Evolution (N. Y).*, vol. 39, no. 4, pp. 783–791, 1985.

[14] M. Kimura, “A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences,” *J. Mol. Evol.*, vol. 16, no. 2, pp. 111–120, 1980, doi: 10.1007/BF01731581.

[15] M. W. Chase *et al.*, “Land plants and DNA barcodes: short-term and long-term goals,” *Philos. Trans. R. Soc. Lond. B. Biol. Sci.*, vol. 360, no. 1462, pp. 1889–1895, Oct. 2005, doi: 10.1098/RSTB.2005.1720.

[16] R. D. H. Barrett and P. D. N. Hebert, “Identifying spiders through DNA barcodes,” *Can. J.*

Zool., vol. 83, no. 3, pp. 481–491, Mar. 2005, doi: 10.1139/Z05-024.

[17] "Potential phylogenetic utility of the nuclear FLORICAULA/LEAFY second intron: comparison with three chloroplast DNA regions in Amorphophallus (Araceae) | Read by QxMD." https://read.qxmd.com/read/15022754/potential-phylogenetic-utility-of-the-nuclear-floricaula-leafy-second-intron-comparison-with-three-chloroplast-dna-regions-in-amorphophallus-araceae (accessed May 11, 2022).

[18] R. L. Small, R. C. Cronn, and J. F. Wendel, "Use of nuclear genes for phylogeny reconstruction in plants," *Aust. Syst. Bot.*, vol. 17, no. 2, pp. 145–170, Apr. 2004, doi: 10.1071/SB03015.

[19] L. Gielly and P. Taberlet, "The use of chloroplast DNA to resolve plant phylogenies: noncoding versus rbcL sequences.," *Mol. Biol. Evol.*, vol. 11, no. 5, pp. 769–777, Sep. 1994, doi: 10.1093/oxfordjournals.molbev.a040157.

[20] J. Shaw *et al.*, "The tortoise and the hare II: relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis," *Am. J. Bot.*, vol. 92, no. 1, pp. 142–166, 2005, doi: 10.3732/AJB.92.1.142.