DNA Barcoding for Some Medicinal Plants in the El-Omayed Biosphere Reserve

Reham N Abdel Azem[1*], Mohammed A Rashed[2], Mahmoud Magdy[2], Esraa A Elsherbeny[1]

1- Genetics Resources Dept, Desert Research Center, Ministry of Agriculture, Cairo, Egypt
2- Genetic Dept, Fac of Agric, Ain Shams Univ, P.O. Box 68, Hadayek Shoubra 11241, Cairo, Egypt

*Corresponding author: rehamabdelazem28@gmail.com

Abstract: DNA barcoding is a valuable tool for molecular identification of plant materials, which is primarily employed to guarantee the nature and therapeutic value of all available medicinal plants. Identification is crucial for the illegal medicinal plant trade, safe use, and preventing adulteration. In this study, DNA barcoding was used to evaluate and confirm the identification of some medicinal plants from the El-Omayed Biosphere Reserve (OBR). Two DNA regions, rbcL and ITS were selected due to their universality, ease of amplification and sequencing, and ability to identify taxa at the generic and species levels. The ITS region demonstrated a higher capacity for species discrimination power than the rbcL region. It assisted in identifying four of eight plants more precisely than rbcL. Together, ITS and rbcL could be used for plant species identification, conservation, and trade control of valuable plant resources.

1 Introduction

Recently, the loss of biodiversity has become a growing concern. The global degradation of natural habitats, ecosystems, and land is a result of climate change, the invasion of alien species, and human-related practices (construction of irrigation canals, roads, and tourist resorts). The North-West Coastal Zone of Egypt and the Omayed Biosphere Reserve (OBR), in particular, are excellent examples of anthropogenic and climate-driven land erosion (El-Raey 2010). In the last 50 years, droughts, pasture conversion to croplands, population growth, paved roads, and agroecological diversification. The Omayed Biosphere Reserve is the only protected area reflecting the ecology of the Mediterranean. It is a significant site for the biodiversity of Egyptian flora in the Mediterrane-

an coastal region, area, comprising about 12% of Egypt's flowering plants (Abdel-Kader 2019, El-Sakaty et al 2022).

Medicinal plants play an essential role in 80% of the world's primary health care (Farnsworth et al 1985, El Beyrouthy and Abi Rizk 2013). Most medicinal plant parts contain active chemicals with direct or indirect therapeutic effects and are used as medicines. Active compounds are substances produced and stored within the bodies of plants (substances) (Phillipson 2001). Egypt contains several diverse medicinal plants, providing a valuable source of plants rich in anticancer, antimicrobial, and antioxidant agents. As both cytotoxic compounds and antioxidant agents. As *Urtica urens* L., *Rosmarinus officinalis*, *Achillea fragrantissima*, *Adonis dentata* Del., *Rumex vesicarius* L. DNA barcode refers to the use of a DNA sequence-based identification protocol that may consist of

a single locus or multiple loci used together as complementary units (Kress and Erickson 2007). DNA barcoding aids in systematics and the identification of species bounder, as well as cryptic species and the adulteration of medicinal plants with non-medicinal substitutes of a closely related species with lower efficiency (Ragupathy 2009). Furthermore, the essential characteristics of DNA barcoding are its universality, specificity on variation, ease of use, and applicability to a diverse array of taxa. There is a substantial divergence between species, but it is conserved within the species so that intra-specific variation is insignificant (Kress et al 2005, Pennisi 2007).

Molecular characterization was performed using two different DNA markers widely used in a DNA barcoding context; the nuclear internal transcribed spacer (ITS) and the chloroplast Rubisco large subunit (rbcL) gene have the ability to discriminate at the generic and species levels and were used for molecular characterization. The rbcL gene has a simple and stable genetic structure, is haploid, seldom undergoes recombination, and is typically transmitted uniparentally. Moreover, universal primers can amplify specific target sequences (Tsubota et al 2001). The ITS is a highly polymorphic region of noncoding DNA spacer between structural ribosomal RNAs genes. The spacer is divided into ITS1 and ITS2, separated by the 5.8S ribosomal gene (Wheeler and Honeycutt 1988).

