Phylogenetic relationships and DNA barcoding of nine endangered medicinal plant species endemic to Saint Katherine protectorate

Ahmed M. Hashim, Aishah Alatawi, Faris M. Altaf, Sameer H. Qari, Mohamed E. Elhady, Gamal H. Osman, Heba H. Abouseada

A high degree of endemism has been recorded for several plant groups collectively in Saint Katherine Protectorate (SKP) in the Sinai Peninsula. Nine endangered endemic plant species in SKP were selected to test the variable abilities of three different DNA barcodes; Ribulose-1,5-Biphosphate Carboxylase/Oxygenase Large subunit (rbcL), Internal Transcribed Spacer (ITS), and the two regions of the plastid gene (ycf1) as well as Start Codon Targeted (SCoT) Polymorphism to find the phylogenetic relationships among them. The three barcodes were generally more capable of finding the genetic relationships among the plant species under study, new barcodes were introduced to the National Centre for Biotechnology Information (NCBI) for the first time through our work. The barcode sequences were efficient in finding the genetic relationships between the nine species. However, SCoT polymorphism could only cluster plant species belonging to the same genus together in one group, but it could not cluster plant species belonging to the same families except for some primers solely. rbcL was the most easily amplified and identified barcode in eight out of the nine species at the species level and the ninth barcode to the genus level. ITS identified all the species to the genus level. Finally, ycf1 identified six out of the eight species, but it could not identify two of the eight species to the genus level.

© 2021 The Authors. Published by Elsevier B.V. on behalf of King Saud University. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).
environmental pressure until extinction (Al-Qurainy et al., 2018). Therefore, the conservation of genetic diversity is important, as genetic diversity provides continuing evolutionary potential against changing environmental conditions (Dobson, 1998).

Medicinal plants play important roles on the therapeutic and economic levels for residents of SKP. The nine target species in this study are collected from their natural habitats for many purposes; for home use, trade and folk medicine by traditional healers and for scientific research, in many cases these plants are collected for grazed and used for fuel wood (Aghakhani Kaaji and Kharazian, 2019; el-Negoumy et al., 1986; El-Wahab et al., 2004; Hashim et al., 2020; IUCN, 2015; Khafaga et al., 2011; Mokni et al., 2019; Moustafa et al., 2015; Omar, 2014; Pieroni et al., 2006; Shaltout et al., 2015; Uritu et al., 2018; Zahran and Willis, 2009). Potential applications in medicine, pharmacy, and in other industrial uses of the nine studied species are mentioned in table S1.

Based on the red list categories of the International Union for Conservation of Nature (IUCN), endemic species found at high conserving hotspots should be documented and inventoried periodically to evaluate their population status and distribution (Ayyad et al., 2000; Baillie and Butcher, 2012). However, botanical records are generally not complete in many biodiversity hotspots, possibly because the taxonomic duty is thwarted by the low discriminatory practice of morphological descriptors for species that are closely related (Francisco-Ortega et al., 2007).

DNA barcoding and genomics approaches are now being applied in many of biodiversity studies, including species identification (Hollingsworth, 2008; Pope et al., 2017; Savolainen and Karhu, 2000), detection of new taxa (Bell et al., 2012), DNA barcoding for big taxonomic groups (Ojeda et al., 2014), addressing detached taxonomic problems (Feau et al., 2011), conservation of species (Vessson et al., 2011), enabling studies on ecology of plant communities via building phylogenetic trees (Joly et al., 2014).

DNA barcode consists of standard short sequence of DNA, which is unique for every species on our planet in principle and can be easily generated. Numerous plastid sequences were considered, validated and recommended for plant species as ideal barcode loci (Ford et al., 2009; Kress et al., 2005; Pennisi, 2007). The two chloroplast genes segments rbcL and matrurate k (matK) are a pair of these recommended sequences in 2009 by the Consortium for the Barcode of Life (CBOL) Plant Working Group as core barcode (CBOL Plant Working Group, 2009). Given the universality of rbcL gene, it has been proposed as a barcode fragment (Hollingsworth et al., 2016). Currently, rbcL genes have been commonly used in family and subclass phylogenetic analysis among different seed plant groups (Chase et al., 2007). However, variation in rbcL sequence mainly exists at the above-species level, and variation is rarely found at the species level, resulting in low capabilities in species discrimination. Moreover, rbcL seems to be more suitable for barcoding lower plants than for seed plants (Chase et al., 2007). ITS barcode is a ribosomal DNA in the nuclear genome, that is widely distributed in all photosynthetic eukaryotes (except ferns). ITS fragment includes ITS1, ITS2 and 5.8 S, with great differences among the three sequences (Chase et al., 2005). ITS barcode characterized by the following advantages: 1) high rate of species identification due to its highly repeated in nuclear genome (Kress and Erickson, 2007); 2) accurately helps in reconstruction of phylogenetic relationships between plant species specifically in lower taxonomic order (Baldwin et al., 1995); 3) there is a large amount of data from ITS barcode have been recorded in GenBank (Hollingsworth et al., 2016). Ycf1 is the second largest gene in chloroplast genome, that encodes a protein of about 1800 amino acids and crucial for viability of plant species (Dong et al., 2015; Kikuchi et al., 2013). Since ycf1 is too long and likewise variable for the design of universal primers, attention has been paid to DNA barcoding purposes at low taxonomic levels, however its high variability indicates its potency in molecular systemic of higher plants (Dong et al., 2012).

