Print ISSN: 0375-9237
Online ISSN: 2357-0350

EGYPTIAN JOURNAL OF BOTANY (EJBO)

DNA barcoding and tribal placement of *Forsskaolea tenacissima* L. (Urticaceae) in Western Saudi Arabia: Insights from *rbcL* and ITS DNA markers

Sameer H. Qari, Muath F. Alrehaili, Ahmed El-Banhawy

PUBLISHED BY

DNA barcoding and tribal placement of *Forsskaolea tenacissima* L. (Urticaceae) in Western Saudi Arabia: Insights from *rbcL* and ITS DNA markers

Sameer H. Qari[1], Muath F. Alrehaili[1], Ahmed El-Banhawy[2]
[1]Biology Department, Aljumum University College, Umm Al-Qura University, Makkah, Saudi Arabia
[2]Botany and Microbiology Department, Faculty of Science, Suez Canal University, Ismailia, Egypt

The Urticaceae family, encompassing 53 genera and 2625 species, exhibits a subcosmopolitan distribution. *Forsskaolea* L. is a small genus of seven species. It is represented in Saudi Arabia by two species *F. tenacissima* L. and *F. viridis* Ehrenb. The current research elucidates the validity of DNA barcoding in species authintican and taxonomic identity of *Forsskaolea tenacissima,* indigenous to the western region of Saudi Arabia. In this investigation, the Internal Transcribed Spacer (ITS) and ribulose bisphosphate carboxylase large subunit (*rbc*L) gene loci were designated as molecular markers to facilitate precise taxonomic documentation and the detection of novel *F. tenacissima* variants sourced from the western region of the Kingdom of Saudi Arabia. They were employing the Basic Local Alignment Search Tool (BLAST). The ITS and *rbc*L exhibited robust amplification across the *F. tenacissima*, underscoring their utility as universal barcodes. ITS region not only aids in confirming the identification of *F. tenacissima* at the generic and species levels but also demonstrates superior discriminatory power in distinguishing between closely related species or variants within the genus. The phylogenetic analysis of the individual datasets of the DNA sequences of *rbc*L and ITS addressed the monophyly of the species under investigation. In contrast, the combined phylogenetic analysis could not address the monophyly of *Forsskaolea* species in Saudi Arabia. The phylogenetic analysis on individual and combined datasets successfully placed *F. tenacissima* within the monophyletic tribe Forsskaolea. This discovery marks the initial inclusion of such information in the Flora of Saudi Arabia.

Keywords: Nanofertilizer, Wheat, Nanochitosan, Carbohydrate, Protein, Elements

ARTICLE HISTORY
Submitted: September 30, 2023
Accepted: May 28, 2024

CORRESPONDANCE TO
Ahmed El-Banhawy,
Botany and Microbiology Department,
Faculty of Science, Suez Canal University, Egypt
Email: ahmedbanhawy@science.suez.edu.eg
Sameer H.-Qari
Bology Department, Aljumum University College, Umm Al-Qura University, Saudi Arabia
DOI:10.21608/ejbo.2024.239918.2515

EDITOR
Prof. Reda Gaafar,
Botany and Microbiology Department,
Faculty of Science, Tanta University, Egypt
Email: redagaafar@gmail.com

INTRODUCTION

The Urticaceae Juss. (Nettle family) comprises 53 genera and 2625 species. (Wu et al., 2013; Kim et al., 2015, Christenhusz & Byng 2016), Displays a subcosmopolitan distribution, with a majority thriving in the moist tropics. Antoine-Laurent de Jussieu categorized it into three groups based on inflorescence morphology. Gaudichaud (1830: 491) later revised the family, organizing the genera into five tribes or subfamilies, including Elatostemateae (as "Elatostemeae"), Urereae, Boehmerieae, Parietarieae, Forsskaoleae (as "Forskalieae"), and Cecropieae, marking the introduction of tribal names (Conn & Hadiah, 2009). Weddell (1854, 1856, 1869) subsequently attempted to classify the genera into natural units based on gender distribution, adopting Gaudichaud's proposed subdivisions as tribes within the Urticaceae. Berg (1977, 1989) and Friis (1989, 1993) delineated Cecropiaceae, Moraceae, and Urticaceae as distinct families. Conn and Hadiah (2009) proposed validating Gaudichaud's subdivisions as tribes, reintroducing Cecropieae to tribal rank, suggesting tribal names including Boehmerieae, Elatostemateae, Forsskaoleeae, Parietarieae, Urticeae, and Cecropieae. Molecular studies reveal the monophyletic nature of Urticaceae. Presently, the family comprises six tribes (Boehmerieae, Elatostemateae, Forsskaoleeae, Parietarieae, Urticeae, Cecropieae), each demonstrated to be monophyletic based on DNA sequence data (Wu et al. 2013, Kim et al. 2015).

Forsskaolea L. represents a small genus comprising seven species: *Forsskaolea angustifolia* Retz., *Forsskaolea candida* L.f., *Forsskaolea griersonii* A.G. Mill. & J.A. Nyberg., *Forsskaolea hereroensis* Schinz., *Forsskaolea procridifolia* Webb., *Forsskaolea tenacissima* L., and *Forsskaolea viridis* Ehrenb. ex Webb. These plants possess non-stinging hairs and feature dot-like accumulations of mineral material on their green parts. The genus received its name as a tribute to Swedish botanist Peter Forsskål (Gilman et al. 1905). *Forsskaolea* leaves is trinerved, alternate, dentate, and crenate. The inflorescence is cymose, bisexual, and arises from the axils, subtended by campanulate, densely pilose involucres of 3-6 bracts. Flowers are minute and unisexual, with pistillate flowers surrounded by a perigynium of staminate flowers. Staminate flowers possess a 3-5 lobed calyx, while pistillate flowers lack calyx lobes. They bear solitary stamens and erect lanate ovaries without style (Plants of the World Online). Achenes are oval to elliptic, compressed, lanate, and surrounded by lanate bracts. *Forsskaolea* species are distributed in the southern Palearctic, from the Canary Isles and southeastern Spain, extending eastwards to Pakistan and western India (Flora of Pakistan).

Forsskaolea tenacissima is abundant in Saudi Arabia, particularly in the southwest (Egazy et al., 1998) and western regions (Qari et al., 2021). Renowned for its rich active compounds (Assaf et al., 2020), it enjoys broad utilization in alternative medicine (Shah et al., 2010; Assaf et al., 2017; Attia et al., 2021), particularly in treating various ailments such as wound healing and gall bladder stone removal (Sher et al., 2017). Given its extensive use in traditional medicine, the authentication and elucidation of evolutionary relationships among F. tenacissima species have become imperative.