This study characterizes eight plant species using rbcL and ITS DNA regions to confirm their identification as economic and/or medicinal plants; the collected specimens are part of the *ex-situ* conservation plan for the human-threatened flora of the Omayed Biosphere Reserve (OBR).

2 Materials and Methods

2.1 Sample collection

In this study, eight unidentified medicinal plants were collected from human-threatened regions of the OBR, the western Mediterranean coastal region of Egypt, and Marsa Matrouh, Egypt. The OBR is situated 80 km west of Alexandria in the western Mediterranean coastal region of Egypt (29°00′–29°18′E and 30°52′–20°38′N). It extends ~30 km along the Mediterranean coast from west El-Hammam to El-Alamein and encompasses approximately 700 km^2 of land (**Fig 1**).

2.2 DNA extraction

The use of 0.25 g ground leaves from eight samples. Total DNA was extracted with a plant DNA extraction kit (Trans, China). Under UV light, the DNA quality was evaluated using 1% Agarose gel electrophoresis, visualized by pre-added Red Safe® (5ul /100 ml) under UV light.

2.3 PCR and sequencing

The ITS region was amplified using the primer pair ITS1 (5'-TCC GTA GGT GAA CCT GCG G-3') and ITS4 (5'-TCC TCC GCT TAT TGA TAT GC-3') (White et al 1990). The rbcL gene was amplified using the primer pairs rbcLaF (5'-ATG TCACCA CAA ACA GAG ACT AAA GC-3') and rbcLaR (5'-GTA AAA TCA AGT CCA CCR CG-3').

PCR of 25 µl reaction mixture (1 µl Template, 1 µl Forward primer, 1 µl Reverse primer, 2.5 µl EasyTaq® buffer (10x), 2 µl dNTPs (2.5 mM each), 0.2 µ l Easy Taq® DNA polymerase, 17.3 µl ddH$_2$O)were performed. A standard PCR profile with 55ºC annealing temperature was used to amplify ITS and 50ºC to amplify rbcL. The result of the amplified products was tested on 1.5% Agarose gel electrophoresis.

When successful, the amplified bands were purified using a PCR product purification kit (GeneDir X). forward and reverse sequences were aligned using BIOEDIT V3 (Hall 1999) to confirm the correctness of the sequences, then blasted on GenBank using default settings, and all closely related species were downloaded for further analysis. The rbcL and ITS sequences were independently aligned, and a maximum-likelihood tree was constructed using MEGA V6 with 1000 bootstrap replicates (Tamura et al 2013).

3 Results and Discussion

The eight species collected from the North-Western Coastal Zone have been identified based on their morphology as follows: 01 *Zygophyllum album*, 02 *Zygophyllum coccineum*,03 *Rumex vesicarius*, 04 *Emex Spinosa*,05 *Rumex pictus*, 06 *Urtica urens*, 07 *Gymnocarpos dentrans*, and 08 *Helianthemum Lippi* as shown in **Figs 2–9**, **Table 1**.
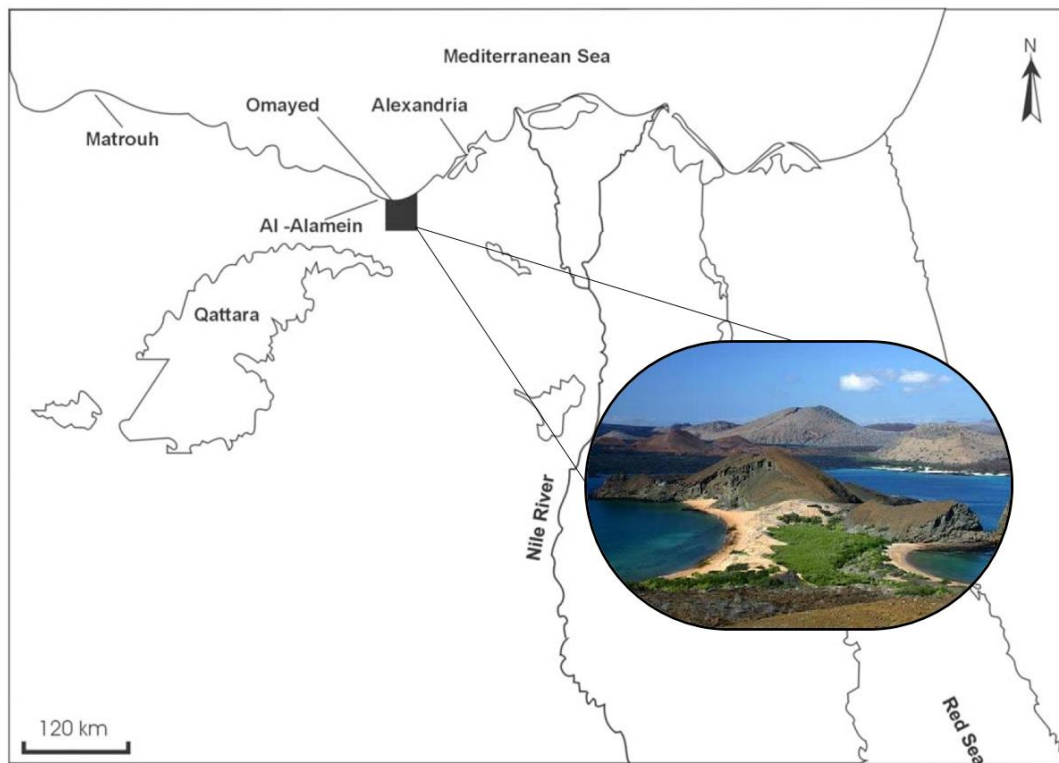
Fig 1. The El-Omayed Biosphere Reserve location

