DNA Barcoding of Five Medicinal Plants from Siwa Oasis, Egypt

Abeer Mohamed

Agricultural Botany Department, Faculty of Agriculture (Saba Basha), Alexandria University, Egypt

Abstract

Siwa Oasis is one of the depressions located in the northern part of the western desert of Egypt covering almost 1000 Km². The oasis land provides habitat for a diverse flora which became endangered after human activity. Collection and surveying of the plant species that exist in the Oasis and the surrounding areas by morphological and molecular means are important approaches. This study utilized DNA barcoding as a rapid, reproducible and economic molecular tool for the discrimination and identification of five medicinal plants in Siwa Oasis. I used rbcL+matK as a core barcode sites. Species identified during the course of this study are *Nitraria retusa*, *Cynara sibthorpiana*, *Capparis spinosa*, *Peganum harmala*, *Pergularia tomentosa*. This study is a first step toward the barcoding of all flora in Siwa Oasis and should help to increase the awareness of plant biodiversity, not only in Siwa Oasis, but also in many other localities of biodiversity in Egypt.

Keywords: DNA barcoding, rbcL, matK, medicinal plants, Siwa Oasis.

1. Introduction

Egypt has a collection of flora and plant biodiversity that was completed by [1]. In Boulos’s study, the total number of vascular plant species in Egypt was estimated to be 2075. Khedr and his colleagues, 2002 [2] reported that the angiosperm flora of Egypt is represented by 120 families, 742 genera, and 2088 species. Many locations including the Mediterranean, Sinai Peninsula, and Gebel Elba, a mountain range that supports Acacia woodland, hold a considerable volume of this diversity. However, the northern Mediterranean fringe, western desert has the lowest plant biodiversity in Egypt. Plant diversity at all of these localities is precious because rare medicinal plants make up a large portion of this diversity. Sinai Peninsula, for example, is considered an important location in the Arabian Desert as it is rich in medicinal plants (veterinary, aromatic, and culinary) [3]. The Mediterranean coastline (26,000 Km) is also an area of high biodiversity where 4,284 known species, many of them endemic, are found [4].

Siwa Oasis is one of the depressions located in the northern part of the western desert of Egypt and covers almost 1,000 Km². Saline lakes, dune fields, gravel plains, cliffs and scattered populations of acacia tree give the region a rich variety of landscapes and provide habitat for a diverse flora. Plants found in this region include the endangered wild cotton and unique varieties of date palm. This diversity became endangered due to many factors associated with human activity such as urban and industrial development which incease waste products, increase of cultivation area, and introduction
of exotic species, grazing and tourism [5]. Therefore, the plant biodiversity currently found at the Oasis of Siwa needs more awareness as well as the development of a protection and conservation plan. Loss of plant diversity continues each day due to human interference in the Siwa Oasis and in most of the natural habitat in Egypt. Half of the diversity in Mediterranean coast plants has disappeared since development began along the shoreline. Collection and surveying of the plant species that exist in the Oasis and the surrounding areas by morphological and molecular means is an important approach. This approach will be the first step to increase awareness of biodiversity and conserve the natural habitat for medicinal and ecological benefits.