Start Codon Targeted (SCoT) polymorphisms are predominant and reproducible new markers developed on the basis of the short-conserved region flanking the translation start codon (ATG) in plant genes (Bhattacharyya et al., 2013; Collard and Mackill, 2008), which makes it advantageous to other molecular marker techniques. Also, they are considered to be more efficient and reliable, as they exhibit a relatively high annealing temperature (Shahlaei et al., 2014). This technique is designed for most plant research laboratories with standard kits by conventional gel electrophoresis with agarose gels and stains (Zhang et al., 2015). Additionally, because of the lower levels of recombination between gene/traits and SCoT polymorphism, compared with random markers such as Simple Sequence Repeats (SSRs), Random Amplified Polymorphic DNA (RAPD), and Inter-Sequence Simple Repeats (ISSRs), SCoT polymorphisms have a direct use in marker-aided breeding programs (Mulipuri et al., 2013). SCoT markers have been effectively used in the identification of cultivars to evaluate genetic diversity in numerous species including sugarcane, rice, grape, mango, potato, garbanzo and cacti (Abouseada et al., 2020; Amirmoradi et al., 2012; Cabo et al., 2014).

In this study, we aimed to identify and classify the nine endemic vascular plant species in SKP using the three DNA barcodes (rbcL, ITS, and ycf1) for the first time concerning most of them. Moreover, to evaluate the potential and difference between the three barcodes. And finally, to construct phylogenetic trees for assessment the genetic relationships among these nine species using the three barcodes as well as SCoT polymorphism technique. These species are extremely rare, endangered, or critically endangered under the criteria of the IUCN. The results obtained provide a scientific basis for the molecular identification, controlled collection, study the evolutionary-relationships and future conservation of these endemic species.

2. Materials and methods

2.1. Plant materials and collection

Nine endangered, endemic, angiosperm plant species were selected from their high-altitude natural habitats in the SKP mountainous region (Fig. 1). The protectorate is located between 28° 30’ to 28° 35’ N and 33° 55’ to 34° 30’ E, and the plateau elevation is between 1,300 to 2,600 m above sea level (Moustafa and Zaghloul, 1996; Moustafa, 2017). The most important consideration for plant species collection was the fact that all collected plants are endemic, endangered, or critically endangered according to the red data book of IUCN (IUCN, 2012), the number of collected individuals per each species was restricted (Almost one individual per species).

Species identification and assignment were independently confirmed prior to the molecular studies and was based on an assessment of morphological descriptors developed by the herbarium section of the Botany Department, Faculty of Science, Ain Shams University, Cairo, Egypt.

The scientific names of the targeted endemic species with their families, life forms, distribution, and their IUCN status are shown in Table 1. The photos of species in their natural habitats are shown in Fig. 2. Plant samples were collected during the spring season of 2019. Immediately after collection, the fresh leaves from the plant samples were placed in silica gel beads for further lab analysis.

2.2. DNA extraction

Genomic DNA was extracted from 100 mg silica-gel dried leaf tissue for each plant species by using a DNeasy Plant Mini Kit (QIA-
GEN, Santa Clarita, CA, USA) according to the manufacturer’s protocol. The concentration of the extracted DNA was measured using a Nanodrop Spectrophotometer (Thermo Fisher Scientific Inc.). DNA concentrations were then adjusted to 10 ng/μL in all samples for subsequent SCoT polymorphism and DNA barcoding PCR amplification.

