Scientific Research Publishing

American Journal of Plant Sciences, 2018, 9, 2757-2774
http://www.scirp.org/journal/ajps
ISSN Online: 2158-2750
ISSN Print: 2158-2742

DNA Barcoding and Identification of Medicinal Plants in the Kingdom of Bahrain

Malabika Roy Pathak[1*], Abdulaziz A. M. Mohamed[1,2], Muhammad Farooq[1]

[1]Department of Life Sciences, Arabian Gulf University, Manama, Kingdom of Bahrain
[2]Ministry of Works and Municipalities Affairs and Urban Planning, Manama, Kingdom of Bahrain
Email: *malabikarp@agu.edu.bh, amabdulkareem@gmail.com

How to cite this paper: Pathak, M.R., Mohamedm A.A.M. and Farooq, M. (2018) DNA Barcoding and Identification of Medicinal Plants in the Kingdom of Bahrain. *American Journal of Plant Sciences*, **9**, 2757-2774.
https://doi.org/10.4236/ajps.2018.913200

Received: November 5, 2018
Accepted: December 23, 2018
Published: December 26, 2018

Abstract

Authentication of medicinally important plants is essential for increasingly demands of herbal remedies worldwide. DNA barcoding technology is currently gaining importance as a reliable tool for plant species identification, although one barcode gene is not enough in the exceptions. Short sequence diversity of standardized specific coding gene regions of *rbcLa* and *matK* of plastid genome together with noncoding ribosomal internal transcribed spacer 2 (ITS2) marker is used as barcode to compare and differentiate plant species. The success of obtaining sequences of the 29 analyzed plants distributed in 21 families using three different barcode genes *rbcLa*, *matK* and ITS2 were 97%, 79% and 75% respectively. Multiple sequence alignment confirmed the medicinal plants at species level by 89.28%, 86.32% and 60.86% obtained through *rbcLa*, ITS2 and *matK* barcodes sequences respectively. The genetic distance between sequence pairs (GD) and percentage identity (PI) is compared to analyze the plant identity at species level. The phylogenic trees constructed to show the relatedness and distance of the analyzed plants in the history of evolution by the analysis of richness of clades. The construction of DNA barcode library of desert medicinal plants is an introductory research arena in Kingdom of Bahrain in helping the routine identification of plants, and developing guidelines for detection of adulterants in herbal medicines as well as protection of biodiversity.

Keywords

DNA Barcode, Medicinal Plants, Phylogenetic Tree, Species Identification

1. Introduction

Medicinal plants are used since ancient time for the treatment and management

of human and animal diseases. The use of medicinal plants in traditional medicine as well as in modern drugs discovery has well documented to maintain world health and to treat chronic diseases [1] [2] [3]. According to WHO, in an estimation around 25% modern drugs are derived from medicinal plants either directly or indirectly and between 70% - 95% of the populations of developing countries and 42% - 80% population of developed countries including the USA and Europe, using traditional medicines as primary health care in different name according to regions [4]. Drug discovery from medicinal plant leads to various target diseases including cancer, HIV/AIDS, Alzheimer's, malaria, and pains etc. and around 60% of the antitumor and anticancer drugs have derived from natural products [5] [6]. Although, drug discovery of medicinal plants explored only a small fraction of the huge diversity of plant metabolism, while most of the secondary bi-products have huge contributions in human society [7].

The Kingdom of Bahrain, geographically consists of several scattered islands in the middle of the Arabian Gulf, prevails semi-desert to desert environment with year wide average day temperature 36˚C (14˚C to 48˚C), scanty of average rainfall (39 to 128 mm per year) and holds several hundred species of flora. The number of vascular plant species reaches 357 of which most of these plants are adapted to the hot, arid and semi-arid environment [8]. Among them, 25% plants are used as a medicinal herb by Bahrainis or by others residents in the Arabian Peninsula or neighboring countries [9]. The usefulness of regional plants around 70% is of native while others introduced through Bedouin culture of the Arab region. The practice is still strong in the rural Arab regions in the treatment of minor ailments with these plants including ulcers, pneumonia, stomach disorders, rheumatism, diabetes, renal problems, and bronchitis [10] [11]. Some of the medicinally important desert plants are multipurpose plants in the Kingdom of Bahrain and some of them are threatened and restricted in distribution due to rapid urbanization, climate change as well as unsustainable utilization of natural resources [12] [13].

In the recent time, the correct identification of medicinal plants is prerequisite for their safe use in new drug discovery. The traditional identification of plants by taxonomist needs collection of proper morphological data during their growing season with reproductive organs such as flowers and fruits, which are often difficult, time consuming, and mostly unavailable during field survey [14] [15]. Moreover, the traditional methods of using the medicinal plant to cure various diseases is common but due to lack of proper taxonomic identification, the herbal industry suffers for substitution and adulteration of medicinal herbs with closely related species [16]. Identification of biological samples using DNA barcoding is a novel method of species identification and study molecular evolution [15]. DNA barcoding is considered as a molecular and bioinformatics tool for species differentiation, identification and discovery of new species at molecular taxonomy level [17]. DNA barcodes are short, specific regions of DNA that can amplify and can be sequenced routinely using universal primers and the recovered standardized short sequence of DNA depicted as a unique identification

marker for species [18]. Gradually, study and comparison of unknown plant barcode sequences with the sequences of the Global DNA reference libraries help to identify unknown plant samples as well as helpful to evaluate, understand, preserve and utilize biodiversity in a widely presented way [19].