DNA barcoding, comprised of a standardized short DNA sequence unique to each species (Kress and Erickson, 2008), has emerged as a valuable tool for biodiversity investigation, monitoring, molecular phylogeny, and evolution (Pei et al., 2017). Its application spans a wide array of biodiversity studies, including species identification (Hollingsworth, 2008; Hosein et al., 2017; Fouad et al., 2019), the discovery of new taxa (Bell et al., 2012; Hashim et al., 2021; Shamso & Fouad, 2018), and solving taxonomic challenges across large taxonomic groups (Ojeda et al., 2014; El-Sherif & Ibrahim, 2020) and detached taxonomic problems (Feau et al., 2011). Moreover, DNA barcoding contributes to species conservation efforts (Yesson et al., 2011; Fouad et al., 2022) and the delineation of plant communities through phylogenetic tree construction (Joly et al., 2014). Various studies have successfully utilized DNA barcoding markers such as *rbc*L, *mat*K, and ITS for species identification and differentiation (Anvarkhah et al., 2013; Faried et.al., 2018; Pathak et al., 2018; El-Banhawy et al., 2020(a and b); Jamdade et al., 2022), as well as for detecting adulteration in medicinal herb products and contamination issues (Han et al., 2016; Bansal et al., 2018). The effectiveness of the ITS and *rbc*L as DNA barcodes has been recognized globally (CBOL, 2009; Hollingsworth et al., 2009).

The investigation into the phylogeny of *F. tenacissima* promises to unveil the biological mechanisms underlying medicinal plants, aiding in establishing DNA-based authentication methods to monitor adulteration and enhance their efficacy in the pharmaceutical industry. This study employs the Basic Local Alignment Search Tool (BLAST) to analyze homology modeling and functional annotation of ITS and *rbc*L sequences across *F. tenacissima*. Furthermore, the study rigorously examines the amplification and sequencing of *rbc*L and ITS gene markers for authentication and phylogenetic inquiry in *F. tenacissima*. Additionally, investigating the phylogeny of *F. tenacissima* will shed light on the evolutionary dynamics of medicinal plants, providing insights into their genetic diversity and evolutionary relationships. This more profound understanding of the phylogenetic relationships within *F. tenacissima* populations will contribute to developing effective conservation strategies and sustainable management practices for these valuable plant resources.

MATERIALS AND METHODS

Plant Materials

Fifty-four leaf samples were collected, with nine samples obtained from each of the six regions in western Saudi Arabia: Makkah, Aljumum, Taif, Al-Madinah, Wadi Al-Fora`a, and Al-Yutamah. Collection took place between 2021 and 2022, spanning latitudes of '21° 21' N to 23° 87'' N and longitudes 39° 36' E to 40° 42'' E (Figure 1). The last author of the current study verified the authenticity of the samples. Before DNA extraction, the samples underwent a thorough washing process under running water to eliminate dirt and debris.

DNA extraction, PCR amplification, and DNA sequencing

Genomic DNA extraction followed the CTAB protocol with modifications described by Tiwari et al. (2012) and EL-Banhawy and Al-Juhani (2019). Additionally, the GenEluteTM Plant Genomic DNA KIT was employed. PCR amplification of two barcode loci, namely chloroplast DNA (cpDNA) *rbc*L and nuclear ribosomal DNA (nrDNA) ITS, was conducted using a Thermal Cycler TC-TE BOE (BOEC, Germany). PCR protocols adhered to procedures outlined by White et al. (1990), Ali et al. (2015), and Maloukh et al. (2017).

Figure 1. Map of the study sites in the western region of Saudi Arabia, (A) geographical boundaries of Saudi Arabia. (B) study area and sample collection in the western region of Saudi Arabia. Google (2023) [Google map for Saudi Arabia].

PCR reactions were performed in 20 µL volumes for ITS, *mat*K, and *rbc*L loci, comprising 10 µL Taq DNA polymerase with 25 µM MgCl2, 10 µM dNTPs, and 1µL of each primer. Amplifications were carried out using an Applied Biosystems®-Veriti-96-well thermal cycler (Thermo Fisher Scientific-Fisher Scientific AS-Postboks 114, Smestad-0309 Oslo–Norway). The thermal cycling program for amplification of the *rbc*L region included an initial denaturation at 94℃ for 5 min, followed by 35 cycles of denaturation at 94℃ for 30 s, annealing at 55℃ for 45 s, extension at 72℃ for 90 s, and a final extension at 72℃ for 10 min. For the *mat*K and ITS regions, thermal cycling involved an initial denaturation at 97℃ for 5 min, followed by 30 cycles of denaturation at 97℃ for 30 s, annealing at 58℃ for 45 s, extension at 72℃ for 45 s, and a final extension at 72℃ for 10 min. The primers utilized in this study are listed in Table 1. PCR products of interest were separated by electrophoresis in 2% agarose gels containing ethidium bromide (0.5 mg/mL) in 1X TBE buffer. A 100-bp-plus DNA ladder served as the molecular size standard. PCR results were documented using a Gel DocTM XR + System with Image LabTM Software (Bio-Rad) (Haglund, 2022).

DNA Sequencing

Following manufacturer recommendations, PCR products underwent purification using the ExoSAP-ITTM kit from Thermo Fisher Scientific (USA). Subsequently, sequencing reactions were carried out using the BigDye XTerminator kit (Applied Biosystem, USA) following the principles outlined by Sanger et al. (1977). Identical primers utilized for both rounds of amplification were also employed for sequencing purposes.

Bioinformatics

Sequence manipulation: The highest precision in DNA sequencing was attained through sequence refinement using Bio-Edit (version 7.2) software. Subsequently, employing the Basic Local Alignment Search Tool (BLAST) method as elucidated by Altschul et al. (1990), each modified sequence was ascribed to a specific taxon by juxtaposing it with nucleotide sequences from GenBank and the Barcode of Life Database (BOLD). Sequences exhibiting high E-values were eliminated from the dataset, while lower E-values indicated closer proximity of the query sequence to the reference sequence within the database (Madden, 2002).

Sequence Alignment: FASTA-formatted files containing sequences were aligned using BioEdit version 7.0 (Hall, 1999). Multiple alignments of DNA sequences were performed utilizing Clustal W 1.83 with default parameters (Tamura et al., 2013). Alignment curation was conducted using the Gblocks method, accessible via the advanced analysis option on http://www.phylogeny.fr/ (Dereeper et al., 2010). Standard sequences sourced from GenBank were employed for aligning species with singular samples. The aligned *rbc*L and ITS sequences were merged to construct two-locus DNA barcodes using the Sequence Matrix program (version 1.9), and the combined alignment was exported as NEXUS files.

Molecular Identification: Identification success was evaluated using the best-close match approach outlined by Taxon DNA (Meier et al., 2006). Each dataset was assigned a threshold value (T) calculated as the percentage of divergence at which 95% of all intra-specific distances were observed.

Table 1. The primer sets that are used for DNA barcoding.