2.3. SCoT PCR analysis

SCoT polymorphism PCR-based amplification was performed using nineteen SCoT primers as developed by (Collard and Mackill, 2008). The primers were synthesized by Macrogen (Seoul, Republic of Korea) in 10 nmol stock concentration, obtained lyophilized, then rehydrated using sterile water to become 100 μM final concentration, and finally stored at –20 °C. The nineteen primers were screened against the nine plant species. Only ten of them gave prominent and reproducible bands with the nine samples, so that continued to the final analyses. PCR was performed on reaction mixtures of 25 μL, 25 ng of genomic DNA, 2X MyTaq red Mix 50 Reactions (BIOLINE, Lot No: MTRX-516204), 2.0 μL of each primer (2.5 μM), and distilled deionized water. The PCR amplification protocol was adjusted as follows: 94 °C for 1 min, 50 °C for 1 min, then 72 °C for 90 s) with a final extension at 72 °C for 7 min. PCR amplification products were resolved by electrophoresis in 2% agarose gel containing ethidium bromide (0.5 μg/mL) in 1X TBE buffer. A 100-bp-plus DNA ladder was used as a molecular size standard. The PCR products were visualized using UV light and photographed using a Gel Doc™ XR + System with Image Lab™ Software (Bio-Rad).

2.4. DNA barcodes PCR analysis

PCR amplification of the three barcode loci; two plastid loci the (rbcL and ycf1) barcodes and one nuclear locus for the rDNA (ITS), were performed in a Thermal Cycler TC-TE BOE 8,089,602 (BOECO, Germany) using three primer pairs, shown in Table 2. PCR reactions were performed in a 50-μL volume containing approximately 50 ng of genomic DNA, 2X MyTaq red Mix 50 Reactions (BIOLINE, Lot No: MTRX-516204), 1.0 μL of each primer (2.5 μM), and distilled deionized water. The reaction conditions were as follows: initial denaturation at 95 °C for 5 min, 40 cycles at 94 °C for 30 s, 58 °C for 30 s, 72 °C for 45 s, and 72 °C for 10 min. The desired PCR products were separated by electrophoresis in 1.5% agarose gels containing ethidium bromide (0.5 μg/mL) in 1X TBE buffer. A 100-bp-plus DNA ladder was used as a molecular size standard.

Table 1
List of the studied endemic species with their families, Arabic names, life forms, chorology, and their IUCN status.

| Family         | Plant species                        | Arabic name | Life form | Distribution       | *IUCN status       |
|----------------|--------------------------------------|-------------|-----------|--------------------|--------------------|
| Solanaceae     | Hyoscyamus boveanus (Dunal) Asch. & Schweinf. | النسار | Chamaephyte | Endemic to Egypt   | Endangered         |
| Caryophyllaceae| Buphonia multisecis Decne.           | النمة       | Hemi-cryptophyte | Endemic to SKP     | (very rare) Endangered |
|                | Silene leucophylla Boiss.            |            | Hemi-cryptophyte | Endemic to SKP     | Endangered         |
|                | Silene schimperiana Boiss.           |            | Hemi-cryptophyte | Endemic to SKP     | Endangered         |
| Lamiaceae      | Nepeta septemcrena Benth.            | النية       | Chamaephyte | Endemic to SKP     | Endangered         |
|                | Phlomis aurea Decene.                | عورور        | Chamaephyte | Endemic to SKP     | Endangered         |
| Polygalaceae   | Polygala sinaica var. sinaica Botsch. | ديل - مغر | Chamaephyte | Endemic to SKP     | Endangered         |
| Primulaceae    | Primula boveana Decne. ex Duby       | فتح الجيل - دماغ | Hemi-cryptophyte | Endemic to SKP     | Critically endangered |
| Rosaceae       | Rosa arabica Crep.                   | النور الدياري | Nano-phanerophyte | Endemic to SKP     | Critically endangered |
The PCR products were visualized using UV light and photographed. Amplicons of the appropriate size were purified from the gel using Gene JET Gel Extraction Kit, Thermo Scientific Catalog number: K0691. The purified PCR products were directly sequenced by Macrogen lab (Seoul, South Korea).

The obtained sequences were submitted to the GenBank and were given the following accession numbers: MT333246, MT333247, MT333248, MT333249, MT333250, MT333251, MT333252, MT333253, MT333254, MT333255, MT333256, MT333257, MT333258, MT333259, MT333260, MT333261, MT333262, MT333263, and MT333264.

2.5. Data analysis

For SCoT polymorphism data analysis, the amplified bands were scored using the free software PyElph1.4 (https://pyelph.software.informer.com/1.4). Only clear bands were scored, while faint bands were neglected. The bands were scored (1) for presence or (0) for absence to create a binary dataset. The capacity of SCoT primers to discriminate between the nine genotypes were determined by calculating the values of polymorphic information content (PIC), Resolving power (Rp) and Marker Index (MI). PIC value for each primer, which reflects the degree of detecting polymorphism, was calculated according to (McGregor et al., 2000) using the formula: PIC = 1- \sum (Pi)^2, where Pi is the proportion of samples carrying the i-th allele of a particular locus. Rp value was calculated according to (Powell et al., 1996) using the following equation; Rp = \Sigma IB, where IB represents the band informativeness, was measured from the following equation, IB = 1 - 2 \times | 0.5 - p |, where P is the frequency of accessions that ports bands (Prevost and Wilkinson, 1999). The number of amplified polymorphic bands was divided by the total number of amplified bands by the same primer or primer combination to calculate the percentage of polymorphism. Jaccard’s coefficient was used to estimate genetic similarity matrix (Jaccard, 1908). Past 4.03 free software (https://past.en.lo4d.com/windows) was used to construct the dendrogram and the principal component analysis (PCA). The dendrogram was created by using cluster analysis and the un-weighted pair group method of the arithmetic averages (UPGMA). The Principal component analysis (PCA) was performed using a D centre module (Jaccard, 1908).