The searching of DNA barcode in a wide range of flowering land plants is delicate as well as challenging enough in comparison to animal barcode region of mitochondrial gene COI, which is not effective in plants [20]. Hybridization, lack of sequence polymorphism, low nucleotide substitution rate and frequent integration of gene flow between sister species are the critical barriers in the selection of universal barcoding gene region in land plants [17]. Moreover, the desert plants adopted to endure in tough conditions of environment and soil; as a result, they possess different survival characteristics and molecular diversity [21]. The success in identification of plant species based on DNA barcode analysis depends on the comprehensive database analysis, otherwise missing and cryptic species cannot be identified, if identified there is a risk of the tested sample to a wrong species [22]. In addition, the presence of shared haplotype between closely related species (having identical DNA sequences) is facing limitations of the technique [18]. Barcoding of vascular plants was mostly focused on markers of chloroplast genes, several markers were tested and with time most commonly used combinations are *rbcL*, *matK*, *trnH-psbA*, with a nuclear internal transcribed spacer (ITS2) established [18] [19] [23] [24].

The survival characteristics of the desert medicinal plants due to the harsh arid environment as well as the archipelago nature (natural as well as artificial) of the Kingdom of Bahrain restricted plant distribution over the time of urbanization [10]. Until recently, there is no report to work on DNA barcode of any plants in Kingdom of Bahrain for identification or any other purpose. We collected 29 medicinally important plant species of Kingdom of Bahrain from different location. The sampling purpose is the molecular identification of collected 29 medicinally important plant using DNA based barcode study. Here, our primary goal is development, comparison and selection of the best DNA barcode marker of the desert medicinal plants in the Kingdom of Bahrain using universally accepted marker genes *rbcLa*, *matK and* ITS2. Moreover, we want to evaluate the taxonomic authentication of species by the barcode sequence analysis. The comparison of barcode DNA sequences of the local medicinally important plant at the genus or species level is based on basic local alignment search tool (BLAST) and global multiple sequence alignment (MSA) using GenBank accessions. The evaluation of sequences of the three different markers for the identification of the desert medicinal plant is helpful to choose the effective marker for future study. We also studied the barcode gap of the analyzed plants at interspecific to investigate phylogenetic relationship. DNA barcode library of desert medicinal plants in the Kingdom of Bahrain is an introductory and important research arena in Kingdom of Bahrain. Additionally, the DNA barcode sequences of desert medicinal plants in the global sequence library (GenBank) would be

useful for plant identification nationally as well as globally.

2. Materials and Methods

2.1. Study Area

Collection of local medicinal plants of Kingdom of Bahrain covered eight different regions as shown in Figure 1.

2.2. Plant Collection, Voucher Preparation and Identification

Plant samples were collected during their vegetative and reproductive growth stage in the two successive year 2016-2017 (January 2016 to April 2016, December 2016-April 2017) by several field trips in the different area, as the most of the annual desert plants, start growing after winter rainfall and perennial plants enter reproductive phase. Following the methodology of Barcode of Life Database (BOLD), whole plants with root system (if small) and part of the plant with branches, leaves, flowers showing maximum morphological characteristics useful for identification were collected for voucher preparation and plant identification. Plants were collected with all detail information of locality, plant habitat, road number, GPS coordinate, elevation level, field photograph, collector name, date, time etc.

Figure 1. Collection sites of medicinal plants for DNA barcode study of medicinal plants in Kingdom of Bahrain (Map Ref: http://www.freerunsca.org/).

Plant vouchers were prepared with the proper sample ID after collection. Plant parts cleaned, and dried using spacers between layers of blotting papers and finally pasted on the Herbarium sheet with proper labelling. Field/Herbarium photograph taken and uploaded in the BOLD specimen submission portal and the records of each plant with pictures are traceable online in Plant Taxomony portal in BOLD system (http://www.boldsystems.org). Plants identified and botanical names assigned based on reference books of Flora of Bahrain by ourselves. In some cases, we resolved the identification problem by consulting with taxonomists. Plants identified using standard identification method by comparing plant habitat, nodal characteristics of stem, leaf shape, leaf type, leaf arrangement, floral type, floral appearance, flower shape (sepal and petal), fruit characteristics etc.

2.3. Tissue Sample Preparation

Young leaves from the collected plant samples used for sample tissue preparation. Clean leaves properly dried either room temperature (thin and small) or heat dried at 37˚C (in case of succulents and thick leaves of desert plants). The leaves stored in sealable plastic packs and kept at room temperature [25].