Primer	Sequence F/R	References
*rbc*LaF *rbc*La-R	5′-ATG TCA CCA CAA ACA GAG ACT AAA GC-3' 5′-GTA AAA TCA AGT CCA CCR CG -3'	**Fazekas *etal*., 2008**
*mat*K3F *mat*K1R	5′-CGT ACA GTA CTT TTG TGT TTA CGA G-3' 5′-ACC CAG TCC ATC TGG AAA TCT TGG TTC-3'	**Fazekas *et al*,2008**
ITS2-F1 ITS2-R1	5′-ATG CGA TAC TTG GTG TGA AT-3' 5′-GAC GCT TCT CCA GAC TAC AAT-3'	**Chen *et al*.,2010**
ITS5-F1 ITS4-R1	5′-GGA AGT AAA AGT CGT AAC AAG G-3' 5′-TCC TCC GCT TAT TGA TAT GC-3'	**White *et al*.,1990**
ITS1-F1 ITS4-R1	5′-TCC GTA GGT GAA CCT GCG G-3' 5′-TCC TCC GCT TAT TGA TAT GC-3'	**White *et al*.,1990**

Phylogenetic Analysis: Bayesian phylogenetic inferences (BI) were generated using MrBayes software (version 3.2) (Ronquist et al., 2012). Three distinct datasets were analyzed and categorized into concatenation datasets of chloroplast DNA (cpDNA) and nuclear DNA (ncDNA) markers, along with single-locus datasets (two datasets). The best nucleotide substitution model (GTR+I+G) for each alignment was selected using PAUP (Swofford and Sullivan, 2003), MrModelblock from MrModeltest (Nylander, 2004), and Akaike information criteria (AIC) (Akaike, 1974). Four chains were employed in each independent Bayesian analysis to test for convergence, with appropriate tree sampling across generations. The Markov chain Monte Carlo (MCMC) Tracer Analysis program (version 1.6.0, 2003) was utilized to determine burn-in by plotting negative log likelihoods against generation time (Rambaut et al., 2013). The fully compatible tree was visualized and edited using TreeGraph 2 software (version 2.0.50-314 beta) (Stöver and Müller, 2010), while posterior probabilities (PP) were employed to gauge clade support. Additional phylogenetic trees were generated using MEGA6 (Tamura et al., 2013) and three other algorithms (neighbor-joining, maximum parsimony, and maximum likelihood). These ordinal phylogenies were compared against the topologies described by APG III (2009) to assess any conflicts.

RESULTS

DNA isolation and PCR amplification

The genomic DNA extraction from *F. tenacissima* leaf samples was successfully done across the six collection sites utilizing the kit-based method without encountering any discernible issues. However, DNA isolates obtained through the CTAB method exhibited diminished purity and fluctuating concentration across most samples. Table 2 shows that fifteen primer combinations were employed in the current study. Nine primer combinations were obtained from previously published work, and six were newly designed. The six newly designed primers failed to yield robust PCR amplification reactions, while two primer combinations of the previously published primers gave a good yield. Significantly, successful amplification of the target DNA sequence in *F. tenacissima* was primarily achieved using the primer sets *rbc*La(F-R) and ITS1-F1/ITS4-R1. Furthermore, Table 2 elucidates primer success rates across diverse collection regions and the corresponding sample numbers from each site.

The *rbc*La(F-R) and ITS1-F1/ITS4-R1 primer sets exhibited high success rates, reaching 90.7% and 94.4%, respectively, as illustrated in Figure 2.

DNA sequencing

The recovery of DNA sequences for *rbc*L and ITS was consistently successful, achieving amplification and sequencing success rates of 94.4% and 100%, respectively. Each region was subjected to three repetitions of primer-based DNA sequencing, ensuring the robustness and reliability of the results. Identical sequences were obtained for corresponding regions across all repetitions, indicating high reproducibility and consistency. Consequently, sequence analysis was conducted for all samples of *F. tenacissima* utilizing each primer, as detailed in Table 3. This meticulous approach facilitated the comprehensive examination and comparison of genetic data across multiple samples, enhancing the validity and depth of our findings.

In this study, all successfully representative sequences of the *rbc*L, and ITS genes were meticulously documented and submitted to the GenBank at NCBI for public access and reference. Currently, the *F. tenacissima* gene sequences for *rbc*L and ITS are accessible under specific accession numbers, as delineated in Table 4.

DNA barcoding

The data outlined in Table 5 illustrate the utilization of the BLAST algorithm for DNA barcoding of *F. tenacissima*. Newly developed cpDNA (*rbc*L) and nrDNA (ITS) marker sequences were employed as barcodes. Both markers exhibited effective performance in genus and species-level identification. Specifically, the *rbc*L marker achieved a maximum query coverage percentage of 95%, while the ITS marker reached 52%. Furthermore, the most significant similarity rates observed were 99% for *rbc*L and 98% for ITS, respectively. These findings underscore the utility and reliability of the *rbc*L and ITS markers in facilitating accurate DNA barcoding and species identification in *F. tenacissima*.

Phylogenetic reconstruction

The evolutionary relationships among *Forsskaolea* species were investigated using two DNA markers, *rbc*L, and ITS. Bayesian Inference (BI) trees constructed based on ITS and *rbc*L sequences revealed the categorization of species into two major clusters.

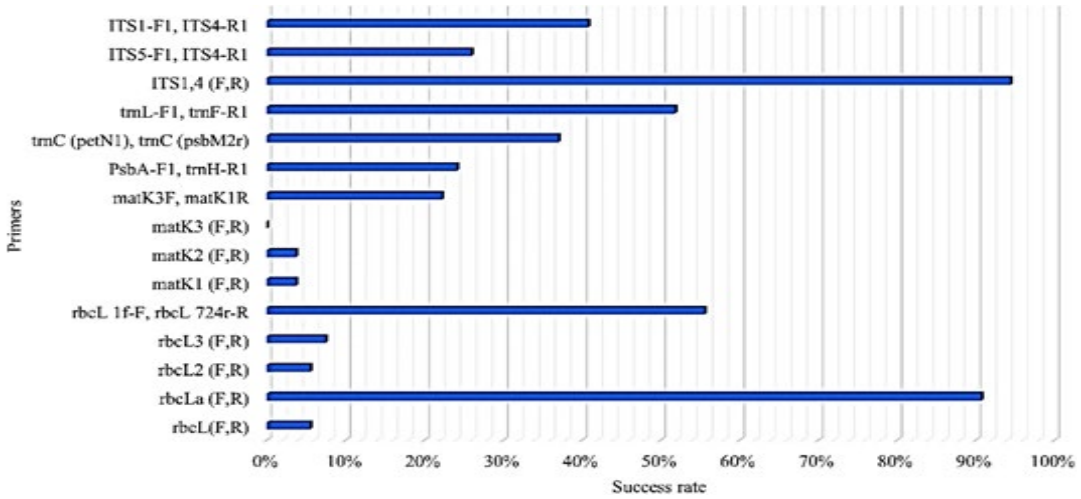

Figure 2. The success % of PCR amplification of designed and published primers bycollection sites.

Table 2. The success percentage of PCR amplification of designed and published primers percollection site.