The resulting DNA barcodes sequences were analysed using different software and online tools, but first the 3’ and 5’ peripheral noisy parts of each sequence were trimmed, and then primers were removed. The local CLUSTAL W (Thompson et al., 1994) free software (http://www.clustal.org/clustal2) was used for the sequence alignment. The aligned sequences were submitted to the NCBI BLAST database (https://blast.ncbi.nlm.nih.gov/Blast.cgi) for species identification and were given the following accession numbers: MT333246, MT333247, MT333248, MT333249, MT333250, MT333251, MT333252, MT333253, MT333254, MT333255, MT333256, MT333257, MT333258, MT333259, MT333260, MT333261, MT333262, MT333263, and MT333264.
alignment of barcode sequences. An ITOL online website (https://itol.embl.de/) was used for phylogenetic tree construction (Letunic and Bork, 2007). Principal component analysis (PCA) and heat map analysis were conducted using R software provided by the online free tool ClustVis 2.0 (https://biit.cs.ut.ee/clustvis). BLASTn was used for aligning DNA barcode sequences against sequences in the National Centre of Biotechnology Information (NCBI) database using default parameters (Altschul et al., 1997).

3. Results

3.1. Molecular phylogeny based on SCoT marker analysis

Nineteen SCoT primers were used in the SCoT marker analysis. After three trials, only ten of them were reproducible and yielded scorable bands as shown in Table 3 and Fig. 3. The total number of amplification products was 197, out of which 196 were polymorphic with a polymorphism percentage of 99.5%.

The number of scorable bands per primer ranged from 12 for primer SCoT 20 to 26 for primer SCoT 33, and the polymorphism percentage ranged from 94.7% for primer SCoT 23 to 100% for primers Scot 1, 6, 11, 12, 14, 15, 16, 20, and 33 (Table 3).

The average polymorphism information content (PIC) of the ten used SCoT primers was 0.81 (Table 3). The PIC ranged from 0.75 to 0.87; the highest PIC values were for primers SCoT 6 and SCoT 15, and the lowest PIC values were for primers SCoT 20 and SC.

The Rp value ranged from 7.1 to 14.7 (Table 3); where the highest RP values were for primers SCoT 11 and SCoT 33, and the lowest RP values were for the primers SCoT 20 and SCoT 6, with an average of 11.9 for the ten Primers.

The SCoT analysis-based dendrogram obtained using the UPGMA method according to Jaccard's coefficient divided the nine species into two main clusters, as shown in Fig. 4A. The first cluster has only sample 9 (Primula boveana). The second group is further subclustered to contain the remaining species. Samples 6 and 7 Silene leucophylla and Silene schimperiana were separated together in one group. The results obtained from PCA (Fig. 4B) were in harmony with dendrogram analysis, and both showed comparable groups. According to Jaccard's similarity matrix (Table 4), the highest similarity (80.6%) was found between Silene leucophylla and Silene schimperiana that belong to the same genus followed by (47.2%) for Nepeta septemcrenata and Phlomis aurea that belong to the same family.

3.2. Plant identification and genetic relationships using DNA barcodes

3.2.1. PCR amplification and identification of the samples

The three barcode primers rbcL, ITS, and ycf1 produced expected product sizes of approx. 600 bp, 900 bp, and 750 bp, respectively, except for Polygala sinaica with the ycf1 barcode, which did not give the expected result even after many trials with different conditions. The BLASTn tool provided by the NCBI was used to identify most of the sequences according to the species or genus level, on the basis of the availability of corresponding sequences in the database. Species identification according to the rbcL barcode revealed that the rbcL gene was successful in identifying eight out of the nine studied plant species at the species level except for Rosa arabica, (Fig. 4A) with a query coverage of 98%-100% and an identity percentage of 94.7%-99.81% (Table 5A). Rosa arabica was only identified to the genus Rosa, due to absence of its rbcL sequence, and as mentioned before, the rbcL sequence published for Rosa arabica from our work is the first on databases. For the ITS barcode, all the plant species were identified to the genus level only (Fig. 5B) with a query coverage of 100% and an identity percentage of 92.65%-99.36% (Table 5B), but it was unable to identify two species (Phlomis aurea and Bufonia multiceps) neither to the species nor to the genus levels.