2.4. Tissue Sub-Sample Preparation

As DNA was extracted using the plate based method, plant tissue sample of particular ID, around 2 - 3 pieces of plant leaves (0.5 cm) were subsampled in 96-well tube strips placed in plant box following the specific guideline of BOLD systems.

2.5. DNA Extraction

DNA extraction, PCR and sequencing work performed at Canadian Center for DNA Barcoding (CCDB), Canada using their standardized protocols. DNA extracted following the Glass fiber plate DNA extraction protocol, using a small amount of dry leaf samples [26] [27]. The leaf tissues were homogenized into fine powder using Tissue Lyser (Qiagen, USA) with rotated rack adopter at 28 Hz for 30 seconds by two times. DNA was extracted using 2X CTAB (250 µl) buffer and by incubating at 65˚C for 90 min. Cell lysates (50 µl) were transferred into 96-well Eppendrpf plate and Plant Binding Buffer (PBB, 100 µl) was added and incubated for 5 min at RT. 96-well Glass Fiber plate (PALL1) used and semi-automated glass fiber filtration method followed, while the corresponding DNA was bound to GF membrane [25]. GF Membrane rinsed by using Protein Wash Buffer (PWB) and the DNA eluted from GF plate to the collection microplate by adding pre-warmed double distilled water (60 µl), covered with cap strip and stored for PCR.

2.6. PCR and Sequencing

Three gene regions (*rbcLa*, *matK*, ITS2) were amplified using CCDB plant protocol [25] [26]. Amplification of the different plant markers required different

primers and PCR recipes in PCR plates shown in Table 1 and Table 2. PCR plate placed into the thermo-cycling block and different PCR thermocycle programs were used for different primer set of three different barcode genes. Amplification of *rbcLa*: 94˚C for 4 min; 35 cycles of 94˚C for 30 sec, 55˚C for 30 sec; 72˚C for 1 min; final extension 72˚C for 10 min; hold at 4˚C. For ITS2: 94˚C for 5 min; 35 cycles of 94˚C for 30 sec, 56˚C for 30 sec, 72˚C for 45 sec; final extension 72˚C for 10 min; hold at 4˚C. Amplification of *matK*: 98˚C for 45 s; 35 cycles of 98˚C for 10 s, 54˚C for 30 s, 72˚C for 40 s; final extension 72˚C for 10 min; hold at 4˚C. PCR products visualized and analyzed on 2% Agarose E-gel ^{R} 96 system (Invitogen) using recommended program and software [27]. Strong amplification of *rbc*L and ITS2 were obtained using low concentration of primers, dNTPs and Taq polymerase with one primer set *rbcLa-F* [28], *rbcLa-R* [29] and another primer set IITS-S2F [30] ITS4 [31]. In case of *matK* amplification, higher concentration of forward primer *matK*-xF [32] and reverse primer *matK*-MALPR1 [33], dNTPs and Taq polymerase were used (Table 2). According to CCDB protocol, diluted PCR replicons used directly for sequencing [26]. PCR products were sequenced on an ABI 3730Xl DNA analyzer (Applied Biosystem, California, USA) following standard procedure. For bidirectional sequencing of *rbcLa* and ITS2, same PCR primer set used separately for bidirectional sequencing at 96˚C for 2 min; 30 cycles of 96˚C for 30 s, 55˚C for 15 s, 60˚C for 4 min; hold at 4˚C. For bidirectional sequencing of *matK*, three different primers used separately depending on PCR products used for sequencing at 96˚C for 2 min; 30 cycles of 96˚C for 30 s, 50˚C for 15 s, 60˚C for 4 min; hold at 4˚C.

2.7. Sequence Editing, Aligning, Assembly and Data Analysis

Chromatographs of all the three markers of those plants recovered, trimmed, edited and aligned by using Codon Code Aligner version 3.7.1-6.0.2 (CodonCode

Table 1. List of PCR primers and sequencing primers used in the DNA barcoding and identification of medicinal plants in Kingdom of Bahrain presented.

Barcode genes	PCR Primers	Primer sequence 5 '-3 '	Annealing Temp for (PCR)	Sequencing Primers	References
rbcL	*rbcLa*-F (forward)	ATGTCACCACAAACAGAGACTAAAGC	55˚C	1) *rbcLa*-F*	[41]
rbcL	*rbcLa*-R (reverse)	GTAAAATCAAGTCCACCRCG		2) *rbcLa*-R*	[37]
ITS2	ITS-S2F (forward)	ATGCGATACTTGGTGTGAAT	56˚C	3) ITS-S2F*	[32]
ITS2	ITS4 (reverse)	TCCTCCGCTTATTGATATGC		4) ITS4*	[42]
matK	*matK*-xF (forward)	TAATTTACGATCAATTCATTC	54˚C	5) *matK*-xF 6) *matK*-MALPR1 (choice I)*	[28]
matk	*matK*-MALPR1 (reverse)	ACAAGAAAGTCGAAGTAT		7) *matK*-1RKIM-f 6) *matK*-MALPR1 (choice II)*	[43]
matK	-	ACCCAGTCCATCTGGAAATCTTGGTTC		7) *matK*-1RKIM-f*	[39]

*Bidirectional Sequencing.