One way to assess this knowledge is through the use of DNA barcoding which is a molecular phylogeny method that uses a short standardized DNA sequence in a well-known gene to identify a plant, animal or fungus as belonging to a particular species [6]. Chloroplast genome sequence data is the main type of DNA used for plant barcoding. This is because the chloroplast genome has a haploid and stable genetic structure where there is no recombination, and it is generally uniparentally inherited [6-7]. In addition, universal primers can be used to amplify target sequences via PCR amplification followed by sequencing [8]. One of the most widely used regions for plant barcoding is \( \textit{rbcL} \) (Ribulose-1, 5–bisphosphate carboxylase/ oxygenase large subunit gene) which is responsible for the production of the large subunit of the enzyme RuBisCo (important for carbon fixation). The portion of plastid gene \( \textit{rbcL} \) consists of a 599 bp region at the 5′ end of the gene is certainly the most sequenced locus among land plants. It has been known for its straightforward recovery and providing a useful backbone to the barcode dataset [9-10]. However, low sequence variation at the family level and below leads to modest discriminatory power, and therefore multigene analyses including genes besides \( \textit{rbcL} \) are required to corroborate the findings [11-13]. The Consortium for the Barcode of Life (CBOL), recommends the use of portions of two plastid coding regions, \( \textit{rbcL} \) and \( \textit{matK} \), taken together, as a barcode for plants (CBOL 2009). The \( \textit{matK} \) barcode region consists of a ca. 841 bp region at the center of the gene. \( \textit{matK} \) is one of the most rapidly evolving coding portions of the chloroplast genome [14]. The high discrimination power at family, genus, and even species levels recommends \( \textit{matK} \) as one of the most versatile candidates for barcoding [12], [15]. Unfortunately, \( \textit{matK} \) can be difficult to amplify and sequence from some taxa and optimization of PCR conditions is needed [16-17]. Since \( \textit{rbcL} \) and \( \textit{matK} \) are coding regions, electronic translation of data sequences from DNA to amino acids can be used to proofread for any errors in sequence assembly, correct sequence orientation and check for the presence of pseudogenes. The ease of sequence assembly in a single alignment facilitates comparative analyses of DNA barcode diversity among taxonomic groups and geographical regions.

There are many applications for barcoding, including the involvement of local researchers in international networks and biodiversity initiatives which should increase the awareness of biodiversity in Egypt. However, only a few studies have been conducted on barcoding specimens from Egypt [18], [19]. In order to develop an effective conservation plan of Egypt biodiversity, intensive studies of many localities and flora as well as the use of effective, quick, and economical molecular identification are needed. In this study, we aim to discriminate and confirm the identification of five medicinal plants at the molecular level, using \( \textit{matK} \) and \( \textit{rbcL} \) regions as well as confirm their morphological identification. Blasting the \( \textit{matK} \) and \( \textit{rbcL} \) sequences against the reference sequences available at The Barcode of Life Data System (BOLD) [20] and Genbank [21] allow us to determine how effective were the two regions to discriminate and identify the species under investigation.

2. Materials and Methods

2.1 Sample collection:
Seeds of many taxa were collected from Siwa Oasis, west desert, Egypt and were identified according to their morphological characteristics [1], [22]. Species identified during the course of this study are \textit{Nitraria retusa}, \textit{Cynara sibthorpiana}, \textit{Capparis spinosa}, \textit{Peganum harmala}, \textit{Pergularia tomentosa}. For
each species we sequenced two specimens. The selection of plant species under consideration was based on availability of tissue samples, good quality DNA and sequence data.

2.2 DNA Extraction:
DNA was extracted using the DNeasy kit solutions and spin column (Qiagen) or spin column (Epoch biolabs). Seed tissue was disrupted in 2 ml microfuge tubes in which one tungsten ball was added. Tubes were placed in the grinding rack of TissueLyser II and tissue was ground at 30 strokes/second for 1.5 minutes. The orientation of the tube was switched and this step was repeated. DNA was released with detergent and chaotropic agents. Proteins, polysaccharides, and cell debris were then eliminated and DNA was purified further by a silica bind-wash-elute procedure in micro-centrifuge spin columns. DNA quality was tested using agarose gel electrophoresis and quantified using spectrophotometry.