Heat maps are based on the similarity matrices generated using the barcodes sequences of the studied plants and previously published barcodes sequences. It reflects how the barcode sequences show possible diversity within the species and genus. Heat map analysis was used to visually differentiate the structure of studied plant samples according to the color intensity. As the color intensity increases, the similarity increases. The analysis revealed the high capacity of ycf1 followed by ITS to differentiate the studied samples according to species and genus (Fig. 15 (A, B and C)).

PCA is also based on the similarity generated by sequences alignment. It can explain the diversity by collecting the most similar samples in one cluster, which could be helpful in finding the closest samples according to sequence similarity. Similar results have been concluded using PCA analysis, where ycf1 and ITS overload rbcL DNA barcode (Fig. 2S (A, B and C))

3.3. Phylogenetic relationships based on plant DNA barcodes analysis

The barcode sequences obtained from the plant species under study were subjected to multiple alignment with each other. Multiple alignment revealed considerable difference between different samples at the genus level (Fig. 6 A-C). With respect to rbcL, ITS, and ycf1 genes, sample 4 (Bufonia multiceps) was the most different among other species per the evolutionary basis. The three barcodes were all successful in combining samples 6 and 7 (Silene leucophylla and Silene schimperiana) in one cluster, which belong to the same genus. Samples 6, 7, and 4 (Silene leucophylla, Silene schimperiana and Bufonia multiceps) all belong to family Caryophyll-
laceae. However, only rbcL and ycf1 combined samples 2 and 3 (Phlomis aurea and Nepeta septemcrenata) which belonged to family Labiatae, together in one cluster.

The rbcL available sequences for the studied species that were previously published on the NCBI database, were used to construct a phylogenetic tree to compare it with the tree obtained from the
sequences of our own specimens. Both phylogenetic trees (Fig. 6.A and Fig. 7) were the same.

4. Discussion

To the best of our knowledge, this is the first work to examine the ability of DNA barcodes and SCoT polymorphism in investigating the genetic relationships between the selected species. In addition, DNA barcoding involves the production of PCR amplicons from particular regions to sequence them and these sequence data are used to identify or “barcode” that organism to make a distinction from other species (Lebonah et al., 2014), one \(rbcL\) sequence for \(Rosa arabica\), and all ITS and \(ycf1\) sequences were novel barcodes for these species to the NCBI database.

SCoT molecular marker was able to differentiate between the nine species; it was also able to combine samples 6 and 7 that belonged to the same genus in one group but was unable to combine different species that belonged to the same family together in one group. Only different SCoT primers were able to combine different species that belonged to the same family together in one group; for example, SCoT primers 1, 12, and 23 combined samples 4, 6, and 7 that belong to the same family together in one group, and SCoT primers 6 and 16 combined samples 2 and 3 that belong to the same family together in one group. (Shahlaei et al., 2014) used ISSR and SCoT molecular markers to study the genetic diversity in \(Lycopersicum esculentum\), and they estimated that the mean PIC values to be 0.142 and 0.088 for SCoT and ISSR respectively, and the mean RP values to be 1.88 and 1.55 for both markers respectively.

DNA barcoding has been proposed as a powerful tool not only for identifying and confirming species but also for finding genetic relationships among the species. The three barcodes we used were all reproducible and produced the expected band size. Only \(ycf1\) with sample 5 (\(Polygala sinaica\)) could not produce the expected band size.

Table 4
Jaccard’s similarity matrix based on the SCoT analysis of the nine plant species.

|          | H. boveanus | P. aurea | N. septemcrenata | B. multiceps | P. sinaica | S. leucophylla | S. schimperiana | R. arabica | P. boveana |
|----------|-------------|----------|------------------|--------------|------------|----------------|----------------|------------|------------|
| H. boveanus | 100%       |          |                  |              |            |                |                |            |            |
| P. aurea   | 27.9%      | 100%     |                  |              |            |                |                |            |            |
| N. septemcrenata | 26.6%    | 47.2%    | 100%             |              |            |                |                |            |            |
| B. multiceps | 12.9%     | 23.3%    | 19.6%            | 100%         |            |                |                |            |            |
| P. sinaica | 24.8%      | 29.9%    | 28.7%            | 21.1%        | 100%       |                |                |            |            |
| S. leucophylla | 24.2%    | 23.6%    | 22.5%            | 40.2%        | 22.5%      | 100%           |                |            |            |
| S. schimperiana | 19.8%    | 24.1%    | 25.7%            | 45.9%        | 23.1%      | 80.6%          | 100%           |            |            |
| R. Arabica | 24.7%      | 20.7%    | 23.7%            | 20.4%        | 25%        | 23.1%          | 22.7%          | 100%       |            |
| P. boveana | 25.8%      | 25.9%    | 28.9%            | 22.6%        | 27.2%      | 20.9%          | 21.5%          | 19.1%      | 100%       |

Table 5
BLASTn results for the barcodes sequences against NCBI databases. Query; the barcode sequence obtained in the current study, Subject; NCBI published sequences, Subject ID; accession number of the NCBI published sequences.

