Saudi Journal of Biological Sciences 29 (2022) 1125–1133

Contents lists available at ScienceDirect

Saudi Journal of Biological Sciences

journal homepage: www.sciencedirect.com

Original article

Molecular identification and phylogenetic analysis of *Aloe shadensis* from Saudi Arabia based on matK, rbcL and ITS DNA barcode sequence

Awad A. Algarni

Department of Biology, College of Science and Arts, Albaha University, Baljurashi, Saudi Arabia

ARTICLE INFO

Article history:
Received 12 July 2021
Revised 8 August 2021
Accepted 13 September 2021
Available online 20 September 2021

Keywords:
Aloe shadensis
rbcL
matK
ITS
Molecular Identification
DNA barcoding

ABSTRACT

The Kingdom of Saudi Arabia thrives with great plant diversity, including rare plants of the family *Asphodelaceae* that have multiple benefits and are still being studied. *Aloe shadensis* is one of these plants that must be preserved and documented in its natural environment. The most appropriate molecular approach currently approved for documentation is the sequencing of some genomic markers. The current study is the first to use genomic markers to record this rare plant. In this study, the plastid genes *matK* (Maturase K), *rbcL* (Ribulose-bisphosphate carboxylase/oxygenase large subunit), and the nuclear region ITS (Internal transcribed spacer) were used to reveal their efficiency in identifying the plant under study. This study is the first to deal with this plant and document it using these genetic markers. The study showed a promising result concerning identifying the sequence of the *matK* gene and ITS region, while the *rbcL* gene did not give a good indicator through the used primers. The obtained sequences of the *matK* gene and the ITS region were determined through two different sets of primers in each case then deposited in GenBank. The evolutionary relatedness of *Aloe shadensis* was established with the different species of *Aloe*. The study showed that the closest species is *Aloe vera* with a similarity of more than 99 %. The study concludes with the possibility of using these genes to correctly identify, distinguish and document the species of *Aloe shadensis*.

1. Introduction

Kingdom of Saudi Arabia possesses enormous plant diversity. Their benefits are not yet fully discovered and hence the need to conserve them. These plants include the family Asphodelaceae and the genus *Aloe*, in particular, some vulnerable species. Therefore, there is a need to document such plants genetically and their phenotypic characteristics to attain the ideal identification of the species (Das et al., 2015). Such studies are of great importance for understanding the characteristics of these species as well as the whole family (Jaiswal et al., 2021). *Aloe shadensis* is distinguished by its presence in Jabal Shada in Albaha region, Saudi Arabia only and distinctive phenotypic characteristics. Many of its characteristics and naming are mentioned in several sources (Chaudhary, 2001; Carter et al., 2011). This study is the first to document this plant genetically. Barcoding is based on determining preserved DNA regions in different species (Wattoo et al., 2016). Many genes were reported useful since the development of DNA barcoding of plants in 2008 (Yu et al., 2021). The most common markers in previous studies include the plastid *matK* and *rbcL*, the nuclear ITS1, and ITS2. Moreover, the combination of these markers is important to confer the species identification correctly (Techen et al., 2014; Ganie et al., 2015; Kress, 2017). The study aimed to evaluate *rbcL*, *matK*, and ITS genetic markers in DNA barcoding of *Aloe shadensis* and provide recommendations about the efficacy in identifying this species properly.

2. Materials and methods

2.1. Plant sampling

The plant samples have collected in 2020 from Jabal Shada area with the coordinates of the sites, Latitude, 19.8541 and Longitude, 41.3110, in Al-Baha, KSA. The plant was recognized by its morphological features and has been confirmed. The plant leaf was used to conduct the study (Fig. 1).

Peer review under responsibility of King Saud University.

Production and hosting by Elsevier

E-mail address: Dr.awad101@gmail.com

https://doi.org/10.1016/j.sjbs.2021.09.053
1319-562X/

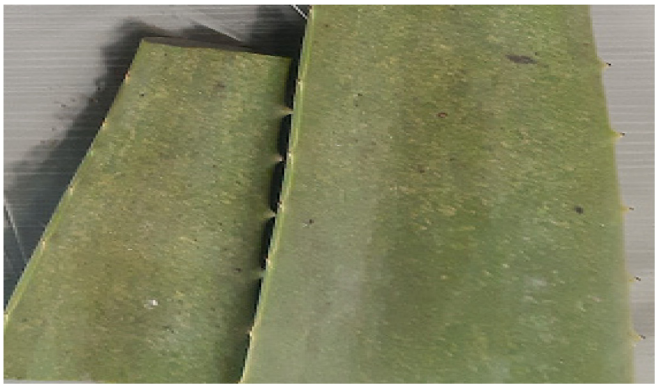

Fig. 1. The used leaves of *Aloe shadensis* for DNA extraction.

Table 3
The maximum likelihood estimate of substitution matrix in the constructed phylogenetic tree of *Aloe shadensis matK* gene sequence by the first pair of primers.

From\To	A	T	C	G
A	–	7.5782	3.2754	9.3227
T	6.0101	–	9.5153	2.8379
C	6.0101	22.0155	–	2.8379
G	19.7435	7.5782	3.2754	–

2.2. Molecular identification

2.2.1. DNA extraction

DNA has been extracted from the plant leaf sample (100 mg) following the kits' manual (NucleoSpin Plant II, Macherey-Nagel). The purity of the extracted DNA was ensured by gel electrophoresis and documentation (Bio-Rad).

Table 1
The primers used for PCR and sequencing in this study.