Marker	Primer Origin	Primer Name	Success percentage per number of samples per site						Average Percentage
			1	2	3	4	5	6	
*rbc*L	Designed	*rbc*L(F,R)	11	0.0	0.0	11	11	0.0	5.5
	Published	*rbc*La (F,R)	88	66	88	100	100	100	90.7
	Designed	*rbc*L2 (F,R)	0.0	0.0	0.0	11	0.0	22	5.5
	Designed	*rbc*L3 (F,R)	0.0	22	11	0.0	11	0.0	7.4
	Published	*rbc*L *1f*-F, *rbc*L, *724r*-R	100	0.0	77	0.0	55	100	55.55
*mat*K	Designed	*mat*K1 (F,R)	0.0	0.0	11	0.0	0.0	11	3.7
	Designed	*mat*K2 (F,R)	0.0	11	0.0	0.0	11	0.0	3.7
	Designed	*mat*K3 (F,R)	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	Published	*mat*K3F, *mat*K1R	33	11	22	0.0	44	22	22.2
trn	Published	*Psb*A-F1, *trn*H-R1	33	22	11	44	11	22	24.07
	Published	*trn*C (petN1), *trn*C *(psb*M2r*)*	66	55	22	55	11	11	37.03
	Published	*trn*L-F1, *trn*F-R1	11	22	66	88	44	77	51.8
ITS	Published	ITS1,4 (F,R)	100	100	88	88	100	88	94.4
	Published	ITS5-F1, ITS4-R1	0.0	77	0.0	0.0	0.0	77	25.92
	Published	ITS1-F1, ITS4-R1	55	0.0	77	44	0.0	66	40.74

Table 3. Sequencing success of DNA barcode loci

Sequence Effectiveness	DNA barcodes loci	
	***rbc*L**	**ITS**
Successful sequencing	17/18 (94.4%)	18/18 (100%)
Unsuccessful sequencing	5.6%	0.00%
Total number of samples	54 (9 replicates for each collection site)	

Table 4. GenBank accession numbers of DNA barcoding of F. tenacissima collected from different sites in western region of Saudi Arabia

No.	Herbarium voucher	Collection site	Accession No. *rbc*L	Accession No. ITS
1	FtMac7	Makkah	OR351932	OR364720
2	FtTai2	Taif	OR351934	OR364722
3	FtMad1	Madinah	-	OR364725
4	FtAlj4	Aljumum	OR351933	OR364721
5	FtWad3	Wadi Al-Fora`a	OR351935	OR364723
6	FtAly8	Al-Yutamah	OR351936	OR364724

Table 5. DNA barcode of *rbc*L and ITS gene loci with Urticaceae using BLASTn.

No.	Species name	Gene loci	Query coverage	Similarity	E-value	Accession number
1	*Australina flaccida*	*rbc*L	95%	98%	0.0	KF138109.1
2	*Australina pusilla*	*rbc*L	95%	99%	0.0	KT626669.1
3	*Debregeasia elliptica*	*rbc*L	95%	97%	0.0	KF138138.1
4	*Debregeasia orientalis*	*rbc*L	95%	98%	0.0	LC690129.1
5	*Droguetia ambigua*	*rbc*L	93%	98%	0.0	AM235161.1
6	*Droguetia iners*	*rbc*L	95%	98%	0.0	KF138154.1
7	*Forsskaolea angustifolia*	*rbc*L	95%	99%	0.0	KF138170.1
8	*Forsskaolea tenacissima*	*rbc*L	95%	99%	0.0	KY656706.1
9	*Parietaria debilis*	*rbc*L	93%	98%	0.0	HM850235.1
10	*Parietaria alsinifolia*	*rbc*L	95%	98%	0.0	MH185902.1
11	*Parietaria micrantha*	*rbc*L	95%	98%	0.0	KF138207.1
12	*Parietaria decoris*	*rbc*L	95%	98%	0.0	OP711536.1
13	*Didymodoxa caffra*	*rbc*L	95%	98%	0.0	KF138151.1
14	*Hemistylus macrostachya*	*rbc*L	95%	97%	0.0	KF138182.1
15	*Pouzolzia guineensis*	*rbc*L	95%	97%	0.0	KF138235.1
16	*Pouzolzia poeppigiana*	*rbc*L	95%	98%	0.0	MH358141.1
17	*Pouzolzia zeylanica*	*rbc*L	95%	97%	0.0	MN205261.1
18	*Boehmeria nivea*	*rbc*L	95%	97%	0.0	MN205257.1
19	*Cecropia pachystachya*	*rbc*L	43%	97%	0.0	MH358056.1
20	*Pourouma guianensis*	*rbc*L	44%	96%	0.0	MH358139.1
21	*Australina flaccida*	ITS	14%	97%	0.0	KF137801.1
22	*Australina pusilla*	ITS	32%	81%	0.0	OM892758.1
23	*Debregeasia elliptica*	ITS	15%	96%	0.0	KF137829.1
24	*Debregeasia orientalis*	ITS	14%	97%	0.0	KF137833.1
25	*Debregeasia longifolia*	ITS	15%	96%	0.0	KF137831.1
26	*Debregeasia saeneb*	ITS	15%	96%	0.0	KF137835.1
27	*Debregeasia squamata*	ITS	15%	96%	0.0	KF137837.1
28	*Droguetia ambigua*	ITS	19%	98%	0.0	KF137843.1
29	*Droguetia iners*	ITS	15%	98%	0.0	KF137844.1
30	*Forsskaolea angustifolia*	ITS	52%	96%	0.0	KF137860.1
31	*Forsskaolea tenacissima*	ITS	51%	98%	0.0	MH357892.1
32	*Forsskaolea candida*	ITS	48%	96%	0.0	KM586475.1
33	*Lichtensteinia globosa*	ITS	31%	82%	0.0	EU434672.1
34	*Boehmeria malabarica*	ITS	15%	96%	0.0	FJ750387.1
35	*Oreocnide frutescens*	ITS	33%	86%	0.0	KF137888.1
36	*Oreocnide pedunculata*	ITS	30%	85%	0.0	MH357912.1
37	*Oreocnide rubescens*	ITS	27%	86%	0.0	MH357913.1
38	*Oreocnide tonkinensis*	ITS	28%	85%	0.0	MH357914.1
39	*Cecropia pachystachya*	ITS	29%	84%	0.0	MK601859.1
40	*Eremophila brevifolia*	ITS	23%	88%	0.0	MN411535.1

The reconstructed phylogenetic tree based on *rbc*L sequences encompassed 54 sequences, comprising five Saudi *Forsskaolea* species, 31 sequences from non-Saudi *Forsskaolea* species sourced from GenBank, and 18 sequences from ten genera serving as outgroups. Despite the comprehensive dataset, the Bayesian inference (BI) analysis of the *rbc*L sequences failed to differentiate between the two *Forsskaolea* species, *F. angustifolia* and *F. tenacissima,* native to Saudi Arabia. However, the phylogenetic analysis incorporating both *rbc*L gene sequences confirmed the monophyly of *F. tenacissima* in Saudi Arabia, marking the first investigation of its kind in the region. Although the branch support may not be notably strong, the credibility of the results remains high due to the reliability and robustness of Bayesian inference (BI) analysis, widely recognized as one of the most reliable methods in phylogenetic studies (Figure 3-1). Similarly, the phylogenetic analysis of the ITS region included six individuals representing six genera of the 54 *Forsskaolea* species, with more than twelve species acting as outgroups. The analysis indicated that the genus *Forsskaolea* is monophyletic (Figure 3-2).