(A) \(rbcL\)

| Query       | Subject species | Subject ID      | Bit score | Query cover | E-value | Identity % |
|-------------|-----------------|-----------------|-----------|-------------|---------|------------|
| H. boveanus | H. boveanus     | MF668605.1      | 939       | 100%        | 0.0     | 98.51%     |
| Phlomis aurea | Phlomis aurea | KY794564.1      | 931       | 99%         | 0.0     | 98.31%     |
| N. septemcrenata | N. septemcrenata | KY794558.1 | 974       | 99%         | 0.0     | 99.38%     |
| Bufonia multiceps | Bufonia multiceps | KX70610.1 | 806       | 100%        | 0.0     | 96.90%     |
| Polygonia sinaica | Polygonia sinaica | KY656729.1 | 896       | 98%         | 0.0     | 98.25%     |
| Silene leucophylla | Silene leucophylla | MK055336.1 | 948       | 98%         | 0.0     | 99.81%     |
| Silene schimperiana | Silene schimperiana | MF668591.1 | 865       | 100%        | 0.0     | 99.58%     |
| Rosa arabica | Rosa canina     | FN689381.1      | 843       | 100%        | 0.0     | 94.42%     |
| Primula boveana | Primula boveana | KY656738.1 | 826       | 100%        | 0.0     | 97.15%     |

(B) ITS

| Query       | Subject species | Subject ID      | Bit score | Query cover | E-value | Identity % |
|-------------|-----------------|-----------------|-----------|-------------|---------|------------|
| H. boveanus | H. bornmulleri | KU295790.1      | 795       | 100%        | 0.0     | 89.74%     |
| P. aurea    | P. lychnitis    | AV792771.1      | 730       | 100%        | 0.0     | 89.82%     |
| N. septemcrenata | N. crassifolia | AJS15307.1 | 198       | 100%        | 4e-47   | 93.08%     |
| Bufonia multiceps | Bufonia parviflora | JNS89044.1 | 891       | 100%        | 0.0     | 91.38%     |
| Polygonia sinaica | Polygonia erioptera | KF805107.1 | 1271      | 100%        | 0.0     | 98.21%     |
| Silene leucophylla | Silene flavescens | KX757520.1 | 1288      | 100%        | 0.0     | 99.03%     |
| Silene schimperiana | Silene armena | KX757619.1 | 1218      | 100%        | 0.0     | 98.41%     |
| Rosa arabica | Rosa canina     | FM164946.1      | 1112      | 100%        | 0.0     | 95.31%     |
| Primula boveana | Primula verticillata | AY680732.1 | 1242      | 100%        | 0.0     | 97.52%     |

(C) \(ycf1\)

| Query       | Subject species | Subject ID      | Bit score | Query cover | E-value | Identity % |
|-------------|-----------------|-----------------|-----------|-------------|---------|------------|
| H. boveanus | H. bornmulleri | KF248009.1      | 695       | 98%         | 0.0     | 96.02%     |
| Phlomis aurea | Colphounia coccinea | MN165115.1 | 1112      | 99%         | 0.0     | 94.01%     |
| N. septemcrenata | Nepeta catara | JF289048.1 | 797       | 100%        | 0.0     | 98.25%     |
| Bufonia multiceps | Psammoulinea tunicoides | NC_045947.1 | 418       | 100%        | 2,16E-112 | 94.30%     |
| Silene leucophylla | Silene littorea | MN365983.1 | 893       | 98%         | 0.0     | 98.051%    |
| Silene schimperiana | Silene littorea | MN365983.1 | 1086      | 90%         | 0.0     | 96.77%     |
| Rosa arabica | Rosa luciae     | MN689791.1      | 1411      | 100%        | 0.0     | 99.36%     |
| Primula boveana | Primula knuthiana | NC_039350.1 | 547       | 100%        | 2,64E-151 | 92.65%     |
band with different trials and conditions, emphasizing that the same DNA sample was used with the other successful barcodes. The use of the \textit{rbcL} barcode, although all \textit{rbcL} sequence for eight of the nine selected species are already published, was to standardize the consistency of our protocol and sequencing efficiency. As due to the scarcity and endangered nature of the selected species, we couldn't obtain more than one individual for each species to obey the rules stated by the organization of the nature protection sector in the Egyptian Environmental Affairs Agency (EEAA). The phylogenetic relationship between the 8 species and \textit{Rosa canina} as \textit{Rosa arabica} wasn't published before according to the \textit{rbcL} sequences available on the NCBI database was shown in the dendrogram (Fig. 7), and it was exactly similar to that obtained from the \textit{rbcL} sequences of our samples (Fig. 6 A), which ensures that our sequences are correct and the sequence efficiency was fine, and could be relied on with the rest of the sequences. Also, the success of the obtained sequences to correlate the specimens to either the species or genus level proves their success.