2.3 DNA amplification and sequencing:
Purified DNA was subject to PCR and subsequent sequencing. The following primer pair was used to amplify\textit{rbcL}: Forward-ATGTCACCACAAGAGACTAAAGC and Reverse-GAAACGGTCTCTCCAACGCAT [23-24]. For \textit{matK}, we used this primer pair: Forward-CGTACAGTACTTTTGTGTTTGAG and Reverse- ACCCAAGTCATCTGGAAATCTTGGTC [16]. PCR amplification was performed in a volume of 20 µl. Reactions contained 1× buffer, 2 mM MgCl₂, 2.5 µM of each dNTP, 0.35 µM of each primer, 20 ng of genomic DNA and 1U of \textit{Taq}. After a 2.5 minute initial denaturation step at 95°C, amplification of \textit{rbcL} proceeded for 34 cycles at 95°C, 58°C, and 72°C for 30 seconds each, and a final elongation step for 10 minutes at 72°C. Amplification of \textit{matK} was similar to \textit{rbcL} except that the annealing temperature was 50°C and the extension temperature was 60°C. The PCR products were visualized using a 1.3% agarose gel. When successful, PCR products were sent for sequencing using the same primer pairs.

2.4 Sequence Alignment/editing and Statistical Tests:
Sequences were assembled into contigs using Sequencher 4.9 (Gene Codes Inc.). The sequences of \textit{rbcL} and \textit{matK} were handled in separate alignments. Due to the absence of indel variation in \textit{rbcL}, sequences were readily aligned and unambiguous. For \textit{matK}, a few indel variations were expected therefore, gaps were allowed during alignment. All alignments were checked for an open reading frame to exclude any pseudogenes. Ambiguous regions were trimmed from the analyses. For each species sequence, we ran a BLAST test against the BOLD [20] and Genbank [21] database for identification of specimens at family, genus, and species levels. From the BLAST result, we downloaded all sequences of genera within the family of interest. To identify the phylogenetic relationship of each specimen, the alignment was analyzed using MEGA6 [25]. The evolutionary history was inferred by using the Maximum Likelihood method. Initial tree(s) for the heuristic search were obtained by applying the Neighbor-Joining method to a matrix of pairwise distances estimated using the Maximum Composite Likelihood (MCL) approach. A discrete Gamma distribution was used to model evolutionary rate differences among sites (4 categories (+G, parameter = 0.1127)). Trees are drawn to scale, with branch lengths measured in the number of substitutions per site.

3. Results and Discussion

The development of an efficient conservation plan of biodiversity requires reliable identification of the life forms in the ecosystem. Morphology-based species identification is the first tool, but could also be limited in some cases [15], [26-27]. DNA barcoding is a rapid, reproducible and economic tool for identifying living organisms and provides additional and complementary support to morpho-species identification [28].
3.1 Morphological identification:
Each specimen was identified at the species level based solely on the morphological characteristics recorded upon collection of specimens. To identify specimens morphologically, we followed Boulos [1]. Specimens were then given a number. All five species we study herein are of considerable importance in the field of human health care.

Sample 2 was morphologically identified as *Nitraria retusa*, a salt-tolerant shrub known for small, edible fruits which belongs to the *Nitrariaceae* family. The plant is native to desert areas of northern Africa [29]. Some studies show their potential use as an antioxidant and antimicrobial agent [30]. Ethanol extract (NRE) from *Nitraria retusa* shows effectiveness against obesity [31].

Sample 3 was identified as *Cynara sibthorpiana* or wild artichoke and is a member of the family *Asteraceae*. Following the discovery of cynarin, many species that belong to the *Cynara* genus have become important as medicinal herbs. This bitter-tasting compound is found in leaves and improves liver and gall bladder function, stimulates the secretion of digestive juices, especially bile, and lowers blood cholesterol levels [32]. Leaves are usually harvested just before flowering and can be used fresh or dried to treat chronic liver and gall bladder diseases, jaundice, hepatitis, arteriosclerosis and the early stages of late-onset diabetes [33-34].

Sample 5 was identified as *Capparis spinosa* and belongs to the family *Capparaceae*. It has many uses as a medicinal plant. The root can be used internally in the treatment of gastrointestinal infections, diarrhea, gout and rheumatism [35]. Externally, it is used to treat skin conditions, capillary weakness, and ease bruising [33]. The flower buds are a laxative, and also used to ease stomach pain, in the treatment of coughs, and externally to treat eye infections [33].