However, the combined phylogenetic tree of both *rbc*L and ITS DNA sequences could not address the monophyly of *Forsskaolea* species in Saudi Arabia (Figure 4). The phylogenetic analysis on individual and combined datasets of rbcL and ITS DNA sequences successfully positioned *F. tenacissima* within the monophyletic tribe Forsskaolea. This discovery marks the initial inclusion of such information in the Flora of Saudi Arabia.

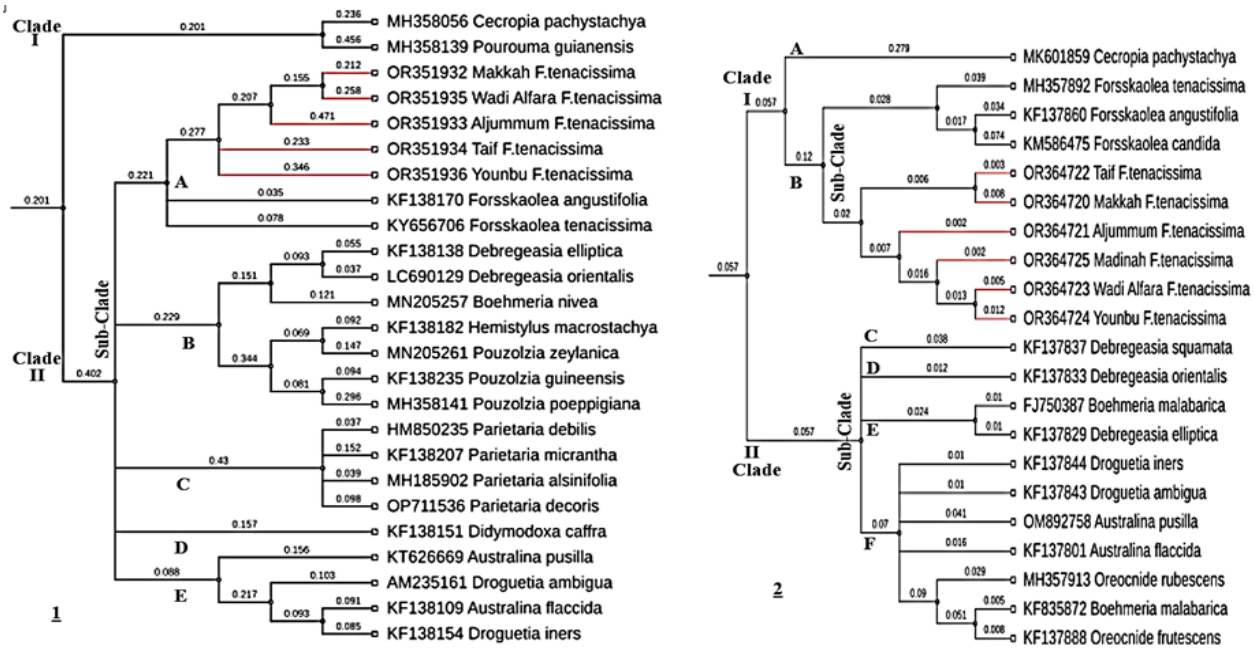

Figure 3. Bayesian all compatible tree inferred from (1): *rbc*L (2): ITS gene loci of *Forsskaolea* species. (I, II): clades, (A-E): sub-clades, (Red coloured branches): Saudi *Forsskaolea*, Numbers above branches is a posterior probability (PP).

DISCUSSION

In this study, morphological characterization and authenticated DNA barcoding were employed for the first time to analyze *F. tenacissima* in the Kingdom of Saudi Arabia. The taxonomic position within the genus and its monophyletic origin were elucidated using two molecular markers: nuclear nrDNA ITS and plastid cpDNA *rbc*L. Fifty-four samples of *F. tenacissima* were subjected to bidirectional sequencing, yielding consensus sequences without discernible issues. Our findings align with previous studies by Lahaye et al. (2008), Gonzalez et al. (2009), Chen et al. (2010), and Parmentier et al. (2013).

Notably, the *rbc*L and ITS genes exhibited higher efficiency (90 to 95%) in amplification and sequencing compared to *mat*K (22%), consistent with El-Banhawy and Al-Juhani (2019). *F. tenacissima* proved challenging to amplify by matK, possibly due to rapid evolution, as Gillman et al. (2010) suggested. Furthermore, the efficiency of *mat*K PCR, conducted using four primer pairs, was lower for *F. tenacissima*, consistent with the findings of Gonzalez et al. (2009) and Kress et al. (2010).

Additionally, we employed four barcode DNA regions (ITS, *mat*K, *trn*, and *rbc*L) to identify 54 samples of *F. tenacissima* collected from six regions. PCR amplification of *mat*K and trn loci demonstrated lower efficacy (22 to 51%) compared to ITS and *rbc*L loci, which exhibited better PCR amplification (95%) and sequencing effectiveness (94% for *rbc*L and 100% for ITS). Nevertheless, the recognition efficiency at the genus (100%) and species (100%) levels were high, consistent with recommendations by Kress et al. (2005), Newmaster et al. (2006), Luo et al. (2010), Ali et al. (2015), and Maloukh et al. (2017) advocating for the use of *rbc*L and ITS barcodes for plant identification. Kress et al. (2005) highlighted the substantial interspecific divergence of ITS, making it the most frequently sequenced region in plant phylogenetic research.

For phylogenetic analysis, a global primer is crucial for the utility of DNA barcoding (Kress et al., 2005). Our results demonstrated that the ITS and *rbc*L barcode regions yielded the best results in PCR sequencing and amplification of F. tenacissima. The low success rate of *trn* and *mat*K may stem from difficulties with secondary structure development, resulting in poor sequence information, excessive copies, and other issues, as observed by Starr et al. (2009) and Xiao et al. (2010). Based on Gonzalez et al. (2009), DNA barcoding may be a promising tool for identifying errors in species identification. Other barcode primers, such as *trn*C (*pet*N1), *psb*A-*trn*H, and *trn*C (*psb*M2r), were attempted for *F. tenacissima* but were not recommended due to low amplification rates Thomas, (2009).