The produced barcodes identified most of the plant species under study, especially \textit{rbcL} that identified 8 out of 9 species to the species level and the ninth species (\textit{Rosa arabica}) to the genus level. The \textit{ITS} barcode identified all the nine species to the genus level, but the \textit{ycf1} barcode failed totally to identify two species (\textit{Phlomis aurea} and \textit{Bufonia multiceps}). This failure is due to the absence of available sequences for those species in the NCBI databases. DNA barcoding was more successful than SCoT in finding the
genetic relationships among different plant species, although both techniques could combine different species belonging to the same genus. However, DNA barcoding exceeded SCoT in combining different species belonging to the same family together in the same group, except for ITS with regard to samples 2 and 3 (Phlomis aurea and Nepeta septemcrenata).

The rbcl and ITS barcodes were used because of the recommendation of the Consortium for the Barcode of Life (CBOL Plant Working Group, 2009) that the chloroplast genes rbcl and matK are the core barcodes of plant species. In addition, the intergenic sequence trnH-psbA along with the nuclear gene ITS act as supplement barcodes (CBOL Plant Working Group, 2009). The rbcl barcode was favored in our study to matK because of its widespread availability, easy amplification, and comparability (Hollingsworth et al., 2016). In addition, the results obtained with the rbcl gene results confirm its efficiency in phylogenetic analysis within the family and subclass of angiosperms (Fazekas et al., 2008). The barcode could combine different species belonging to the same families together in one group. In contrary to (Gonzalez et al., 2009; Kress and Erickson, 2007; Newmaster et al., 2006) rbcl sequences were able to find variation at the species level and did not result in poor abilities of species discrimination as they proposed, but it was able to differentiate between samples 6 and 7 and identify them solely as Silene leucophylla and Silene schimperiana and was able to group them together. So, our study strongly recommends rbcl as a core plant barcode.

ITS identified all the plant species to the genus level (100%), but this result may be due to sequence unavailability on the databases for the species under study. ITS was previously recommended as a core DNA barcode for identifying land plant species (Chase et al., 2007; Kang et al., 2017). Ycf1 showed the least amplification percentage of 88.9%, compared with both rbcl and ITS (100%). In addition, ycf1 had the least identification ability, which was due to the unavailability of sequences in the databases. However, ycf1 produced a phylogenetic tree similar to that produced by rbcl, and its combined samples 4, 6 and 7 that belonged to the same family together in the same group. In addition, ycf1 also combined samples 2 and 3 belonging to the same family together in one group. The least amplification percentage of the ycf1 gene may be attributed to the absence of ycf1 in some taxa (Dong et al., 2015).

5. Conclusions

In conclusion, in this study, 19 new barcodes sequences were introduced for the first time to the NCBI databases. The efficacy of DNA barcoding to identify different plant species is great and powerful. It will add greatly to taxonomic and evolutionary studies that would be carried on these wild plant species. Also, DNA barcodes sequences were able to efficiently cluster the studied species into the appropriate groups. SCoT molecular marker was unable to cluster plant species belonging to the same family together, it was only able to cluster plant species belonging to the same genus. DNA barcodes are the most powerful tool for plant molecular identification, genetic diversity study, and would help and support in genetic conservation programs.

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.

Acknowledgements

Authors would like to express sincere gratitude to Prof. Dr. Soad Hassan, Professor of Plant taxonomy and flora, Department of Botany, Ain Shams University, for helping in plant species identification.

Funding

“This research received no external funding”
Fig. 6. (A-C) Phylogenetic tree for the nine plant species based on DNA barcodes sequences alignments (A) rbcL, (B) ITS and (C) ycf1.

Fig. 7. Phylogenetic tree for the nine plant species based on the rbcL sequences available on the NCBI databases.
essential oil of Phlomis floccosa D Don. Biomed. Chromatogr. 33, https://doi.org/10.1002/bmc.4603 e4603.

Moustafa, A.E.-R.A., El-Azeem, H., Omran, M.A., Nasr, S.A.M., Abdel-Nabi, L.M., Teleh, Z.A., 2015. Biochemical Effect of Nepeta Septempetacnea Growing in South Sinai, Egypt. American J. Ethnomed. 2.