Loci	Primers	Direction	Sequence	Reference
rbcL	*rbcL* 1F	F	5′- ATGTCACCACAAACAGAAAC-3′	Lledo et al. (1998)
	rbcL 724r	R	5′- TCGCATGTACCTGCAGTAGC-3′	
matK	*matK*-XF	**1st**	F	5′-TAATTTACGATCAATTCATTC-3′	Dunning and Savolainen (2010)
	matK-MALP-R1		R	5′-ACAAGAAAGTCGAAGTAT-3′	
	1R_ kim	**2nd**	F	5′- ACCCAGTCCATCTGGAAATCTTGGTTC -3′	
	3F_ kim		R	5′- CGTACAGTACTTTTGTGTTTACGAG -3′	
ITS 1	ITS 5a	F	5′-CCTTATCATTTAGAGGAAGGAG-3′	Stanford et al. (2000)
	ITS 4	R	5′-TCCTCCGCTTATTGATATGC -3′	
ITS 2	ITS S2F	F	5′-ATGCGATACTTGGTGTGAAT-3′	Chiou et al. (2007)
	ITS S3R	R	5′-GACGCTTCTCCAGACTACAAT-3′	

Table 2
The plant under study, *Aloe shadensis*, and the percentage of similarity with the closest species in the GenBank based on the *matK* gene sequence by the first pair of primers.

No.	Species description	Scientific name	Aligned sequence (bp)	Coverage (%)	E value	Similarity (%)	Accession
1	*Aloe shadensis*	*Aloe shadensis*	803	100	0	100	MZ458425
2	*Aloe vera* voucher *Aloe vera*	*Aloe vera*	800	99	0	99.88	KX377524.1
3	*Aloe vera* voucher CMPR8774	*Aloe vera*	800	99	0	99.88	KY556640.1
4	*Aloe vera*	*Aloe vera*	800	99	0	99.88	JQ276402.1
5	*Aloe vera* isolate 2	*Aloe vera*	799	99	0	99.75	AY323726.1
6	*Aloe vera* voucher Av3	*Aloe vera*	795	99	0	99.87	KP072727.1
7	*Aloe vera*	*Aloe vera*	795	99	0	99.87	MW176075.1
8	*Aloe vera* voucher Av1	*Aloe vera*	794	99	0	99.75	KP072725.1
9	*Aloe vera*	*Aloe vera*	798	99	0	99.50	AJ511390.1
10	*Aloe compressa* var. *compressa*	*Aloe compressa* var. *compressa*	799	99	0	99.50	AY323721.1
11	*Aloe gneissicola*	*Aloe gneissicola*	799	99	0	99.50	AY323720.1

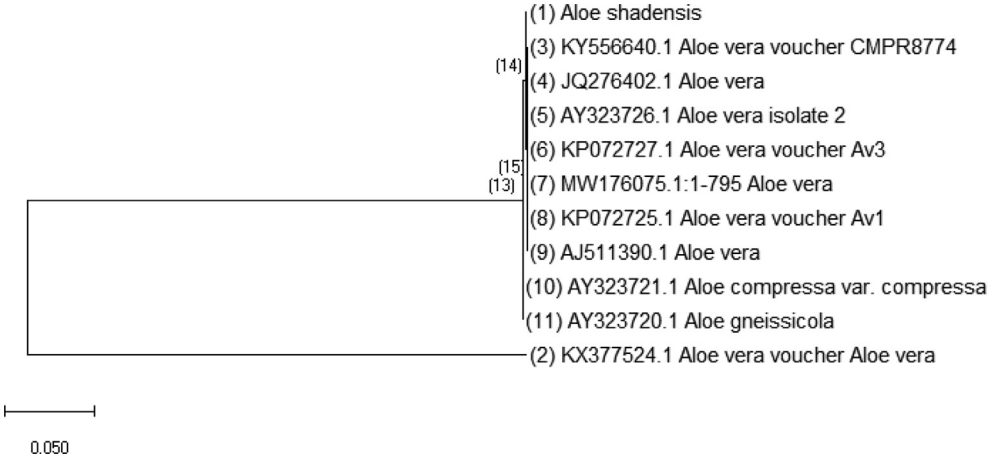

Fig. 2. Phylogenetic tree by neighbor-joining of Aloe shadensis with the closest species based on the matK gene sequence by the first pair of primers.

Aloe vera voucher Aloe vera chloroplast, complete genome
Sequence ID: **KX377524.1** Length: 152875 Number of Matches: 1
Range 1: 1959 to 2758

Score	Expect	Identities	Gaps	Strand	Frame
1533 bits(797)	0.0()	799/800(99%)	0/800(0%)	Plus/Minus	

```
Query  1     ATGGAAATCTTGGTTCAAATCCTTCAATGCCGGATTCAAGATGTTCCTTTTTTGCATTTA  60
Sbjct  2758  ............................................................  2699

Query  61    TTGCGATTCTTTCTTCATGAATATCATAATTGTAATAGTCTTCTCATTACTCAGAACAAA  120
Sbjct  2698  ............................................................  2639

Query  121   TCTATTTATGTTTTTTCAAATGAAAATAAAAGACTATTTCAGTTACTATACAATTCTTAT  180
Sbjct  2638  ............................................................  2579

Query  181   GCTTTTGAATGTGAATTTTTATTAGttttttttCGTAAACAATCTTATTATTTACGATTA  240
Sbjct  2578  ............................................................  2519

Query  241   ACATCTTCTGCAACTTTTCTTGAACGAACCCATTTCTATAGAAAAATAGAACATCTTCGA  300
Sbjct  2518  ............................................................  2459

Query  301   ATAGAACATTTTTTCGTAGTATGTCGTAACTATTTTCATAGAACTCGATGGTTCTTCAAA  360
Sbjct  2458  ............................................................  2399

Query  361   AATCCTTTCATGCATTATGTTCGATATCAAAGAAAGGCAATTGTTGCTTCAAGGGGGACT  420
Sbjct  2398  ............................................................  2339

Query  421   CATTTTCTGATGAAGAAATGGAAATCCCATTTTGTCAATTTCTGGCAATATTATTTTCGC  480
Sbjct  2338  ............................................................  2279

Query  481   TTTTGGTCTCGACCGTACAGAATTCATATAAATCATTTATCAAACTATTCCTTCTATTTT  540
Sbjct  2278  ............................................................  2219

Query  541   CTAGGTTATTTTTCAAGTCTACTAATAAATTCTTCGGCGGTAAGGAATCAAATGTTAGAG  600
Sbjct  2218  ............................................................  2159

Query  601   AATTCATTTCTAATGGATACCGTTACTAAAAAATTTGATACCATAGTCCCAGTTATTCTT  660
Sbjct  2158  ..............................G.............................  2099

Query  661   CTTATTGAATCCTTGTCTAAAGCTAAATTTTGTACCGTATCAGGCCATCCTATTAGTAAG  720
Sbjct  2098  ............................................................  2039

Query  721   CCGATCTGGGCCGATTTCTCAGATTCTGATATTATTGATCGATTTGGTCGGATATGTAGA  780
Sbjct  2038  ............................................................  1979

Query  781   AATCTTTCTCATTATCACAG  800
Sbjct  1978  ....................  1959
```