Table 2. Combinations and concentrations of reagents used for PCR reactions of *rbcL*, *matK* and ITS2 for analysis of medicinal plants of Kingdom of Bahrain presented.

Reagents	*rbcL*	*matK* Volume per reaction (µl)	ITS2
10% Trehalose	6.25	---	6.25
20% Trehalose	----	1.875	---
ddH_2O	2.00	2.60	2.00
10X Buffer	1.25	0.75	1.25
50 mM $MgCl_2$	0.625	0.225	0.625
10 µM Primer Forward	0.125	0.375	0.125
10 µM Primer Reverse	0.125	0.375	0.125
10 mM dNTPs	0.0625	0.15	0.0625
*DNA Polymerse (5U/µl)	0.06	0.15	0.06
DNA Template (20 - 40 ng/µl)	2.00	1.0	2.00
Total Reaction volume	**12.5**	**7.5**	**12.5**

*Platimum DNA Polymerase (Invitrogen, Carlsbad, California, USA) used for PCR reactions.

Co., Massachusetts, USA) by CCDB and received DNA sequences for further analysis and study. The nucleotide sequences of different plants aligned using Basic Local Alignment Search Tool (BLAST) algorithm provided by National Centre for Biotechnology Information (NCBI) and European Bioinformatics Institute (EMBL-EBI). The level of similarity between the samples studied by BLAST algorithm using blastclust (http://www.ncbi.nih.gov), which automatically and systematically clusters the nucleotide and protein sequences based on pairwise distance. It is a query method of genus and species of the samples from which sequences separated by the smallest genetic distances in the matrix [18]. The FASTA files of barcode sequences of *rbcL*, *matk* and ITS2 prepared for comparison with reference data sites. Different BLAST matching tools used widely to optimize the molecular data mining and for the identification of plants at genus as well as species level. This comparison brought out the genus identification and in some cases species identification. FASTA files of nucleotide sequences explored MUSCLE program [34]. The nucleotide sequences were aligned; analyzed and phylogenetic tree was constructed using various algorithm programs in Molecular Evolutionary Genetics Analysis (MEGA 7) software [35]. Sequence editing and alignment were generated by MUSCLE [34] to construct phylogenetic tree and evolutionary distance measurement by Kimura 2-parameter of studied plants [36]. The nucleotide sequences of coding region of chloroplast gene *rbcLa*, *matK* and nuclear ITS2 regions deposited in GenBank (**Table 4**).

3. Results and Discussion

3.1. PCR and Sequencing Success

The success of sequencing of barcode markers evaluated based on the number of

plant species that successfully generated a sequence for a particular marker. The sequence recovery and analysis of three DNA barcodes of 29 medicinal plants collected from different parts of Kingdom of Bahrain showed (**Figure 1**). DNA barcodes genes of *rbcLa*, *matK* and ITS2 of 29 different medicinal plants representing 21 different families analyzed. The primers of *rbcLa*, *matK*, and ITS2 are used for PCR amplification, sequencing (**Table 1**) and PCR reagents (**Table 2**). For PCR amplification and sequencing of *rbcLa* and ITS2 same primers worked but in bidirectional sequencing of *matK* PCR products, 70% PCR products sequenced by choice I combination while 30% sequencing worked choice II (**Table 1**). The detail sequencing results of 29 medicinal plants of 21 different families using *rbcLa*, *matK* and ITS2 represented (**Table 3**). The success of sequence obtained of the 29 plants of three different barcodes genes *rbcLa*, *matK* and ITS2 were 97%, 79% and 75% respectively (**Table 3**). In bidirectional sequencing for *rbcLa*, forward primer gave high quality sequencing result in 92% plants while reverse primer gave 78%. In case of *Limonium axillare* in Plumbaginaceae and *Mesembryanthemum nodiflorum* of Aizoaceae, we got moderate quality sequencing product from forward primer and no product from reverse primer. High quality sequencing in *matK* and ITS2 obtained by 68% and 72% plants respectively using forward primer and 65% and 68% using reverse primer respectively. The high sequencing success rate of *rbcLa* (90%) than *matK* (70%) and ITS2 (41%) was observed in Amazonian trees while eight different DNA markers were tested [14]. In another DNA barcode study to build a community phylogeny of tropical trees, sequencing success was higher in *rbcL* region (90%) in comparing *matK* (68%) [37]. Sequencing success of *rbcL* (95%), *matK* (76%) and ITS2 (89%) from fresh specimen of vascular plants from Churchill was reported [38]. Similarly, the low success rate of *matK* in compare to *rbcL* reported in the speedy assessment of species abundance in taxonomically poorly known area or in cryptic population [39], while, the sequence success of 88% and 90% in case of *matK* region is reported respectively by [20] [23]. The sequencing success of *matK* around 85% observed in standard and nested multiplex-tandem PCR [32]. In general, the lower success of *matK* sequence recovery also reflected due to difficulties in primer selection and binding of primers as explained [38] [39]. The low success of in sequence recovery in some genera may be relatively thick leaves of desert plants, slower desiccation and consequent DNA degradation [38]. It has been noted that *rbcLa* is much easier to sequence than *matK* and others, so the selection of *rbcLa* is an important option in barcoding study [40]. In several comparative studies, the variation in the success of sequencing results of different DNA may depend on the nature of plants, their habitat and ecological behavior [41]. The sequences of three different barcode genes submitted in GenBank (**Table 4**).