Fig 2. sample of *Zygophyllum album*

Fig 3. Sample of *Zygophyllum coccineum* L

Fig 4. Sample of *Rumex pictus forssk*

Fig 5. Sample of *Rumex vesicarius L*

Fig 6. Sample *of Emex spinosa*

Fig 7. Sample *Urtica urens L.*

Fig 8. Sample of *Helianthemum lippii*

Fig 9. Sample of *Gymnocarpos decandrus Forrsk*

3.1 Molecular identification and DNA barcoding

The rbcL region's BLAST results distinguished all samples to the species level, except species number 6. In greater detail, species (01) matched *Tetraena coccinea* (MH598889) with a pairwise identity of 100%, species (02) matched *Tetraena coccinea* (MH598889) with a pairwise identity of 99.64%, species (03) matched *Rumex nepalensis* (KX015758) with a pairwise identity of 99.63%, species (04) matched *Rumex nepalensis* (KX015758) with a pairwise identity of 99.28%, species (05) matched *Rumex spinosus* (KX282712) with a pairwise identity of 99.81%, species (06) matched *Urtica sp.* (KF138265) with a pairwise identity of 99.81%, species (07) matched *Helian-*

themum lippii (KX282785) with a 100% pairwise identity. Comparatively, species (08) matched *Gymnocarpos sclerocephalus* (KX283025) with a pairwise identity of 99.80%.

The ITS BLAST results distinguished each sample to the species level. In particular, species (01) matched *Tetraena fontanesii* (HE602441) with a pairwise identity of 98.80%, species (02) matched *Tetraena fontanesii* (HE602441) with a pairwise identity of 100%, species (03) matched *Rumex vesicarius* (KJ004288) with a pairwise identity of 98.52%, species (04) matched *Rumex vesicarius* (KJ004288) with a pairwise identity of 99.64%, species (05) matched *Rumex spinosus* (MW433670) with a pairwise identity of 100%, species (06) matched *Embryophyte environmental* (KM515510) with a pairwise identity of 99.85%, species (07) matched *Helianthemum stipulatum* (MN638885) with a pairwise identity of 99.85%, species (08) matched *Gymnocarpos decandrus* (KF815496) with a pairwise identity of 99.56% (**Table 2**).

The top 5 best-matched sequences from the NCBI database were retrieved on the basis of the BLAST results in **Table (2)** and aligned with our results for all samples. The diversity indices of each sampled species per genetic region were calculated independently to identify the best marker reflecting the greatest diversity among all.