Fig. 3. The Sequence alignment of *Aloe shadensis* (Query) and the first next species *Aloe vera* (Subject) based on *matK* gene sequence by the first pair of primers.

Table 4
The plant under study, *Aloe shadensis*, and the percentage of similarity with the closest species in the GenBank based on the *matK* gene sequence by the second pair of primers.

No.	Species description	Scientific name	Aligned sequence (bp)	Coverage (%)	E value	Similarity (%)	Accession
1	*Aloe shadensis*	*Aloe shadensis*	825	100	0	100	MZ438002
2	*Aloe vera* voucher *Aloe vera*	*Aloe vera*	800	100	0	99.76	KX377524.1
3	*Aloe vera*	*Aloe vera*	837	100	0	99.76	JQ276402.1
4	*Aloe vera* isolate 2	*Aloe vera*	837	100	0	99.64	AY323726.1
5	*Aloe compressa* var. *compressa*	*Aloe compressa* var. *compressa*	838	100	0	99.40	AY323721.1
6	*Aloe gneissicola*	*Aloe gneissicola*	837	100	0	99.40	AY323720.1
7	*Aloe purpurea* voucher WV 99-067	*Aloe purpurea*	838	100	0	99.40	KX270418.1
8	*Aloe purpurea* voucher MAU 0014447	*Aloe purpurea*	838	100	0	99.40	KX270416.1
9	*Aloe macra* voucher WS 92-729	*Aloe macra*	838	100	0	99.40	KX270415.1
10	*Aloe deserti* voucher NMK:19.10021	*Aloe deserti*	838	100	0	99.40	KU748261.1
11	*Aloe nyeriensis* voucher NMK:24.10026	*Aloe nyeriensis*	838	100	0	99.40	KU748259.1

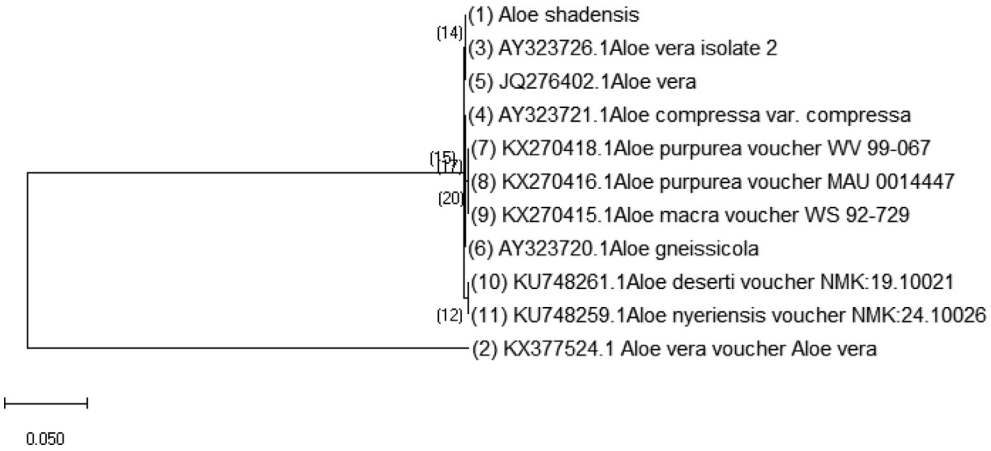

Fig. 4. Phylogenetic tree by neighbor-joining of *Aloe shadensis* with the closest species based on the *matK* gene sequence by the second pair of primers.

Table 5
The maximum likelihood estimate of substitution matrix in the constructed phylogenetic tree of *Aloe shadensis matK* gene sequence by the second pair of primers.

From\To	A	T	C	G
A	–	8.7625	3.8645	8.7967
T	7.0988	–	8.1423	3.4057
C	7.0988	18.4623	–	3.4057
G	18.3357	8.7625	3.8645	–

2.2.2. Amplification

The primers used in the amplification are listed in Table 1. Master Mix Phire Plant (Thermo Scientific) was used. The reaction mixture and cycling instructions were followed as described. GeneAmp cycler 9700 was used for amplification (Applied Biosystems). The PCR product was subjected to ExoSAP-IT (GE-Healthcare) protocol for cleanup. Before sequencing, the purity of the PCR product was ensured by gel electrophoresis (1.2% agarose). A 100 bp DNA marker (Bioatlas) was used as a molecular standard. A Bio-Rad documentation system has been used to visualize the gel under UV light.

2.2.3. Sequencing

Sequencing was done following the kit's protocol (BigDyeTerminator v3.1, Applied Biosystems). The same PCR primers have been used for sequencing. The reaction has done in An ABI-3500 DNA-Analyzer (Applied Biosystems).