Moustafa, A.E.-R.A., Zaghloul, M.S., 1996. Environment and vegetation in the montane Saint Catherine area, south Sinai Egypt. J. Arid Environ. 34, 331–349. https://doi.org/10.1006/jare.1996.0114.

Moustafa, A.G., 2017. Long Term Monitoring of Rosa arabica Populations as a Threatened Species in South Sinai, Egypt [WWW Document]. URL https://www.semanticscholar.org/paper/Long-Term-Monitoring-of-Rosa-arabica-Populations-as-Moustafa/a9209f600f5eef6928a0df021c7866cea10018f1 (accessed 4.2.20).

Moustafa, A.R.A., Klopatek, J.M., 1995. Vegetation and landforms of the Saint Catherine area, southern Sinai Egypt. J. Arid Environ. 30, 385–395. https://doi.org/10.1006/jare.1995.0031.

Mulpuri, S., Muddanuru, T., Francis, G., 2013. Start codon targeted (SCoT) polymorphism in toxic and non-toxic accessions of Jatropha curcas L. and development of a codominant SCAR marker. Plant Sci. 207, 117–127. https://doi.org/10.1016/j/plantsci.2013.02.013.

Newmaster, S.G., Fazekas, A.J., Ragupathy, S., 2006. DNA barcoding in land plants: evaluation of rbcL in a multigene tiered approach. Can. J. Bot. 84, 335–341. https://doi.org/10.1139/b06-047.

Ojeda, D.L., Santos-Guerra, A., Oliva-Tejera, F., Jaen-Molina, R., Cajupe-Castells, J., Marrero-Rodriguez, A., Cronk, Q., 2014. DNA barcodes successfully identified Macaronesian Lotus (Leguminosae) species within early diverged lineages of Cape Verde and mainland Africa. AoB PLANTS 6, plu050–plu050. https://doi.org/10.1093/aobpla/plu050.

Omar, K.A., 2014. Evaluating the effectiveness of in-situ conservation on some endemic plant species in South Sinai, Egypt. Am J Life Sci 2, 164–175.

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

Pieroni, A., Giusti, M.E., de Pasquale, C., Lenzarini, C., Censorii, E., Gonzáles-Tejero, M.R., Sánchez-Rojas, C.P., Ramiro-Gutiérrez, J.M., Skoula, M., Johnson, C., Sanchez-Rojas, C.P., Ramiro-Gutiérrez, J.M., Skoula, M., Johnson, C., Muscolo, C., Minocha, S.C. (Eds.), Molecular Biology of Woody Plants: Volume 1, Forestry Sciences. Springer, Netherlands, Dordrecht, pp. 395–406. https://doi.org/10.1007/978-94-017-2311-4_16.

Shahlaei, A., Torabi, S., Khosroshahi, M., 2014. Efficacy of SCoT and ISSR markers in assessment of tomato (lycopersicum esculentum Mill.) genetic diversity. https://doi.org/10.12692/ijh/5.2.14-22

Shaltout, K.H., Ahmed, D.A., Shabana, H.A., 2015. Population structure and dynamics of the endemic species Phlomis aurea Decne in different habitats in southern Sinai Peninsula Egypt. Global Ecol. Conservat. 4, 505–515. https://doi.org/10.1016/j.gecco.2015.10.002.

Tackholm, V., 1974. Students’ flora of Egypt. Thompson, J.D., Higgins, D.G., Gibson, T.J., 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22, 4673–4680.

Urutu, C.M., Mihai, C.T., Stanciu, G.-D., Dodi, G., Alexa-Stratulat, T., Luca, A., Leon-Constantin, M.-M., Stefanescu, R., Bild, V., Melnic, S., Tamba, B.L., 2018. Medicinal Plants of the Family Lamiaceae in Pain Therapy: A Review [WWW Document]. Pain Res. Manage. https://doi.org/10.1155/2018/7801543.

Yesson, C., Bármenas, R.T., Hernandez, H.M., De La LUZ RUIZ-MAQUEDA, M., Prado, A., Rodriguez, V.M., 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. https://doi.org/10.1111/j.1755-0998.2011.03009.x

Zahran, M.A., Willis, A.J., 2009. The Sinai Peninsula, in: Zahran, M.A., Willis, A.J. (Eds.), The Vegetation of Egypt, Plant and Vegetation. Springer Netherlands, Dordrecht, pp. 213–249. https://doi.org/10.1007/978-1-4020-8756-1_5

Zhang, J., Xie, W., Wang, Y., Zhao, X., 2015. Potential of Start Codon Targeted (SCoT) Markers to Estimate Genetic Diversity and Relationships among Chinese Elymus sibiricus Accessions. Molecules 20, 5987–6001. https://doi.org/10.3390/molecules20045987.