In the case of *rbcL*, the full alignment length was 522 because the longest total length for species 03 and 04 was 533 bp as opposed to 516 bp for species 06. The total GC % ratio was 44%, with the highest GC % ratio found in species 04 (44.70), as opposed to species 02 (40.90). There was a total of 3,375 segregation sites. The highest number of segregating sites was recorded for species 06 (8 sites), while the lowest number was recorded for species 01 (no sites). The alignment of species 08 recorded the greatest number of haplotypes (Hap) with four, whereas the alignment of species 01 only recorded a single haplotype. In contrast to species 01, species 04 and 05 possessed the highest haplotype diversity (Hd) of 0.73 (no sites). Samples 06 and 08 had the highest nucleotide diversity per site (Pi) of 0.005 compared to species 01 and 07, which had no Pi value (**Table 3).**

In the ITS case, the total length was 622 bp, and the longest sequence was 720 bp (species 01), as opposed to sampling 06's length of 559 bp. The total GC % ratio was 61%, the highest GC percentage was recorded for species 03 at 68%, and

the lowest percentage was recorded for species 06 at 53.90% (**Table 4**).

The number of segregation site (s) in species 01 and 02 was 88, whereas, in species 06, there was only 6. (11). for species 01 and 02, the number of haplotypes (Hap) was 11, whereas, for species 08, the number of Hap was the lowest (5). Species 01 and 02 had the greatest haplotype diversity (Hd) at 1.00, while species 05 was estimated to have the least (0.81). Species 01 had the highest nucleotide diversity per site (Pi) at 0.065, while sample 6 had the lowest at (0.005). Higher divergence of the ITS marker has been widely reported, and its high potential as a barcoding region for systematic and evolution research in plants has made ITS a widely cited barcoding region (Fazeli-Nasab et al 2020).

3.4 Comparative phylogenetic analysis

The *rbcL* phylogenetic analysis revealed that species 01 and 02 belonged to the same cluster as Zygophyllum and *Tetraena* species, although *Tetraena coccineum* was more closely related to them. Species 03 and 04 were grouped with four different *Rumex* species and specifically belonged to *Rumex hastatus*, which was the cluster's most closely related species. Alternatively, species 05 and 07 were determined to be *Rumex Spinosa* and *Helianthemum Lippi*, respectively. Species 06 belonged to the Urtica genus, while species 08 belonged to the *Gmnocarpos sclerocephalus* species of the *Gmnocarpos* genus (**Fig 10**).

ITS phylogenetic analysis revealed that species 01 and 02 formed a cluster with single species, *Tetraena fontanesii,* with a bootstrap support of 98%. Species 03 and 04 were clustered with a single species, *Rumex vesicarius,* with a bootstrap value of 96%. With a bootstrap value of 91%, species 05 was grouped with three other *Rumex spinosa* species. Species 06 was clustered with *Embryophyte environmental* species. Species 07 was clustered with two other species that were *Helianthemum stipulatum*, with a bootstrap value of 85%. Species 08 was grouped with three same species, *Gmnocarpos decandrum* with a bootstrap value of 99% (**Fig 11**).

By comparing the ITS and rbcl phylogenetic analyses, we discovered that the ITS tree was more effective in identifying samples than the *rbcl* sequence. This result is consistent with those of previous researchers (Chase et al 2005, El-Atroush et al 2015, Kress et al 2005,), and rbcL was not suitable for DNA barcoding for this species, although ITS was more efficient in differentiating between genera and some species, it cannot be used as single DNA barcoding due to the variation within species.

Table 1. Common medicinal uses of the species of interest in this study

	Scientific name	Family	Common name
1	*Helianthemum lippii* (**Forssk.**)	Cistaceae	It is used in the traditional treatment of cancer because it contains the cyclanite group and contains antioxidants
2	*Zygophyllum album L.*	Zygophyllaceae	Anthelmintic Gout Blood and heart tonic Ophthalmic
3	*Zygophyllum coccineum L*		
4	*Urtica urens L.*	Urticaceae	annual, dwarf, and burning nettle
5	*Rumex pictus forssk*	Polygonaceae	used for stomach disorders and to relieve colic, stomach burn, toothache,
6	*Rumex vesicarius L.*		
7	*Emex spionsa L*		
8	*Gymnocarpos decandrus Forsk.*	Caryophyllaceae	antidiabetic potential, anti-inflammatory, analgesic and diuretic activities