2.2.4. Data analysis

The assembled sequences were used to classify the plant. This was conducted through the GenBank BLAST tool and the BOLD database systems. The final sequences have then been deposited in GenBank. Phylogenetic analysis has been carried out by MEGA-X software (Kumar et al., 2018).

3. Results

The results showed that the primers used in the amplification process for the genetic markers under study mentioned in Table 1 gave fragments of molecular weights as expected, except for the *rbcL* gene, and the result was negative, and therefore its sequence was not determined. The rest of the markers were positive, as in the order of Table 1 the *matK* gene, followed by the ITS region, using two pairs of primers in each case. The molecular weights of

Fig. 5. The Sequence alignment of *Aloe shadensis* (Query) and the first next species *Aloe vera* (Subject) based on *matK* gene sequence by the second pair of primers.

Table 6
The plant under study, *Aloe shadensis*, and the percentage of similarity with the closest species in the GenBank based on the ITS region sequence by the ITS1 pair of primers.

No.	Species description	Scientific name	Aligned sequence (bp)	Coverage (%)	E value	Similarity (%)	Accession
1	*Aloe shadensis*	*Aloe shadensis*	404	100	0	100	MZ420214
2	*Aloe retrospiciens* voucher Grace163	*Aloe retrospiciens*	367	90	0	98.91	KJ557911.1
3	*Aloe dorotheae* voucher Grace202	*Aloe dorotheae*	367	90	0	98.91	KJ557867.1
4	*Aloe camperi* voucher Grace153	*Aloe camperi*	367	90	0	98.91	KJ557857.1
5	*Aloe sinkatana* voucher Grace135	*Aloe sinkatana*	367	90	0	98.91	KC893738.1
6	*Aloe vera* voucher Grace220	*Aloe vera*	366	90	0	98.91	KC893746.1
7	*Aloe minima* voucher Klopper & Abbot464	*Aloe minima*	367	90	0	98.64	KJ557469.1
8	*Aloe trichosantha* voucher Grace151	*Aloe trichosantha*	366	90	0	98.37	KJ557922.1
9	*Aloe fleurentinorum* voucher A.G.Miller&D.Long3459	*Aloe fleurentinorum*	366	90	0	98.09	KJ557872.1
10	*Aloe ankoberensis* voucher Grace148	*Aloe ankoberensis*	367	90	0	98.09	KJ557848.1
11	*Aloe thorncroftii* isolate P7710	*Aloe thorncroftii*	357	88	0	98.88	KF013327.1

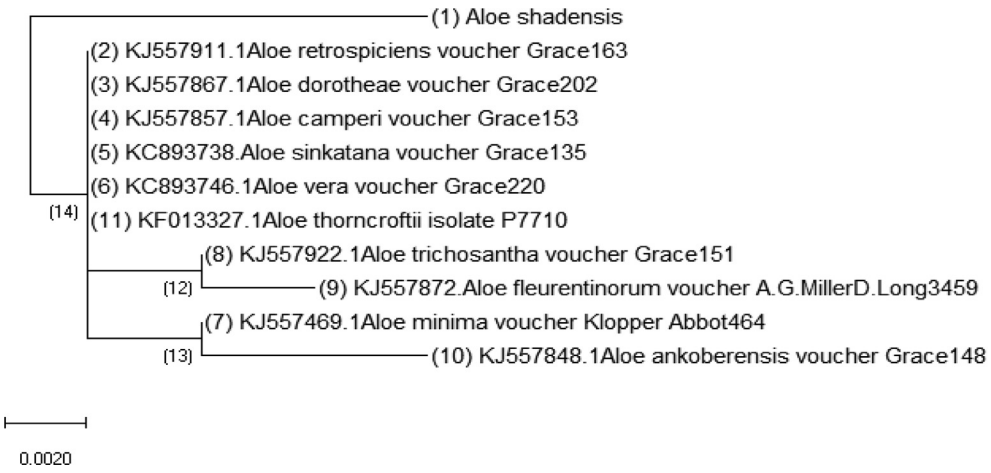

Fig. 6. Phylogenetic tree by neighbor-joining of *Aloe shadensis* with the closest species based on the ITS region sequence by the ITS1 pair of primers.

Table 7
The maximum likelihood estimate of substitution matrix in the constructed phylogenetic tree of *Aloe shadensis* ITS region sequence by the ITS1 pair of primers.

From\To	A	T	C	G
A	–	5.8594	11.6306	9.5684
T	6.2565	–	9.1819	12.0895
C	6.2565	4.6258	–	12.0895
G	4.9518	5.8594	11.6306	–

matK amplicon by first and second pair were 900 and 850, while the weights for the ITS region were 800 and 500 using ITS1 and ITS2 primers, respectively. The partial sequences of these genetic markers were obtained using the same primers, then four assembled sequences were obtained and deposited in the GenBank. The sequences were used through the BOLD system to identify the plant, and also the closest species in the GenBank were obtained. The phylogenetic trees were established, and the plant was compared in each case with the neighboring species. *Aloe shadensis* was deposited and documented here in this study with genetic markers for the first time. The results showed that through any identification system and for all the markers used, the plant genus is *Aloe*. *Aloe vera* was the most closely related species with greater extent than the other species of *Aloe*. Below are the results of each genetic marker used in order. The sequence alignment was performed in the MEGAX program using ClustalW, and evolution trees were established with Bootstrap 1000 replicates. In the case of the *matK* sequence (803 bp) using the first pair of primers (*matK*-XF and *matK*-MALP-R1) and using the BOLD system for identification, the closest species was *Aloe vera* with a percentage of 99.88%. By using the BLAST tool in the gene bank, it was also shown that the closest species is *Aloe vera* with the same percentage of 99.88% (Table 2). The phylogenetic tree for this sequence was constructed

Aloe retrospiciens voucher Grace163 internal transcribed spacer 1, partial sequence; 5.8S ribosomal RNA gene, complete sequence; and internal transcribed spacer 2, partial sequence