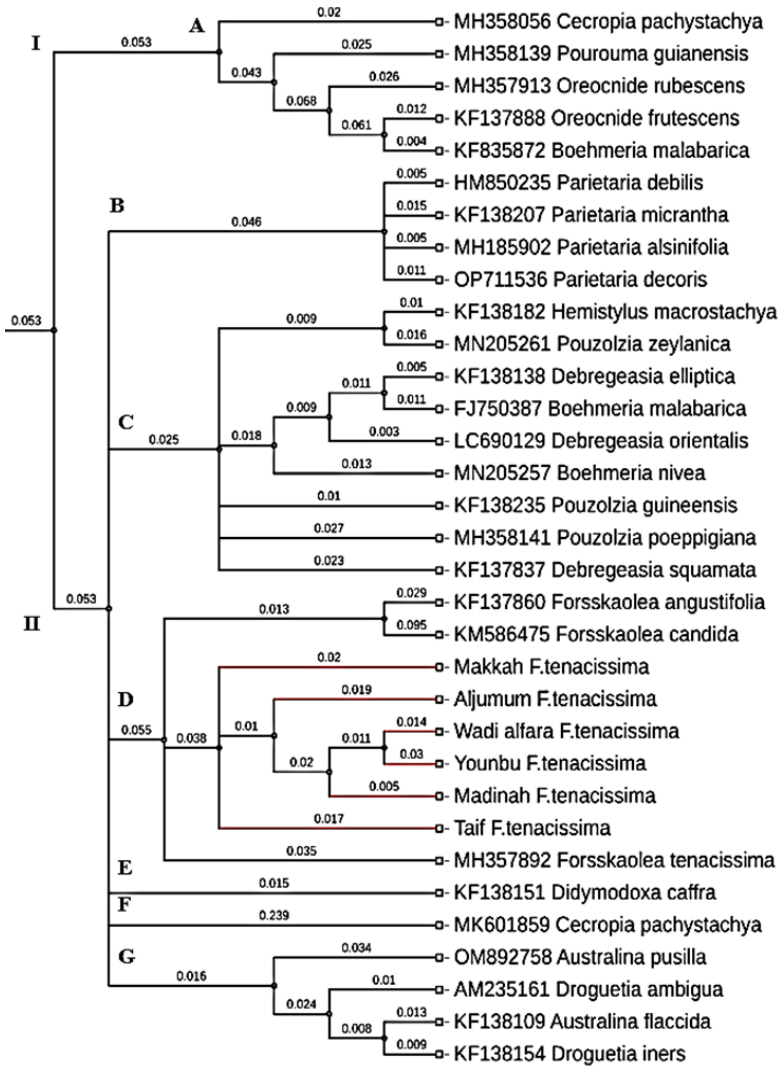

Figure 4. Bayesian all compatible tree inferred from ITS gene loci of *Forsskaolea* species. (I, II): clades, (A-G): sub-clades, (Red coloured branches): Saudi *Forsskaolea,* Numbers above branches is a posterior probability (PP).

The ITS and *rbc*L regions demonstrated excellent PCR amplification efficacy, high sequence quality, and robust species discrimination capacity. Hence, ITS and *rbc*L are deemed the optimal options for *F. tenacissima* barcoding. Consequently, Bayesian inference analyses significantly resolved both generic and intrageneric levels of phylogenetic resolution for F. tenacissima. Our results suggest that DNA barcoding is a practical method for resolving phylogenetic relationships at generic and species levels.

CONCLUSIONS

The current study accurately identified *F. tenacissima* and confirmed its descriptive taxonomy at the species and genus levels within the family Urticaceae. DNA barcoding efficacy for species identification enhances nucleotide information availability in DNA databases. This study utilized *rbc*L and ITS markers to distinguish and validate the identity of a popular medicinal plant used in traditional Arabian medicine, a significant step toward establishing DNA-based surveillance standards for medical counterfeiting in international and national commerce. Thus, the ITS and *rbc*L barcode loci employed in this study provided reliable results for species identification and the reconstruction of *F. tenacissima* phylogenetic relationships. The phylogenetic analysis of the individual datasets of the DNA sequences of *rbc*L and ITS succeeded to address the monophyly of the species under investigation, while the combined phylogenetic tree of both *rbc*L and ITS DNA sequences could not address the monophyly of *Forsskaolea* species in Saudi Arabia. Moropver, the phylogenetic

analysis on individual and combined datasets of *rbc*L and ITS DNA sequences successfully positioned *F. tenacissima* within the monophyletic tribe Forsskaolea. This discovery marks the initial inclusion of such information in the Flora of Saudi Arabia.

REFERENCES

Akaike, H. (1974): A new look at the statistical model identification. *IEEE transactions on automatic control*, 19(6):716-723.

Ali, M.A., Al-Hemaid, F.M., Lee, J., Hatamleh, A.A., Gyulai, G. and Rahman, M.O. (2015): Unraveling systematic inventory of Echinops (Asteraceae) with special reference to nrDNA ITS sequence-based molecular typing of *Echinopsabuzinadianus*. *Genetics and Molecular Research*, 14(4):11752-11762. doi: 10.4238/2015.October.2.9.

Al-Juhani, W.S. (2019): Evaluation of the capacity of the DNA barcode ITS2 for identifying and discriminating dryland plants. *Genetic Molecular Research*, 18(1):32-46.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990): Basic local alignment search tool. *Journal of molecular biology*, 215(3):403-410.

Anvarkhah, S., Khajeh-Hosseini, M., Mohassel Hasan Rashed, M., Davari- Edalat-Panah, A. and Hashemi, H. (2013): Identification of three species of genus Allium using DNA barcoding. *Internacional Journal of Agriculture andCrop Sciences*, 5(0):1195-1203.

Assaf, H.K., Nafady, A.M. and Kamel, M.S. (2017): Investigation of the saponifiable and unsaponifiable matter compositions by GC/MS and the antioxidant hepatoprotective activities of aerial parts of *Forsskaolea tenacissima* Linn. *Pharmaceutical Journal of Innovative Drug Research and Development*, 2(1):22-32.

Assaf, H.K., Nafady, A.M., Allam, A.E., Hamed, A.N.E. and Kamel, M.S. (2020): Phytochemistry and biological activity of family Urticaceae: a review (1957- 2019). *Journal of Advanced Biomedical and Pharmaceutical Sciences*, (3):150-176.

Attia, H.G., Aleraky, M., Youns, M. and Abdou, R. (2021): Cytotoxicity of endophytes of *Calotropis procera*, *Solanum nigrum* and *Forsskaolea tenacissima*. *Indian Journal of Pharmaceutical Education and Research*, 55(3): 872-879.

Bansal, S., Thakur, S., Mangal, M., Mangal, A.K. and Gupta, R.K. (2018): DNA barcoding for specific and sensitive detection of *Cuminum cyminum* adulteration in *Bunium persicum*. *Phytomedicine*, 50:178-183. doi: 10.1016/j.phymed.2018.04.023

Bell, D., Long, D.G., Forrest, A.D., Hollingsworth, M.L., Blom, H.H. and Hollingsworth, P.M. (2012): DNA barcoding of *European Herbertus* (*Marchantiopsida,Herbertaceae*) and the discovery and description of a new species: DNA barcoding herbertus. *Molecular Ecology Resources*,12:36-47.

Berg, C. C. (1977): Urticales, their differentiation and systematic position. *Plant Systematics and Evolution. Supplementum* 1: 349–374.

Berg, C. C. (1989): Systematics and phylogeny of the Urticales. In 'Evolution, systematics, and fossil history of the Hamamelideae. Vol. 2: higher Hamamelideae'. Systematics Association special volume no. 40B. (Eds PR Crane, S Blackmore) pp. 193–220. (Clarendon Press: Oxford).

CBOL Plant Working Group (2009): DNA barcode for land plants are. *Proceedings of the National Academy of Sciences of the United States of America*, 106(31):12794-12797.