Table 2. Blast results for the rbcL and ITS genes for eight samples. Including sample code, the accession number, organism name and percentage of pairwise

Sample	rbcL				ITS	
	Accession no.	Species	PI%	Accession no.	Species	PI%
S01	MH598889	*Tetraena coccinea*	100.00%	HE602441	*Tetraena fontanesii*	98.80%
S02	MH598889	*Tetraena coccinea*	99.64%	HE602441	*Tetraena fontanesii*	100%
S03	KX015758	*Rumex nepalensis*	99.63%	KJ004288	*Rumex vesicarius*	98.52%
S04	KX015758	*Rumex nepalensis*	99.28%	KJ004288	*Rumex vesicarius*	99.64%
S05	KX282712	*Rumex spinosus*	99.81%	MW433670	*Rumex spinosus*	100.00%
S06	KF138265	*Urtica sp.*	99.81%	KM515510	*Embryophyte environmental*	99.85%
S07	KX282785	*Helianthemum lippii*	100.00%	MN638885	*Helianthemum stipulatum*	99.85%
S08	KX283025	*Gymnocarpos sclerocephalus*	99.80%	KF815496	*Gymnocarpos decandrus*	99.56%

Table 3. Statistics of the alignment of the rbcL gene, total sites, percentage of the GC content, number of the segregation sites (s), haplotype diversity (Hd), number of haplotypes (Hap), nucleotide diversity per site (Pi), GC content and nucleotide diversity (π)

Sample	Sites(bp)	GC%	S	Hap	Hd	Pi	π
S01	526	41.10%	0	1	0.00	0.000	n.a.
S02	522	40.90%	3	2	0.55	0.003	1.00
S03	533	44.60%	2	3	0.67	0.001	0.08
S04	533	44.70%	2	3	0.73	0.002	0.18
S05	518	44.50%	5	3	0.73	0.004	0.53
S06	516	43.60%	8	3	0.47	0.005	0.76
S07	525	44.20%	1	2	0.55	0.001	n.a.
S08	503	43.60%	6	4	0.67	0.005	0.40
Full Alignment	522	44.10%	3.375	2.6	0.55	0.003	0.49

Table 4. Statistics of the alignment of the ITS region, total sites, percentage of the GC content, number of the segregation sites (s), haplotype diversity (Hd), number of haplotypes (Hap), nucleotide diversity per site (Pi), GC content and nucleotide diversity (π)