Sequence ID: **KJ557911.1** Length: 667 Number of Matches: 1

Range 1: 301 to 667

Score	Expect	Identities	Gaps	Strand	Frame
683 bits(355)	0.0()	363/367(99%)	0/367(0%)	Plus/Plus	

```
Query  17   ATGCGCTACTTGGTGTGAAGTGCAGAATCCCGCGATCCATCGTGTTTTTGAACGCAAGTT  76
Sbjct  301  .....A.............T.................A......A..............  360

Query  77   GCGCCCGAGGCCACCCGGCCGAGGGCACGCCTGCCTGGGCGTCACGCCTCGCGTCGCTCC  136
Sbjct  361  ............................................................  420

Query  137  GCCTACCTCGCCCTGAGCACCCCGTGCTCTTATGGCGGCGGGCGGCGGATGCGGAGATTG  196
Sbjct  421  ............................................................  480

Query  197  GCCCTCCGTGCCTTGCGGTGCGGTGGGTCGAAGTGTCGGTCGCCGGCCGGGCTTGGCACG  256
Sbjct  481  ............................................................  540

Query  257  GTGAGTGGTGGACGGACTAGCTCCCGAGCGCCGGACGCCGTGAAAAACCATCCCGATGCC  316
Sbjct  541  ............................................................  600

Query  317  GGGTACATGACAGAGATGATAAGAACCCACACCGAAGGGCGCAGCGCGCCATCGGATCGC  376
Sbjct  601  ............................................................  660

Query  377  GACCCCA  383
Sbjct  661  .......  667
```

Fig. 7. The Sequence alignment of *Aloe shadensis* (Query) and the first next species *Aloe retrospiciens* (Subject) based on ITS region sequence by the ITS1 pair of primers.

Table 8
The plant under study, *Aloe shadensis*, and the percentage of similarity with the closest species in the GenBank based on the ITS region sequence by the ITS2 pair of primers.

No.	Species description	Scientific name	Aligned sequence (bp)	Coverage (%)	E value	Similarity (%)	Accession
1	*Aloe shadensis*	*Aloe shadensis*	386	100	0	100	MZ424222
2	*Aloe vera* voucher B0709	*Aloe vera*	365	94	0	99.73	MN519271.1
3	*Aloe vera* bio-material AHP_Lot_147	*Aloe vera*	364	94	0	99.45	MK087867.1
4	*Aloe nyeriensis* voucher NMK:EA 13614	*Aloe nyeriensis*	365	94	0	98.90	MT137508.1
5	*Aloe kedongensis* voucher NMK:EA 13642	*Aloe kedongensis*	365	94	0	98.63	MT137515.1
6	*Aloe tormentorii* isolate ALTO2	*Aloe tormentorii*	357	92	0	98.60	KX689271.1
7	*Aloe volkensii* voucher NMK:EA 13619	*Aloe volkensii*	364	94	0	98.36	MT137509.1
8	*Aloe tormentorii* isolate ALTO1	*Aloe tormentorii*	354	91	0	98.31	KX689270.1
9	*Aloe retrospiciens* voucher Grace163	*Aloe retrospiciens*	327	84	2.00E-171	99.69	KJ557911.1
10	*Aloe dorotheae* voucher Grace202	*Aloe dorotheae*	327	84	2.00E-171	99.69	KJ557867.1
11	*Aloe camperi* voucher Grace153	*Aloe camperi*	327	84	2.00E-171	99.69	KJ557857.1

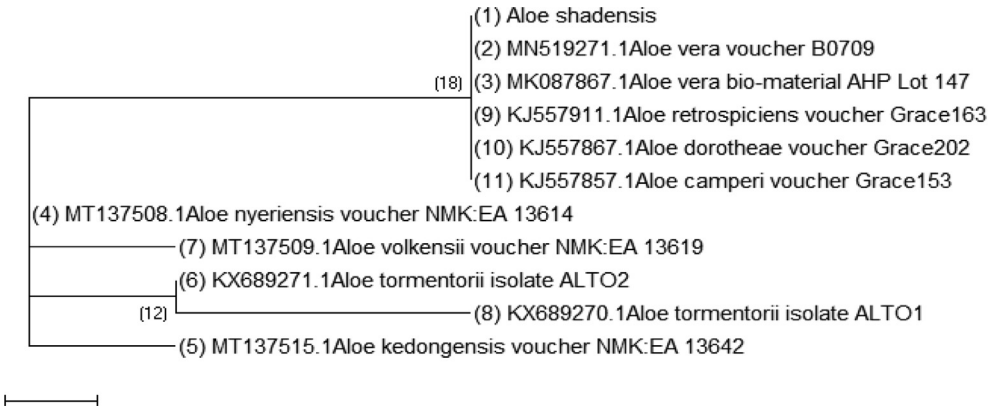

Fig. 8. Phylogenetic tree by neighbor-joining of *Aloe shadensis* with the closest species based on the ITS region sequence by the ITS2 pair of primers.

Table 9
The maximum likelihood estimate of substitution matrix in the constructed phylogenetic tree of *Aloe shadensis* ITS region sequence by the ITS2 pair of primers.