Chen, S., Yao, H., Han, J., Liu, C., Song, J., Shi, L. and Zhu, Y. (2010): Validation of the ITS region as a novel DNA barcode for identifying medicinal plant species. *PLoS One*, 5(1):1-8. doi: 10.1371/journal.pone.0008613

Christenhusz, M. J. & Byng, J. W. 2016. The number of known plants species in the world and its annual increase. Phytospecies, 231: 201-2017.

Conn, B. J., Hadiah, J. T. (2009): Nomenclature of tribes within the Urticaceae. Kew Bulletin

64(2): 349–352.

EL-Banhawy A. and Al-Juhani W. (2019): DNA barcoding and phylogeny of *Phlomis aurea* (Lamiaceae) endemic to Sinai Peninsula, Egypt Pak. J. Bot 51 (4), 1263-1271.

EL-Banhawy A.; Acedo C.; Qari, S. and Elkordy, A. (2020-a): Molecular identification and phylogenetic placement of *Rosa arabica* Crép. (Rosaceae), a critically endangered plant species. *Life*, 10:335-350.

EL-Banhawy A.; Uluer DA.; Fayed AA.; Mohamed M, Faried A. (2020-b): DNA Barcoding and Phylogenetic Placement of the Genus Euphorbia L. (Euphorbiaceae) in Egypt, Biology and Life Sciences Forum 4 (1), 58.

El-Sherif, N., & Ibrahim, M. (2020). Implications of rbcL and rpoC1 DNA barcoding in phylogenetic relationships of some Egyptian Medicago sativa L. cultivars. Egyptian Journal of Botany, 60(2), 451-460.

Egazy, A.K.H., El-D Emerdash, MA and Osni, H.A.H. (1998): Vegetation, species diversity and floristic relations along an altitudinal gradient in south- west Saudi Arabia. *Journal of Arid Environments*, 38(1):3-13.

Faried A.; El-BanhawyA.; and Elqahtani M. (2018): Taxonomic, DNA barcoding and phylogenetic reassessment of the Egyptian *Ephedra* L. (Ephedraceae). *Catrina*: The International Journal of Environmental Sciences 17 (1), 1-13.

Feau, N., Vialle, A., Allaire, M., Maier, W. and Hamelin, R.C. (2011): DNA barcoding in the rust genus *Chrysomyxa* and its implications for the phylogeny of the genus. *Mycologia*, 103:1250-1266.

Flora of Pakistan: http://www.efloras.org/florataxon.aspx?flora_id=5&taxon_id=112946 Retrieved 2024-03-01.

Fouad, A., Hafez, R., & Hosni, H. (2019). Authentication of three endemic species of the Caryophyllaceae from Sinai Peninsula using DNA barcoding. *Egyptian Journal of Botany*, *59*(2), 483-491.

Fouad, A. S., Hamed, A. B., Amer, W. M., & Hafez, R. M. (2022). Barcoding of Some Plant Species Using the rbcL Gene in the Mediterranean Oolitic Sand Dunes West of Alexandria, Egypt. *Egyptian Journal of Botany*, *62*(1), 159-168.

Friis, I. (1989): Urticaceae. In: Polhill, RM. (Ed.). *Flora of tropical East Africa*, Rotterdam, Balkema, 65 pp.

Friis, I. (1993): Urticaceae. In: Kubitzki, K. (ed.) *The Families and Genera of Vascular Plants II*. Springer Verlag, Berlin, pp. 612–629.

Gilman, D. C.; Peck, H. T.; Colby, F. M., eds. (1905): *Forskål, Peter* . New International Encyclopedia (1st ed.). New York: Dodd, Mead.

Gillman, L.N., Keeling, D.J., Gardner, R.C. and Wright, S.D. (2010): Faster evolution of highly conserved DNA in tropical plants. *Journal of EvolutionaryBiology*, 23:1327-1330.

Gonzalez, M.A., Baraloto, C., Engel, J., Mori, S.A., Petronelli, P., Bernard Riéra,B., Roger, A., Thébaud, C. and Chave, J. (2009): Identification of Amazonian trees with DNA barcodes. *PLoS one*, 4:e7483.

Haglund, L. (2022): Optimised PCR protocol for ten microsatellite primers (SSRs) in *Fragaria vesca* Facilitating future work analysing genetic diversity and developing efficient conservation strategies. Linköping University, Department of Physics, Chemistry and Biology. pp 13.

Hall, T. A. (1999): BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic acids symposium series*, 41(41):95-98.

Han, J., Pang, X., Liao, B., Yao, H., Song, J. and Chen, S. (2016): An authenticity survey of herbal medicines from markets in China using DNA barcoding. *Scientific Reports*, 6:1-9. doi: 10.1038/srep18723

Hollingsworth, M.L., Andra Clark, ALEX, Forrest, L.L., Richardson, J.,Pennington, R.T., Long, D. G., Cowan, R., Chase, M.W., Gaudeul, M. and Hollingsworth, P.M. (2009): Selecting barcoding loci for plants: evaluation of seven candidate loci with species-level sampling in three divergent groups of land plants. *Molecular Ecology Resources*, 9(2):439-457.

Hollingsworth, P.M. (2008): DNA barcoding plants in biodiversity hot spots: Progress and outstanding questions. *Heredity*, 101(1):1-2. doi: 10.1038/hdy.2008.16

Hosein, F.N., Austin, N., Maharaj, S., Johnson, W., Rostant, L., Ramdass, A.C. and Rampersad, S.N. (2017): Utility of DNA barcoding to identify rareendemic vascular plant species in Trinidad. *Ecology and Evolution*, 7:7311– 7333.

Jamdade, R., Mosa, K.A., El-Keblawy, A., Al Shaer, K., Al Harthi, E., Al Sallani,M., ... & Mahmoud, T. (2022): DNA barcodes for accurate identification of selected medicinal plants (Caryophyllales): Toward barcoding flowering plants of the United Arab Emirates. *Diversity*, 14(4):262.

Joly, S., Davies, T.J., Archambault, A., Bruneau, A., Derry, A., Kembel, S.W., Peres-Neto, P., Vamosi, J. and Wheeler, T.A. (2014): Ecology in the age of DNA barcoding: the resource, the promise and the challenges ahead. *Molecular Ecology Resources*, 14:221-232.

Kim, C., Deng, T., Chase, M., Zhang, D. G., Nie, Z. L., Sun, H. (2015): Generic phylogeny and character evolution in Urticeae (Urticaceae) inferred from nuclear and plastid DNA regions. Taxon 64(1): 65–78. http://dx.doi.org/10.12705/641.20

Kress, W.J. and Erickson, D.L. (2008): A Two-Locus Global DNA Barcode for Land Plants: The Coding rbcL Gene Complements the Non-Coding trnH-psbASpacer Region. *PLoS ONE*, 2(6): e508. doi: 10.1371/journal.pone.0000508

Kress, W.J., Erickson, D.L., Swenson, N.G., Thompson, J., Uriarte, M. andZimmerman. JK (2010): Advances in the use of DNA barcodes to build a community phylogeny for tropical trees in a Puerto Rican Forest Dynamics Plot. *PLoS ONE*, 5(11):e15409.