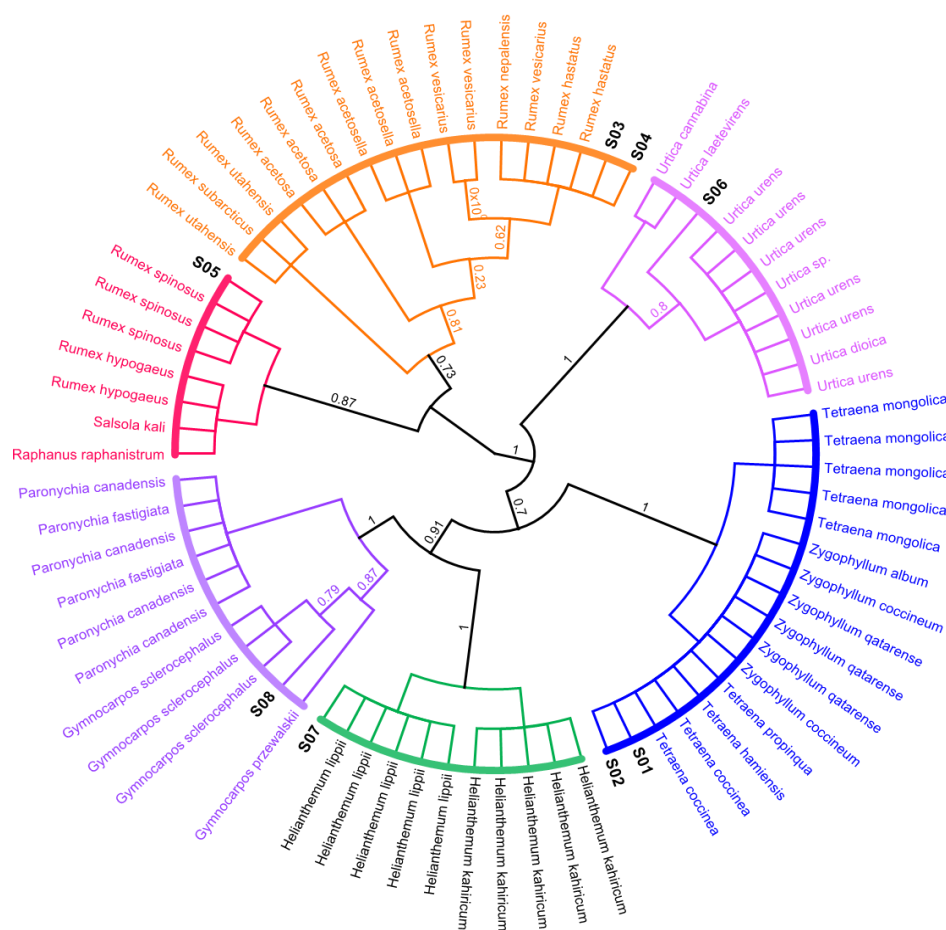| Samples	Sites(bp)	GC%	S	Hap	Hd	Pi	π
S01	720	60.50%	88	11	1	0.065	0.18323
S02	623	61.20%	88	11	1	0.065	0.18323
S03	583	68.00%	53	8	0.95	0.043	0.29978
S04	579	67.90%	56	8	0.95	0.045	0.3039
S05	559	67.20%	59	7	0.81	0.05	0.48888
S06	642	53.90%	11	8	0.92	0.005	0.15143
S07	673	63.20%	23	10	0.98	0.009	0.16014
S08	601	56.10%	29	5	0.85	0.019	0.50132
Full Alignment	622	61.70%	50.8	8.5	0.93	0.038	0.28399

Fig 10. Phylogenetic tree between different species of El-Omayed Biosphere Reserve, using the rbcL sequence-based maximum-likelihood method
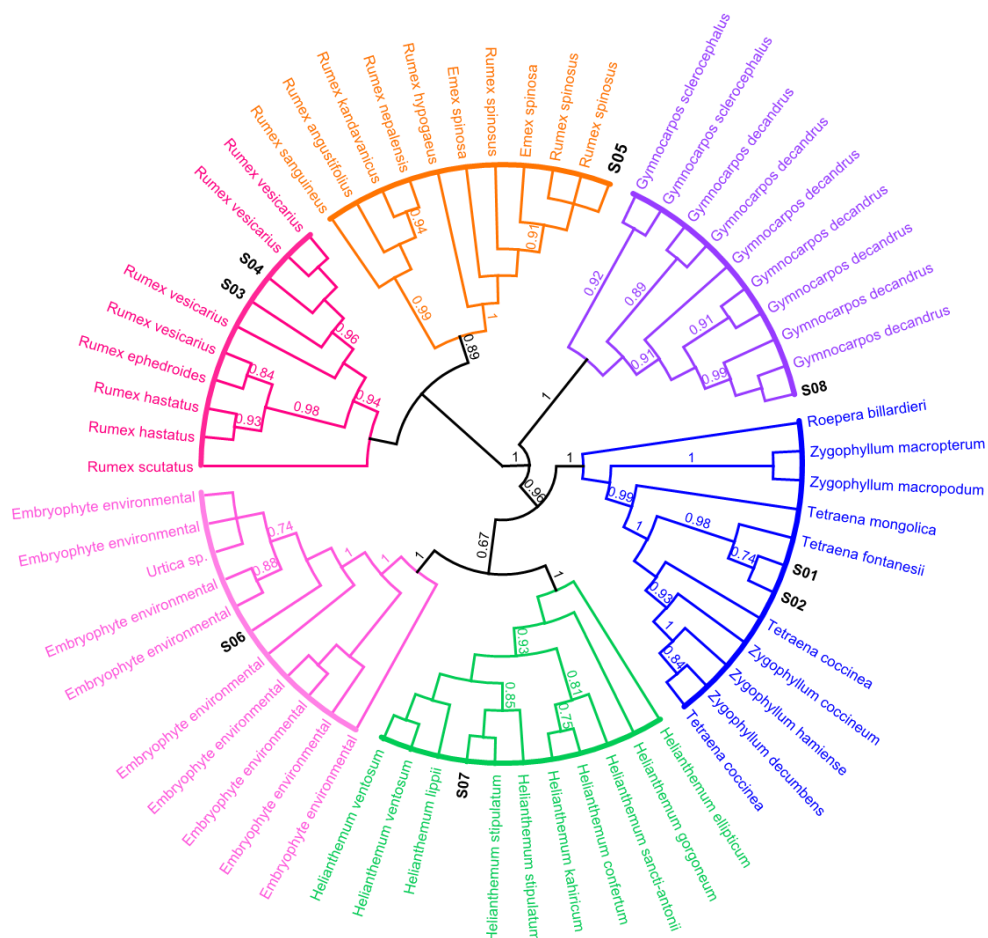
Fig 11. Phylogenetic tree between different species of El-Omayed Biosphere Reserve, shows a match between the present study and sequence in the GenBank database, using the maximum-likelihood method based on the ITS sequences