From\To	A	T	C	G
A	–	2.7125	5.8786	6.3531
T	3.1065	–	37.4319	6.1764
C	3.1065	17.2718	–	6.1764
G	3.1954	2.7125	5.8786	–

with the closest species (Fig. 2) with the estimated substitution matrix (Table 3). The overall mean distance amongst *Aloe shadensis* and the individuals represented in the phylogenetic tree was 0.1 with a standard error (S.E. equals 0.0). Fig. 3 shows the alignment of the sequence of *Aloe shadensis* and its closest species, *Aloe vera*. In the case of the *matK* sequence (825 bp) using the second pair of primers (1R_ kim and 3F_ kim) and using the BOLD system for identification, the closest species was *Aloe vera* with a percentage of 100%. By using the BLAST tool in the gene bank, it was also shown that the closest species is *Aloe vera* with a percentage of 99.76% (Table 4). The phylogenetic tree for this sequence was constructed with the closest species (Fig. 4) with the estimated substitution matrix (Table 5). The overall mean distance amongst *Aloe shadensis* and the individuals represented in the phylogenetic tree was 0.25 with a standard error (S.E. equals 0.0). Fig. 5 shows the alignment of the sequence of *Aloe shadensis* and its closest species, *Aloe vera*. In the case of the ITS region sequence (404 bp) using the ITS1 pair of primers (ITS 5a and ITS 4) and using the BOLD system for identification, the closest species was *Aloe sinkatana* with a percentage of 100%. By using the BLAST tool in the gene bank, it was also shown that the closest species is *Aloe retrospiciens* with a percentage of 98.91% (Table 6). The phylogenetic tree for this sequence was constructed with the closest species (Fig. 6) with the estimated substitution matrix (Table 7). The overall mean distance amongst *Aloe shadensis* and the individuals represented in the phylogenetic tree was 0.13 with a standard error (S.E. equals 0.0). Fig. 7 shows the alignment of the sequence of *Aloe shadensis* and its closest species, *Aloe retrospiciens*. In the case of the ITS region sequence (386 bp) using the ITS2 pair of primers (ITS S2F and ITS S3R) and using the BOLD system for identification, the closest species was *Aloe sinkatana* with a percentage of 100%. By using the BLAST tool in the gene bank, it was also shown that the closest species is *Aloe vera* with a percentage of 99.73% (Table 8). The phylogenetic tree for this sequence was constructed with the closest species (Fig. 8) with the estimated substitution matrix (Table 9). The overall mean distance amongst *Aloe shadensis* and the individuals represented in the phylogenetic tree was 0.43 with a standard error (S.E. equals 0.01). Fig. 9 shows the alignment of the sequence of *Aloe shadensis* and its closest species, *Aloe vera*. Table 10 shows a summary of the closest species to *Aloe Shadensis* using the two identification systems. An Alignment of both sequences of the *matK* gene was performed using both pairs of primers, and a sequence of 777 bp was obtained (Fig. 10). Another alignment of the two ITS region sequences was also performed, and a sequence of 332 bp was obtained (Fig. 11). By matching the final *matK* sequence and also using both identification systems, the results showed that *Aloe* vera is the closest species to *Aloe shadensis*, with a percentage of more than 99%. Also, the ITS region gave a similarity of more than 99 % to *Aloe vera* in the GenBank system. In the BOLD system, *Aloe dorotheae* was given, followed by other species of *Aloe*, then *Aloe vera* ranked 89th. In all these species, the sequence covering ratio has decreased from 322 in the first species to 194 in the case of *Aloe vera*.

Aloe vera voucher B0709 5.8S ribosomal RNA gene, partial sequence; internal transcribed spacer 2, complete sequence; and large subunit ribosomal RNA gene, partial sequence

Sequence ID: **MN519271.1** Length: 480 Number of Matches: 1
Range 1: 21 to 385

Score	Expect	Identities	Gaps	Strand	Frame
687 bits(357)	0.0()	364/365(99%)	1/365(0%)	Plus/Plus	

```
Query  9    CGAGTTTTTGA-CGCAAGTTGCGCCCGAGGCCACCCGGCCGAGGGCACGCCTGCCTGGGC  67
Sbjct  21   ...........A.............................................  80

Query  68   GTCACGCCTCGCGTCGCTCCGCCTACCTCGCCCTGAGCACCCCGTGCTCTTATGGCGGCG  127
Sbjct  81   ............................................................  140

Query  128  GGCGGCGGATGCGGAGATTGGCCCTCCGTGCCTTGCGGTGCGGTGGGTCGAAGTGTCGGT  187
Sbjct  141  ............................................................  200

Query  188  CGCCGGCCGGGCTTGGCACGGTGAGTGGTGGACGGACTAGCTCCCGAGCGCCGGACGCCG  247
Sbjct  201  ............................................................  260

Query  248  TGAAAAACCATCCCGATGCCGGGTACATGACAGAGATGATAAGAACCCACACCGAAGGGC  307
Sbjct  261  ............................................................  320

Query  308  GCAGCGCGCCATCGGATCGCGACCCCAGGTCAGGCGGGAACACCCGCTGAGTTTAAGCAT  367
Sbjct  321  ............................................................  380

Query  368  ATCAA  372
Sbjct  381  .....  385
```

Fig. 9. The Sequence alignment of *Aloe shadensis* (Query) and the first next species *Aloe vera* (Subject) based on ITS region sequence by the ITS2 pair of primers.

Table 10
Summary of the closest species to *Aloe shadensis* by GenBank and BOLD identification systems.

Genetic Marker	GenBank Identification	BOLD Identification
matK(1st pair of primers)	*Aloe vera*	*Aloe vera*
matK(2nd pair of primers)	*Aloe vera*	*Aloe vera*
ITS1	*Aloe retrospiciens*	*Aloe sinkatana*
ITS2	*Aloe vera*	*Aloe sinkatana*

4. Discussion

Combining coding and noncoding genetic markers is the best approach to DNA barcoding of plants. The most commonly investigated markers in many studies are the plastid conserved rbcL gene and the more variable matK gene. Recently, the (ITS) region also has proven a useful variable marker (Kress, 2017). So, this study was planned to investigate the efficiency of these markers in the