Sample 7 was identified as *Peganum harmala* and is a member of the family *Nitrariaceae*. *P. harmala* has been shown to have antibacterial and anti/protozoal activity [36]. Seed extracts also show effectiveness against various tumor cell lines, both in vitro and in vivo [37-38].

Sample 9 was identified as *Pergularia tomentosa*, which is a perennial herbaceous plant belongs to the family *Apocynaceae*. *P. tomentosa* is commonly used as a laxative and abortive of skin diseases and a depilatory [39]. *P. tomentosa* has been shown to have potent antioxidant and cytotoxic properties [40]. Piacente and his colleagues [41] reported that *P. tomentosa* inhibits Na+/K+ ATPase activity and displays cytotoxic properties, preventing the morphological changes seen in cancer cell lines. Martin and his colleagues [42] reported cytotoxic activity of *P. daemia*, a related species, against cell lines derived from an ovarian cancer.

3.2 Molecular verification of species identification:
Sequences of *rbcL* (536-585 bp) or *matK* (722-767 bp) from each species were BLASTed against Genbank and BOLD. From the BLAST results, we downloaded the sequences with query cover of ≥99%, that belonged to the family suggested by morphological observations for each species. Confirmation of the morphological identification based on our barcoding results was most effective for family level taxonomic assignment or discrimination, less effective at the genus level, and least effective at the species level, an observation similar to that made by Elansary [18].

The *matK* phylogenetic analysis of sample 2 supports its inclusion in the family *Nitrariaceae*, and specifically places the sample within the genus *Nitraria* (Figure 1A). The position of sample 2 as sister to the other *Nitraria* species was highly supported (bootstrap support of 94%). *matK* BOLD blast results show that the sequence of sample 2 has 98.26% similarity to all three sister species *Nitraria sphaerocarpa*, *Nitraria sibirica* and *Nitraria roborowskii*. However, the *rbcL* analysis (Figure 1B) reveals powerful discrimination and suggests that sample 2 not only belongs to the *Nitrariacea* family but is also closely related to *Nitraria retusa* (bootstrap support of 86%). Both Genbank and BOLD support 100% identity between sample 2 and *Nitraria retusa* for *rbcL*. The similarity was 99.65% to the three sister species mentioned above. This finding shows that the *rbcL* region has a better resolution toward species identification than the *matK* region, as fewer *Nitraria* sequences were available for *matK* than *rbcL* in Genbank and BOLD.
Sample 3 was morphologically identified as the species *Cynara sibthorpiana*, which belongs to the family *Asteraceae*. The phylogenetic analysis of sample 3 using *matK* or *rbcL* shows very high discrimination support to the family *Asteraceae*. However, the genus and so the species level discrimination using *rbcL* region was very low (figure 2B) compared to *matK* analysis (Figure 2A). The blast results using *rbcL* region showed 100% similarity with at least 6 other genus other than *Cynara* genus. As a result, sample 3 was not supported to belong to either one of them (bootstrap support of 0%) (figure 2B). Although the *matK* phylogenetic tree is more resolved than the *rbcL* tree, there is not enough resolution to say that sample 3 is certainly related to a single species included in the alignment (Figure 2A). On the other hand, the *matK* blast results from BOLD show a 100% similarity to an early-release (not published) sequence belong to species *Cynara cardunculus* ssp. *cardunculus* As a result of not having clear discrimination power of sample 3 using either *matK* or *rbcL*, we kept the identification of specimen no 3 as *Cynara sibthorpiana* until further analysis with other marker. This finding support the conclusion that *rbcL* mostly has modest discriminatory power and is not suitable to be applied as a solo marker for DNA barcoding (Chase et al, 2005, Rubinoff et al, 2006, Group 2009, Xiwen, et al, 2015). Another example support this finding is the analysis results of sample 7 sequences in which *matK* phylogenetic analysis support it as a member of the family *Nitrariaceae*, and specifically to genus *Peganum* (Figure 4A). Sample 7 was highly supported to *Peganum harmala* (bootstrap support of 98%). This identification of sample 7 as *Peganum harmala* was supported by *rbcL* phylogenetic analysis (figure 4B), but the resolution toward species identification was less than *matK* region.