Kress, W.J., Wurdack, K.J., Zimmer, E.A., Weigt, L.A. and Janzen, D.H. (2005):Use of DNA barcodes to identify flowering plants. *Proceedings of the National Academy of Sciences of the United States of America*, 102(23):8369- 8374. doi: 10.1073/pnas.0503123102

Lahaye, R., Van der Bank, M., Bogarin, D., Warner, J., Pupulin, F., Gigot, G., Maurin, O., Duthoit, S., Barraclough, T.G. and Savolainen, V. (2008)DNA barcoding the floras of biodiversity hotspots. *Proceedings of the National Academy of Sciences of the United States of America*, 105:2923- 2928.

Luo, K., Chen, S., Chen, K., Song, J.Y., Yao, H., Ma, X.Y., Zhu, Y.J., Pang, X.H., Yu, H., Li, HW and Liu, Z. (2010): Assessment of candidate plant DNA barcodes using the Rutaceae family. *Science China*, 53(6):701-708.

Madden, T. (2002): The BLAST sequence analysis tool. 1. GenBank: The nucleotide sequence database. 1-16.

Maloukh, L., Kumarappan, A., Jarrar, M., Salehi, J., El-wakil, H. and Rajya Lakshmi, T.V. (2017): Discriminatory power of rbcL barcode locus for authentication of some of United Arab Emirates native plants. *3Biotech*, 7(2):1-7. doi:10.1007/s13205-017-0746-1

Meier, R., Shiyang, K., Vaidya, G., and Ng, PK (2006): DNA barcoding and taxonomy in Diptera: a tale of high intraspecific variability and low identification success. *Systematic biology*, 55(5):715-728.

Newmaster, S.G., Fazekas, A.J. and Ragupathy, S. (2006): DNA barcoding in land plants: Evaluation of *rbcL* in a multigene tiered approach. *Canadian Journal of Botany*, 84(3):335-341. doi:10.1139/B06-047

Nylander, J.A. (2004): Bayesian phylogenetics and the evolution of gall wasps.Ph.D. Thesis, Faculty of Science and Technology, Acta UniversitatisUpsaliensis, pp. 43.

Parmentier, I., Duminil, J., Kuzmina, M., Philippe, M., Thomas, D.W., Kenfack, D., Chuyong, G.B., Cruaud, C. and Hardy, O.J. (2013): How effective are DNA barcodes in the identification of African rainforest trees? *PLoS One*, 8: e54921.

Pathak, M.R., Mohamed, A.A.M. and Farooq, M. (2018): DNA Barcoding and Identification of Medicinal Plants in the Kingdom of Bahrain. *American Journal of Plant Sciences*, 09(13):2757-2774.

Pei, N., Chen, B. and Kress, W.J. (2017): Advances of community-level plant DNA barcoding in China. *Frontiers in plant science*, 8:225.

Plants of the World Online. Royal Botanic Gardens, Kew. 2024: https://powo.science.kew.org/taxon/urn:lsid:ipni.org:names:40809-1.Retrieved 1 March 2024.

Qari, S.H., Alrefaei, A.F., Filfilan, W. and Qumsani, A. (2021): Exploration of the medicinal flora of the Aljumum region in Saudi Arabia. *Applied Sciences (Switzerland)*, 11(16):1-17.

Ronquist, F., Teslenko, M., Van Der Mark, P., Ayres, D. L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M.A. and Huelsenbeck, JP (2012): MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic biology*, 61(3):539-542.

Sanger, F., Nicklen, S. and Coulson, A.R. (1977): DNA sequencing with chain- terminating inhibitors. *Proceedings of the national academy of sciences*, 74(12):5463-5467.

Shah, S.W.A., Kamil, S., Ahmad, W. and Ali, N. (2010): Spasmogenic, spasmolytic and antihypertensive activity of *Forsskalea tenacissima* L. *African Journal of Pharmacy and Pharmacology*, 4(6):381-385.

Shamso, E., & Fouad, A. (2019). Dicliptera aegyptiaca (Acanthaceae), A New Species from Egypt Supported by Morphological Characters and rbcl-based DNA Barcoding. *Egyptian Journal of Botany*, *59*(2), 475-482.

Sher, A.A., Afzal, M. and Bakht, J. (2017): Pharmacological evaluation of different extracts of *Forsskaolea tenacissima*. *Indian Journal of Pharmaceutical Sciences*, 79(2):257-266.

Starr, J.R., Naczi, R.F. and Chouinard, B.N. (2009): Plant DNA barcodes and species resolution in sedges (*Carex*, Cyperaceae). *Molecular Ecology Resources*, 9:151-163.

Stöver, B.C. and Müller, K.F. (2010): TreeGraph 2: combining and visualizing evidence from different phylogenetic analyses. *BMC bioinformatics*, 11:1-9.

Swofford, D.L. and Sullivan, J. (2003): Phylogeny inference based on parsimony and other methods using PAUP. *The phylogenetic handbook: a practical approach to DNA and protein phylogeny cáp*, 7:160-206.

Tamura, K., Stecher, G., Peterson, D., Filipski, A. and Kumar, S. (2013): MEGA6 : Molecular evolutionary genetics analysis version 6.0. *Molecular Biology and Evolution*, 30(12):2725-2729. doi: 10.1093/molbev/mst197

Thomas, C. (2009): Plant barcode soon to become reality. *Science*, 325:526.

Weddell, H. A. (1854): Revue de la famille de Urticacees. *Annales des Sciences Naturelles: Botanique Series* 4(1): 173–212.

Weddell, H. A. (1856): *Monographie de la famille des Urticées*. Archives du Museum National d'Histoire Naturelle Paris 9, G. & J. Baudry, Paris, 591 pp.

Weddell, H. A. (1869): Urticaceae. In: Candolle, A.P. de (Ed.) *Prodromus Systematis Naturalis Regni Vegetabilis* 16, Paris, pp. 32–235. http://dx.doi.org/10.5962/bhl.title.286

White, T.J., Bruns, T., Lee, S. and Taylor, J. (1990): Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. PCR protocols. Academic Press, Inc. 1:315-322. doi:10.1016/b978-0-12-372180-8.50042-1

Wu, Z-Y., Monro, A. K., Milne, R. I., Wang, H., Yi, T-S., Liu, J., Li, D.-Z., (2013): Molecular phylogeny of the nettle family (Urticaceae) inferred from multiple loci of three genomes and extensive generic sampling. *Mol. Phylogenet. Evol.* 69: 814–827. http://dx.doi.org/10.1016/j.ympev.2013.06.022.

Xiao, L.Q., Moller, M. and Zhu, H. (2010): High nrDNA ITS polymorphism in the ancient extant seed plant Cycas: incomplete concerted evolution and the origin of pseudogenes. *Molecular Phylogenetics and Evolution*, 55:168-177.

Yesson, C., Bárcenas, R.T., Hernández, H.M., De La Luz Ruiz-Maqueda, M., Prado, A., Rodríguez, V.M. and Hawkins, J.A. (2011): DNA barcodes for *Mexican cactaceae*, plants under pressure from wild collecting: DNA barcodes for *Mexican cactaceae*. *Molecular Ecology Resources*, 11:775-783.