Table 5. Comparative species identification table, including morphological inspection and the phylogenetic results of rbcL and ITS trees

Species	Morphology	rbcL	ITS
S01	*Zygophyllum album*	*Tetraena coccinea**	*Tetraena fontanesii*
S02	*Zygophyllum coccineum*	*Tetraena coccinea**	*Tetraena fontanesii*
S03	*Rumex vesicarius*	*Rumex hastatus*	*Rumex vesicarius*
S04	*Rumex pictus*	*Rumex hastatus**	*Rumex vesicarius*
S05	*Rumex spinosa*	*Rumex spinosa**	*Rumex spinosa*
S06	*Urtica urens*	*Urtica sp.**	*Embryophyte environmental*
S07	*Helianthemum Lippi*	*Helianthemum Lippi* X hyb	*Helianthemum stipulatum*
S08	*Gmnocarpos decandrum*	*Gmnocarpos sclerocephalus* X hyb	*Gmnocarpos decandrum*

* Failed

The first instance was observed for species 01 and 02, where the species *Tetraena fontanesii* was absent from the rbcL tree, and the cluster consisted of five undifferentiated species of the genus Tetraena. In contrast, in the ITS tree, the division between samples related to multiple accessions of the same species *Tetraena fontanesii* was 98% bootstrap supported. Consequently, the definition of the two species 01 and 02 would be *Tetraena fontanesii*, given that both *Tetraena* and *Zygophyllum* were used in multiple references for the same genus (Alzahrani and Albokhari 2018). In the second instance, involving species 03 and 04, where no cluster division was observed between the samples and their sisters from other species in the rbcL tree, the cluster consisted of four undifferentiated species. In contrast, species 03 and 04 were grouped with the species *Rumex vesicarius* in the ITS tree with bootstrap support of 96%, accordingly, to this result, it is defined as *Rumex vesicarius*. While classified as species 05, its morphology, rbcL, and ITS sequences were identical to those of *Rumex spinosa.* Species 06, the ITS tree was better in its definition as it appeared in the same cluster as its sister species in the ITS tree, which corresponded to the *Embryophyte environment,* whereas in the rbcL tree, it appeared alone in a cluster with different sister species. According to the results of both rbcL and ITS sequences, species 07 and 08 are hybrids consisting of *Helianthemum lippi* x *Gmnocarpos sclerocephalus* and *Helianthemum lippi* × *Gmnocarpos decandrum*, respectively **(Table 5).**

4 Conclusion

In conclusion, the current preliminary assessment data would be beneficial for DNA barcoding of endangered medicinal plants in the OBR region of Egypt's northern coast. The ITS was more effective at identifying the eight plant species than the rbcL sequences because it had a higher resolution for species identification. Whether *in situ* or *ex-situ*, serious measures should be taken to preserve the endangered plant species in the OBR area. Additionally, a local database of authenticated DNA barcodes should be created so that plant DNA barcoding techniques can be used and applied effectively.