3.2. Identification and Comparison of Barcodes

The performance of the three-barcode markers for identification of plant genus and species level compared with Global multiple sequence alignment (MSA) and

Table 3. List of bidirectional sequence results of *rbcL*, *matK* and ITS2.

No.	Sample ID	Plant Name	Family	Barcode Regions		
				rbcL	ITS2	*matK*
1	BAHMP001110117	*Launaea nudicaulis*	Asteraceae	HQ (1,2)	HQ (3,4)	HQ (5,6)
2	BAHMP003110117	*Alhagi graecorum*	Fabaceae	HQ (1,2)	HQ (3,4)	HQ (5,6)
3	BAHMP004110117R	*Cressa cretica*	Convolvulaceae	HQ (1,2)	-	HQ (7,6)
4	BAHMP007280316R	*Lycium shawii*	Solanaceae	HQ (1,2)	-	HQ (7,6)
5	BAHMP008280316H	*Anastatica hierochuntica*	Brassicaceae	HQ (1,2)	HQ (3,4)	-
6	BAHMP011280316R	*Francoeuria undulata*	Asteraceae	HQ (1), MQ (2)	HQ (3,4)	HQ (7,6)
7	BAHMP013280316H	*Limonium axillare*	Plumbaginaceae	MQ (1)	-	-
8	BAHMP014241216R	*Cynanchum varians*	Apocynaceae	HQ (1,2)	HQ (3,4)	HQ (7,6)
9	BAHMP015241216B	*Malva parviflora*	Malvaceae	HQ (1,2)	HQ (3,4)	-
10	BAHMP018090416	*Herniaria hemistemon*	Caryophyllaceae	HQ (1,2)	HQ (3,4)	HQ(6)
11	BAHMP019311216H	*Teucrium polium*	Lamiaceae	HQ (1,2)	MQ (3)	HQ (7,6)
12	BAHMP020040117R	*Euphorbia serpens*	Euphorbiaceae	HQ (1), MQ (2)	HQ (3,4)	MQ (7,6)
13	BAHMP025241216R	*Andrachne telephioides*	Phyllanthaceae	HQ (1,2)	HQ (3,4)	HQ (7,6)
14	BAHMP026311216R	*Savignya parviflora*	Brassicaceae	HQ (1,2)	HQ (3,4)	MQ (6)
15	BAHMP027311216H	*Senecio glaucus*	Asteraceae	HQ (1,2)	HQ (3,4)	HQ (7,6)
16	BAHMP029311216R	*Dipcadi erythraeum*	Asparagaceae	HQ (1,2)	-	HQ (7,6)
17	BAHMP031040117	*Tribulus terrestris*	Zygophyllaceae	HQ (1,2)	HQ (3,4)	HQ (5,6)
18	BAHMP035140117R	*Spergularia marina*	Caryophyllaceae	HQ (1,2)	HQ (3,4)	-
19	BAHMP036160117	*Convolvulus arvensis*	Convolvulaceae	HQ (1,2)	HQ (3,4)	HQ (5,6)
20	BAHMP037110117R	*Sesuvium portulacastrum*	Aizoaceae	HQ (1,2)	-	HQ (7), MQ (6)
21	BAHMP039200117	*Cynomorium coccineum*	Cynomoriaceae	LQ (1,2)	HQ (3,4)	-
22	BAHMP043040217R	*Cistanche tubulosa*	Orobanchaceae	-	HQ (3), MQ (4)	MQ (7,6)
23	BAHMP044050217H	*Frankenia pulverulenta*	Frankeniaceae	HQ (1,2)	HQ (3,4)	HQ (7,6)
24	BAHMP045050217	*Phyla nodiflora*	Verbenaceae	HQ (1,2)	HQ (3,4)	HQ(5,6)
25	BAHMP050250317	*Ziziphus spina-christi*	Rhamnaceae	HQ (1,2)	HQ (3,4)	HQ (7,6)
26	BAHMP057250317	*Erodium laciniatum*	Geraniaceae	HQ (1,2)	HQ (3,4)	-
27	BAHMP065120417	*Prosopis juliflora*	Fabaceae	HQ (1,2)	-	HQ (7,6)
28	BAHMP068120417	*Mesembryanthemum nodiflorum*	Aizoaceae	MQ (1), HQ (2)	-	MQ (7), HQ (6)
29	BAHMP069120417	*Leptadenia pyrotechnica*	Apocynaceae	HQ (1,2)	HQ (3,4)	HQ (7)