Sequence ID: Query_53913 Length: 837 Number of Matches: 1
Range 1: 1 to 777

Score	Expect	Identities	Gaps	Strand	Frame
1413 bits(765)	0.0()	774/778(99%)	2/778(0%)	Plus/Plus	

```
Query  27   ATGCCGGATTCAAGATGTTCCTTTTTTGCATTTATTGCGATTCTTTCTTCATGAATATCA  86
Sbjct  1    .........G..-.............................................  59

Query  87   TAATTGTAATAGTCTTCTCATTACTCAGAACAAATCTATTTATGTTTTTTCAAATGAAAA  146
Sbjct  60   ............................................................  119

Query  147  TAAAAGACTATTTCAGTTACTATACAATTCTTATGCTTTTGAATGTGAATTTTTATTAGt  206
Sbjct  120  ............................................................  179

Query  207  ttttttttCGTAAACAATCTTATTATTTACGATTAACATCTTCTGCAACTTTTCTTGAACG  266
Sbjct  180  ............................................................  239

Query  267  AACCCATTTCTATAGAAAAATAGAACATCTTCGAATAGAACATTTTTTCGTAGTATGTCG  326
Sbjct  240  ............................................................  299

Query  327  TAACTATTTTCATAGAACTCGATGGTTCTTCAAAAATCCTTTCATGCATTATGTTCGATA  386
Sbjct  300  ............................................................  359

Query  387  TCAAAGAAAGGCAATTGTTGCTTCAAGGGGGACTCATTTTCTGATGAAGAAATGGAAATC  446
Sbjct  360  ............................................................  419

Query  447  CCATTTTGTCAATTTCTGGCAATATTATTTTCGCTTTTGGTCTCGACCGTACAGAATTCA  506
Sbjct  420  ............................................................  479

Query  507  TATAAATCATTTATCAAACTATTCCTTCTATTTTCTAGGTTATTTTTCAAGTCTACTAAT  566
Sbjct  480  ............................................................  539

Query  567  AAATTCTTCGGCGGTAAGGAATCAAATGTTAGAGAATTCATTTCTAATGGATACCGTTAC  626
Sbjct  540  ............................................................  599

Query  627  TAAAAAATTTGATACCATAGTCCCAGTTATTCTTCTTATTGAATCCTTGTCTAAAGCTAA  686
Sbjct  600  ...G........................................................  659

Query  687  ATTTTGTACCGTATCAGGCCATCCTATTAGTAAGCCGATCTGGGCCGATTTCTCAGATTC  746
Sbjct  660  ............................................................  719

Query  747  TGATATTATTGATCGATTTGGTCGGATATGTAGAAATCTTTCTCATTATCACAG-GGA  803
Sbjct  720  ...........................................................C...  777
```

Fig. 10. The Sequence alignment of *Aloe shadensis* by both pairs of primers of *matK* gene.

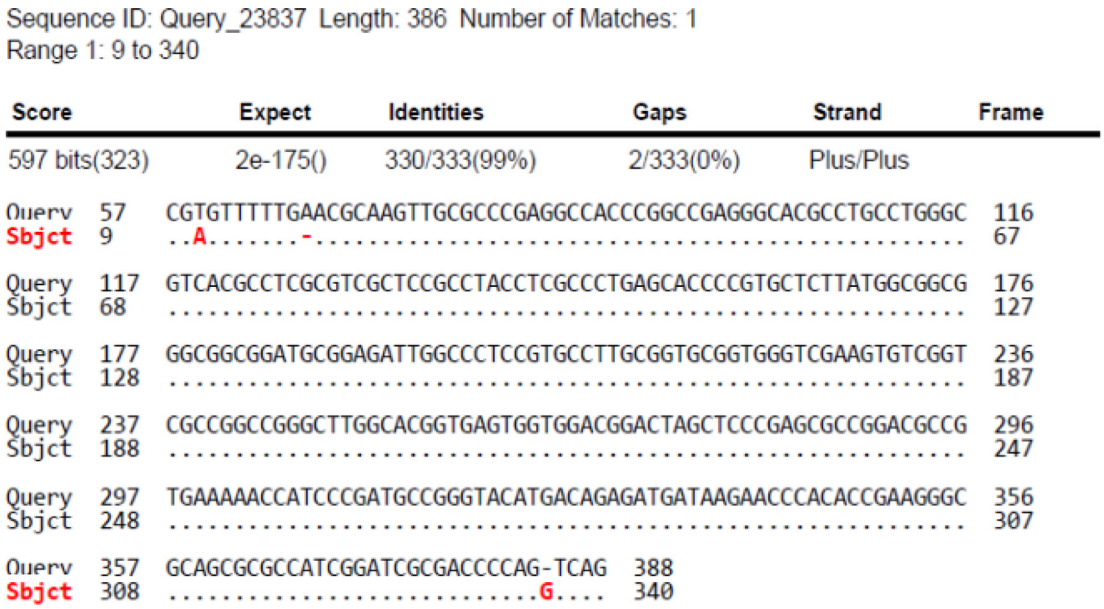

Fig. 11. The Sequence alignment of *Aloe shadensis* by both pairs of primers of ITS region.