Reference

Abdel-Kader FH (2019) Assessment and monitoring of land degradation in the northwest coast region, Egypt using earth observations data. *The Egyptian Journal of Remote Sensing and Space Science* 22, 165-173. https://doi.org/10.1016/j.ejrs.2018.02.001

Alzahrani DA, Albokhari EJ (2018) Taxonomic revision of Saudi Arabian *Tetraena* Maxim and *Zygophyllum* L. (Zygophyllaceae) with one new variety and four new combinations. *Bangladesh Journal of Plant Taxonomy* 25, 19-43. https://doi.org/10.3329/bjpt.v25i1.37177

Chase MW, Salamin N, Wilkinson M, et al (2005) Land plants and DNA barcodes: short-term and long-term goals. *Philosophical Transactions of the Royal Society B: Biological Sciences* 360, 1889-1895. https://doi.org/10.1098/rstb.2005.1720

El Atroush H, Magdy M, Werner O (2015) DNA barcoding of two endangered medicinal plants from Abou Galoom protectorate. *Life Science Journal* 12, 101-109. https://api.semanticscholar.org/CorpusID:92985340

El Beyrouthy M, Abi-Rizk A (2013) DNA fingerprinting: The new trend in fighting the adulteration of commercialized and cultivated medicinal plants. *Advances in Crop Science and Technology* 1, e107. https://rb.gy/68oeur

El-Raey M (2010) Impacts and implications of climate change for the coastal zones of Egypt. In: Michel D, Pandya A (Eds.), Coastal Zones and Climate Change, pp 31-50. https://www.jstor.org/stable/resrep10902.9

El-Sakaty S, Magdy M, Rizk S, et al (2022) Ex-situ Conservation of the Micro and Macro Flora of Omayed Biosphere Reserve (OBR): A Survey Report. *Egyptian Journal of Botany* 62, 149-158. https://doi.org/10.21608/ejbo.2021.46939.1573

Farnsworth NR, Akerele O, Bingel A, et al (1985) Medicinal plants in therapy. *Bulletin of the World Health Organization* 63, 965-981. PMCID: PMC2536466

Fazeli-Nasab B, Sayyed RZ, Farsi M, et al (2020) Genetic assessment of the internal transcribed spacer region (ITS1. 2) in *Mangifera indica* L. landraces. *Physiology and Molecular Biology of Plants* 26, 107-117. https://doi.org/10.1007/s12298-019-00732-x

Hall TA (1999) BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series* 41, 95-98.

Kress WJ, Erickson DL (2007) A two-locus global DNA barcode for land plants: the coding *rbcL* gene complements the non-coding trnH-psbA spacer region. *PLoS One* 2, e508. https://doi.org/10.1371/journal.pone.0000508

Kress WJ, Wurdack KJ, Zimmer EA, et al (2005) Use of DNA barcodes to identify flowering plants. *Proceedings of the National Academy of Sciences* 102, 8369-8374. https://doi.org/10.1073/pnas.0503123102

Pennisi E (2007) Taxonomy Wanted, A barcode for plants. *Science* 318, 190–191. https://doi.org/10.1126/science.318.5848.190

Phillipson JD (2001) Photochemistry and medicinal plants. *Phytochemistry* 56, 237-243. https://doi.org/10.1016/S0031-9422(00)00456-8

Ragupathy S, Newmaster SG, Murugesan M, et al (2009) DNA barcoding discriminates a new cryptic grasss pecies revealed in an ethnobotany study by the hill tribes of the Western Ghats in southern India. *Molecular Ecology Resources* 9, 164-171. https://doi.org/10.1111/j.1755-0998.2009.02641.x

Tamura K, Stecher G, Peterson D, et al (2013) MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. *Molecular Biology and Evolution* 30, 2725-2729. https://doi.org/10.1093/molbev/mst197

Tsubota H, Akiyama H, Yamaguchi T, et al (2001). Molecular phylogeny of the Sematophyllaceae (Hypnales, Musci) based on chloroplast rbcL sequences. *The Journal of the Hattori Botanical Laboratory* 90, 221-240. https://doi.org/10.18968/jhbl.90.0_221

Wheeler WC, Honeycutt RL (1988) Paired sequence difference in ribosomal RNAs: Evolutionary and phylogenetic implications. *Molecular Biology and Evolution* 5, 90–96. https://doi.org/10.1093/oxfordjournals.molbev.a040480