Primers used for sequencing were 1) *rbcLa*-F, 2) *rbcLa*-R, 3) ITS-S2F, 4) ITS4, 5) *matK*-xF, 6) *matK*-MALPR1, 7) *matK*-1RKIM-f and level of sequencing results mentioned as HQ: high quality sequencing and MQ: moderate quality sequencing.

presented (Table 5). The obtained sequence of the three barcode genes, *rbcLa*, *matK* and ITS2 analyzed following BLAST to confirm the taxonomic identification of the plants at the molecular level, using online database sites of International Nucleotide Sequence Database Collaboration (INSDC). The genetic distance (GD) between sequence pair and percent identity value (PI) after global multiple sequence alignment is helpful for comparative study, species identification and to score relatedness and distance. BLAST search data is helpful for

Table 4. List of submitted Genbank Accession Numbers of *rbcLa*, ITS2 and *matK* nucleotide sequences presented with plant Sample ID and plant name.

No.	Sample ID	Plant Name	GenBank Accession numbers		
			rbcLa	ITS2	*matK*
1	BAHMP001110117	*Launaea nudicaulis*	MH093899	MH191253	MH168726
2	BAHMP003110117	*Alhagi graecorum*	MH107147	MH191258	MH168727
3	BAHMP004110117R	*Cressa cretica*	MH115436	-	MH168728
4	BAHMP007280316R	*Lycium shawii*	MH115437	-	MH168729
5	BAHMP008280316H	*Anastatica hierochuntica*	MH115438	MH191259	-
6	BAHMP011280316R	*Francoeuria undulata*	MH115439	MH191260	MH168730
7	BAHMP013280316H	*Limonium axillare*	MH115440	-	-
8	BAHMP014241216R	*Chyancum varians*	MH168725	MH191261	MH211036
9	BAHMP015241216B	*Malva parviflora*	MH115441	MH203147	-
10	BAHMP018090416	*Herniaria hemistemon*	MH115442	MH203148	MH211035
11	BAHMP019311216H	*Teucrium polium*	MH115443	MH203149	MH211034
12	BAHMP020040117R	*Euphorbia serpens*	MH115444	MH203150	MH211042
13	BAHMP025241216R	*Andrachne telephioides*	MH115445	MH201309	MH211043
14	BAHMP026311216R	*Savignya parviflora*	MH115446	MH203151	MH211041
15	BAHMP027311216H	*Senecio glaucus*	MH115447	MH203152	MH211040
16	BAHMP029311216R	*Dipcadi erythraeum*	MH133120	-	MH211039
17	BAHMP031040117	*Tribulus terrestris*	MH133121	MH203153	MH211038
18	BAHMP035140117R	*Spergularia marina*	MH133122	MH203154	-
19	BAHMP036160117	*Convolvulus arvensis*	MH133123	MH203155	MH211037
20	BAHMP037110117R	*Sesuvium portulacastrum*	MH133124	-	MH211044
21	BAHMP039200117	*Cynomorium coccineum*	MH168717	MH203156	-
22	BAHMP043040217R	*Cistanche tubulosa*	-	MH203157	MH211045
23	BAHMP044050217H	*Frankenia pulverulenta*	MH168718	MH203158	MH211046
24	BAHMP045050217	*Phyla nodiflora*	MH168719	MH203159	MH211047
25	BAHMP050250317	*Ziziphus spina-christi*	MH168720	-	MH211048
26	BAHMP050250317R	*Ziziphus spina-christi*	-	MH203160	-
27	BAHMP057250317	*Erodium laciniatum*	MH168721	MH203161	-
28	BAHMP065120417	*Prosopis juliflora*	MH168722	-	MH211049
29	BAHMP068120417	*Mesembryanthemum nodiflorum*	MH168723	-	MH211050
30	BAHMP069120417	*Leptadenia pyrotechnica*	MH168724	MH203162	MH211051

genus and species identification of plants at the molecular level, which is one of the main objectives of this study. This is helping in the process of plant taxonomical query in the regional and global accomplishment. Using NCBI database in BLAST analysis, 100%, 92.85% and 86.95% plant genus of the analyzed medicinal plants confirmed by obtained barcode sequences of ITS2, *rbcLa* and *matK* sequences respectively (**Table 5**). The use of *rbcLa* and ITS2 barcodes worked fine in compare to *matK*. Overall, 97% plants species were correctly

Table 5. The comparison of nucleotide sequences of three barcode genes of *rbcLa*, *matK* and ITS2 of medicinal plants in the Kingdom of Bahrain.