barcoding of the rare plant *Aloe shadensis* for the first time. The primers used in the amplification of the investigated markers under study were adequate, excluding the pairs used for the *rbcL* gene, which may require more specificity for successful amplification. This is not much of a hindrance here as we have other markers to confirm the definition of plant genus, and however, it is not of much use on its own to distinguish plant species (Kang et al., 2017). The remainder markers, including the *matK* gene and the (ITS) region, were partially sequenced by the same primers used in amplification in each case. The results confirmed the identity of the plant genus, *Aloe*. In many studies, the *matK* gene and the (ITS) region genetic markers have been proved very efficient in distinguishing the plant species and hence being promise candidates in the barcoding of plants (Han et al., 2016; Yu et al., 2021). The sequence of the *matK* gene using both pairs of primers showed effectiveness in identifying the species, as many species of *Aloe* were found similar to it, although the closest was *Aloe vera*, due to the complete sequencing of its genome. Also, by performing the alignment of the two sequences of *Aloe shadensis* and obtaining a common sequence with high identity, it gave the same results in the identification, which supports its use as a distinguishing marker of the *Aloe shadensis* plant. The *matK* gene often contains many variable pieces and is considered a distinctive marker for the differentiation of plant species (Dong et al., 2012). In the case of the (ITS) region, the sequences obtained using the ITS1 and the ITS2 primers and also their alignment revealed some variations concerning *Aloe* species relatedness to *Aloe shadensis*. This was also another support to the power of species discrimination through this region. The successful ITS2 region sequence was achieved by the two different sets of primers. The sequence of the ITS1 region was not represented here equally well as in the case of the ITS2 region sequence. Usually, the ITS1 region may encounter a little difficulty in sequencing despite the successful amplification, and recently, the ITS2 region has been paid more attention (Wang et al., 2016; Yu et al., 2021). In Conclusion, the study proved the efficiency of the *mat*K gene and the ITS2 region as reliable markers in the barcoding of *Aloe shadensis*. Even though the results of the *rbcL* gene and the ITS1 region were not very promising, they can give better results with more designed specific primers. The study recommends the importance of using these genetic markers in documenting this rare plant and conducting a further study on this subject.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Carter S., Lavranos J.J., Newton L.E., Walker C.C., 2011. Aloes-The definitive guide; Kew Publishing in association with the British Cactus and Succulent Society, London.

Chaudhary, S., 2001. Flora of the kingdom of Saudi Arabia: Illustrated. Ministry of Agriculture and Water, Riyadh.

Chiou, S.-J., Yen, J.-H., Fang, C.-L., Chen, H.-L., Lin, T.-Y., 2007. Authentication of medicinal herbs using PCR-amplified ITS2 with specific primers. Planta Med. 73 (13), 1421–1426.

Das, A., Haque, S.M., Ghosh, B., Nandagopal, K., Jha, T.B., 2015. Morphological and genetic characterization of micropropagated field grown plants of Aloe vera L. Plant Tissue Cult. Biotechnol. 25 (2), 231–246. https://doi.org/10.3329/ptcb.v25i2.26257.

Dong, W., Liu, J., Yu, J., Wang, L., Zhou, S., Moustafa, A., 2012. Highly variable chloroplast markers for evaluating plant phylogeny at low taxonomic levels and for DNA barcoding. PLoS ONE 7 (4), e35071. https://doi.org/10.1371/journal.pone.0035071.

Dunning, L.T., Savolainen, V., 2010. Broad-scale amplification of matK for DNA barcoding plants, a technical note. Bot. J. Linn. Soc. 164, 1–9. https://doi.org/10.1111/j.1095-8339.2010.01071.x.

Ganie, S.H., Upadhyay, P., Das, S., Sharma, M.P., 2015. Authentication of medicinal plants by DNA markers. Plant Gene. 4, 83–99. https://doi.org/10.1016/j.plgene.2015.10.002.

Han, J., Huang, Q.L, Wu, W.R, Ma, X.Y, Yang, J.F, Zhan, R.T., 2016. Molecular identification of six species in Aloe L. based on DNA barcoding. 47, 1950-1955. https://doi.org/10.7501/j.issn.0253-2670

Jaiswal, S.K., Mahajan, S., Chakraborty, A., Kumar, S., Sharma, V.K., 2021. The genome sequence of Aloe vera reveals adaptive evolution of drought tolerance mechanisms. Iscience 24 (2), 102079. https://doi.org/10.1016/j.isci.2021.102079.

Kang, Y., Deng, Z., Zang, R., Long, W., 2017. DNA barcoding analysis and phylogenetic relationships of tree species in tropical cloud forests. Sci. Rep. 7 (1), 1–9. https://doi.org/10.1038/s41598-017-13057-0.

Kress, W.J., 2017. Plant DNA barcodes: Applications today and in the future. J. Syst. Evol. 55 (4), 291–307. https://doi.org/10.1111/jse.12254.

Kumar, S., Stecher, G., Li, M., Knyaz, C., Tamura K., 2018. MEGA X: molecular evolutionary genetics analysis across computing platforms. Mol. Biol. Evol. 35 (6), 1547-1549. https://doi.org/10.1093/molbev/msy096.

Lledo, M.D., Crespo, M.B., Cameron, K.M., Fay, M.F., Chase, M.W., 1998. Systematics of plumbaginaceae based upon cladistic analysis of rbcL sequence data. Syst. Bot. 23 (1), 21. https://doi.org/10.2307/2419571.

Stanford, A.M., Harden, R., Parks, C.R., 2000. Phylogeny and biogeography of Juglans (Juglandaceae) based on matK and ITS sequence data. Am. J. Bot. 87 (6), 872–882.

Techen, N., Parveen, I., Pan, Z., Khan, I.A., 2014.DNA barcoding of medicinal plant material for identification. Curr. Opin. Biotechnol. 25, 103-110. https://doi.org/10.1016/j.copbio.2013.09.010.

Wang, X.Y., Zheng, S.H., Liu, Y., Han, J.P., 2016. ITS2, a Better DNA Barcode than ITS in Identification of Species in Artemisia L. Chin. Herb Med. 8 (4), 352–358. https://doi.org/10.1016/S1674-6384(16)60062-X.

Wattoo, J.I., Saleem, M.Z., Shahzad, M.S., Arif, A., Hameed, A., Saleem, M.A., 2016. DNA Barcoding: Amplification and sequence analysis of rbcl and matK genome regions in three divergent plant species. Adv. Life Sci. 4 (1), 03–07.

Yu, J., Wu, X.i., Liu, C., Newmaster, S., Ragupathy, S., Kress, W.J., 2021. Progress in the use of DNA barcodes in the identification and classification of medicinal plants. Ecotoxicol. Environ. Saf. 208, 111691. https://doi.org/10.1016/j.ecoenv.2020.111691.