The blast results of sample 5 from Genbank were similar to BOLD identification blast. The *matK* phylogenetic analysis of sample 5 sequences has shown more powerful discrimination and strongly supported to the family *Capparaceae*, and specifically to genus *Capparis* (figure 3A). Sample 5 was highly supported to either of two species *Capparis spinosa* or *Capparis sandwichiana* (bootstrap support of 98%). On the other hand, *rbcL* analysis did not show a clear cut of support to either the family *Burmanniaceae* or *Capparaceae*. In addition, using *rbcL*, sample 5 was not identified to belongs to which genus; *Capparis*, *Gymnosiphon*, or *Geomitra* (figure 3B).

Sample 9 was identified morphologically as *Pergularia tomentosa* which belongs to the family *Asteraceae* and subfamily *Asclepiadoideae*. Both *matK* and *rbcL* analysis were capable to discriminate sample 9 at the level of family and subfamily to be *Asteraceae* and *Asclepiadoideae* respectively. When come to the genus discrimination, our blast results in BOLD database show similarity with Private (not-published) sequences, feature that is not available in Genbank database. In BOLD database, sample 9 blast of either *matK* or *rbcL* sequence shows 99.87% and 100% similarity, respectively, to the species *Pergularia daemia* which is not published yet. The blast in Genbank database and so our phylogenetic analysis is also showing P. *daemia* (figure 5A, B). On Genbank, the blast pick up a partial sequence of *matK* from P. *daemia* of which query cover was only 95%. Therefore the *matK* analysis includes p. *daemia* branch, but the analysis is not highly supported toward this species (bootstrap support 44%) because the full sequence is not available yet. Despite the weak support to P. *daemia*, it is the highest bootstrap support compared to the other genus on the same clade with sample 9. This was exception from our standardize use of blast results of ≥ 99% similarity to be included in phylogenetic analysis. On the other hand, *rbcL* sequence blast pick up a partial sequence from P. *daemia* of which query cover was 97%. The *rbcL* phylogenetic analysis shows support toward P. *daemia*, bootstrap support 71%. The analysis shows more support (bootstrap support 82%) toward P. *tomentosa*, of which query cover was 99%. Therefore phylogentic analysis of sample no. 9 *rbcL* sequence supports its identity as *P. tomentosa*. This finding is similar to the case of sample 2 and supports the conclusion that the more *rbcL* sequences published compare to *matK* will improve the power of identification of plant species.
4. Conclusions

Blasting the matK and rbcL sequences against the reference sequences available at The BOLD [20] and Genbank [21] allow us to determine how effective were the two regions to discriminate and molecularly identify the species under investigation. Confirmation of the morphological identification based on rbcL or matK barcoding results was most effective at the family level taxonomic assignment, less effective at the genus level, and least effective at the species level. Our results showed that matK was more powerful than rbcL in discriminating and identifying samples at the species level and rbcL was not suitable as a solo marker (Figures 2-4). Only at the circumstances, with more available rbcL sequences on Genbank or BOLD, rbcL marker offers better resolution in species identification as it was the case for samples no. 2 and 9 (Figures 1 and 5). Sequence blast against BOLD database allow us to know some information about unpublished data that give us insight about the identity of our samples for future studies.

5. Acknowledgment

Special Thanks to Prof. Damon Little (NYBG) for technical support and advice. Deep thanks Prof. Hussam E. Elwakil (Alexandria University, Egypt) for the morphological identification and providing the samples. Special thanks to Dr. Elizabeth McCarthy (UCR) for her comments on the manuscript.

References


[34] Tropical Plant Database, 1996-present. RAINTREE. [online] Available at: http://www.rain-tree.com/artichoke.htm


[40] AL Jabri, S., 2013 Chemical and Bio-analytical Studies on Pergularia tomentosa and Species from the Mentha Genus. Ph.D. University of Leicester.