Identification level using MSA	Barcode genes			Plant Identification (%)
	rbcLa	*matK*	ITS2	
Plant genus	*rbcLa*			92.85
Plant genus		*matK*		86.95
Plant genus			ITS2	100
Plant species	*rbcLa*			89.28
Plant species		*matK*		60.86
Plant species			ITS2	86.32
Plant species	*rbcLa*+	*matK*+	ITS2	28.57
Plant species	*rbcLa*+	*matK*+		21.42
Plant species	*rbcLa*+		ITS2	28.57
Plant species		*matK*+	ITS2	0
Plant species	*rbcLa* only			10.71
Plant species			ITS2 only	10.71

The percentage of plant identification at species and genus level calculated on the number of identified plants (genus/species level) based on total plants considered for the BLAST study.

identified either anyone of the barcodes (*rbcLa*, *matK* or ITS2) using BLAST survey. 28.57% plant species were confirmed by the analysis of all the three barcode sequences (*rbcLa*, *matK* and ITS2). 21.42% plant species were confirmed by the analysis of sequences of *rbcLa* and *matK* while 28.57% plant species were confirmed by the analysis of sequences of *rbcLa* and ITS2 combinations. 10.71% plant species were identified only by *rbcLa* (*Limonium axillare*, *Herniaria hemistemon*, *Dipcadi erythraeum*) and only by ITS2 sequences separately (*Cynanchum variance*, *Cyanomorium coccineum*, *Cistanche coccineum*). Only one barcode gene specifically confirmed the species identification using GenBank BLAST tool of those plants. BLAST analysis confirmed 100% medicinal plant genus and 89.28%, 86.32% and 60.86% plant species of medicinal plants by *rbcLa*, ITS2 and *matK* barcodes respectively. Similarly using BLAST analysis of sequence in the genetic distance analysis (GD), 99% success rate of genus identification and 73% and 75% species identification for *rbcL* and *matK* respectively observed in the identification of African rainforest trees was reported [18]. Similarly, the selection of *rbcL* and *matK* as core plant barcode marker reported by several workers and they have universal discriminatory power [14] [17] [42]. For identification of medicinal plant, the importance ITS2 together with *rbcL*, *matK* reported and justified that ITS2 is a valuable DNA barcode gene for the identification of closely related species [16] [30] [43]. Moreover, DNA polymorphism is typically low in coding plastid genes such as *rbcL* and *matK* but is frequent in noncoding ITS2 region and is helpful for species identification [18].

3.3. Phylogeny Study

The comparative analysis of molecular sequence data is essentially important for reconstructing evolutionary history of plants. The discriminatory influence of three barcode genes was considered in pairwise distance matrix to construct phylogenetic tree using in MEGA7 software [34]. The evolutionary history was inferred using pairwise distances in Maximum Composite Likelihood (MCL) [35]. The phylogenetic relatedness and evolutionary history drown by means of neighbor-Join (NJ) method [44] using nucleotide sequence alignment of *rbcLa* and ITS2 (**Figure 2** and **Figure 3**). The evolutionary distance analysis conducted using Kimura 2-parameter [36] method of base substitution per site and mentioned

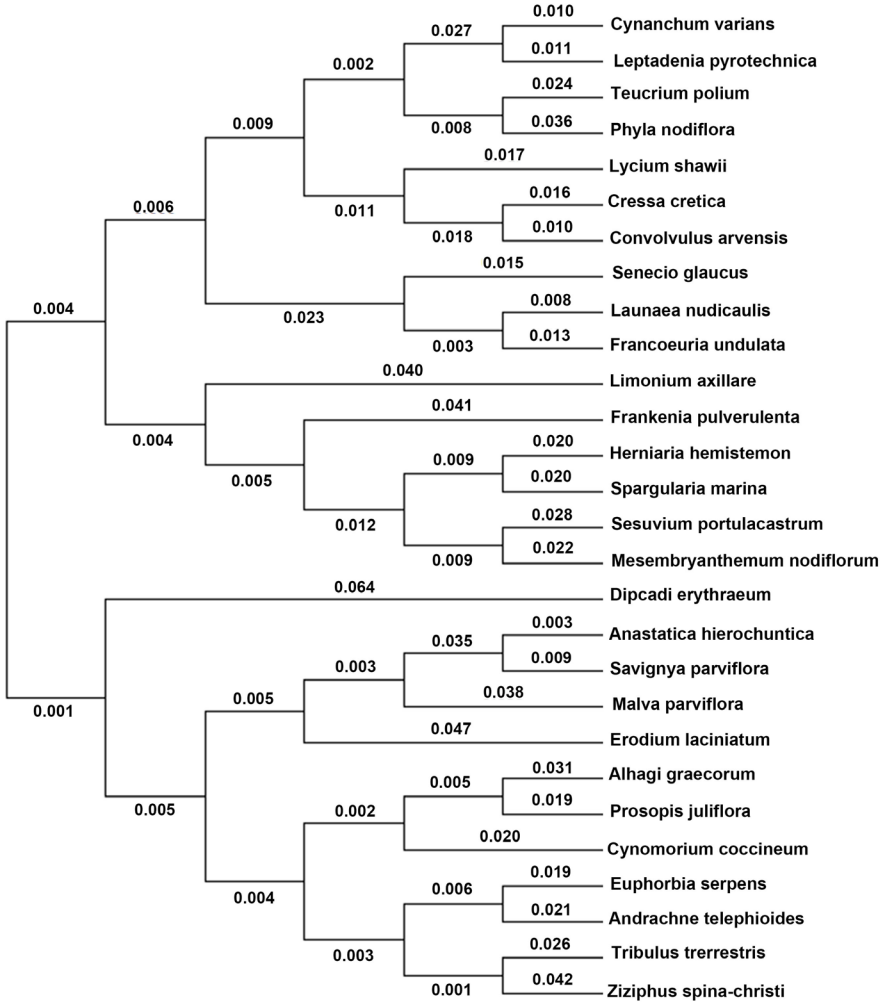

Figure 2. Molecular Phylogenetic affinity analysis of *rbcL* gene sequences of medicinal plants of Kingdom of Bahrain. The evolutionary history inferred using the Neighbor-Joining method [35]. The optimal tree with the sum of branch length = 0.88772833 is shown. The tree drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances computed using the Kimura 2-parameter method [36]and are in the units of the number of base substitutions per site. The analysis included 28 nucleotide sequences of 28 plants and evolutionary analyses were conducted in MEGA7 [34].

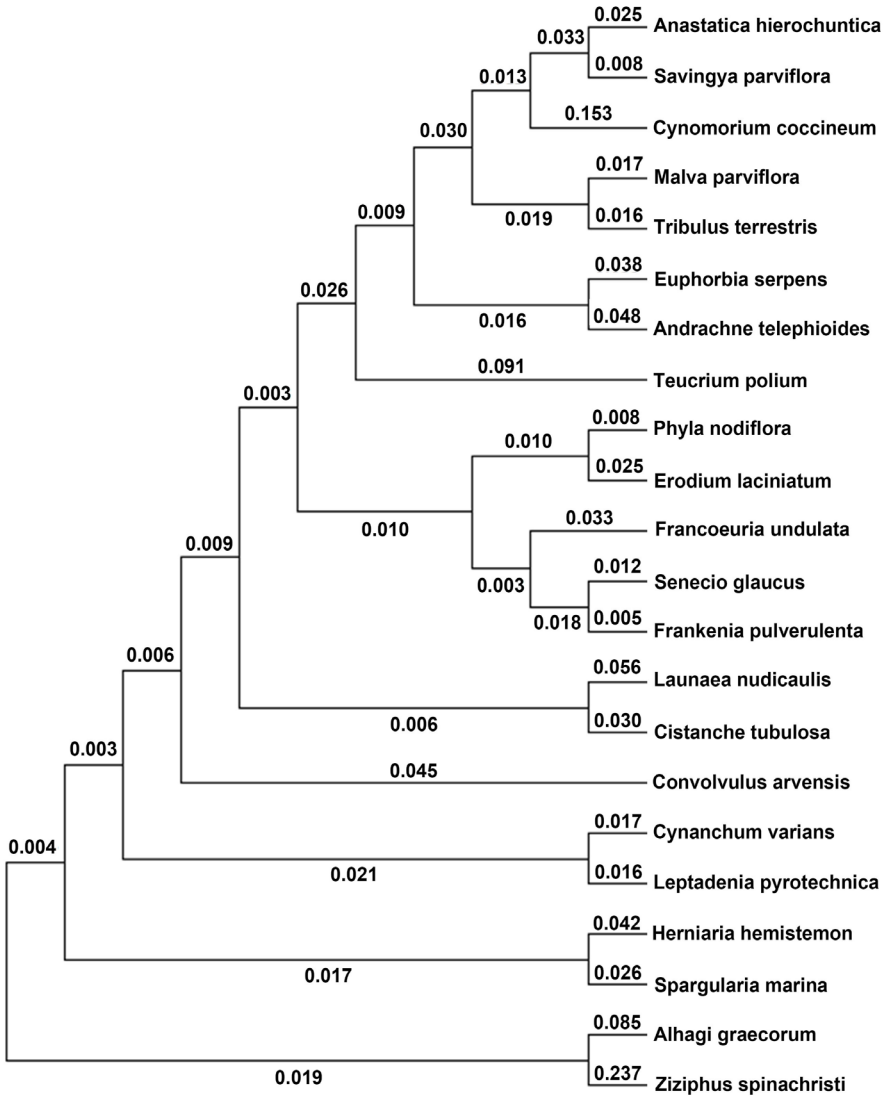

Figure 3. Molecular Phylogenetic affinity analysis of ITS2 sequences of medicinal plants of Kingdom of Bahrain. The evolutionary history inferred using the Neighbor-Joining method [35]. The optimal tree with the sum of branch length = 1.30627545 is shown. The tree drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances computed using the Kimura 2-parameter method [36] and are in the units of the number of base substitutions per site. The analysis included 22 nucleotide sequences of 22 plants and evolutionary analyses were conducted in MEGA7 [34].

by units. The pairwise distance matrix value indicates their closeness during evolution and ultimately helps to draw their phylogenetic tree and the history of evolution. The clades formed in the trees were mostly mixture of several plant species and the big and small branches showed their relatedness and distances clearly among the plant species. *Cynanchum varians* and *Leptadenia pyrotechnica* showed their closeness and common ancestry as they belong to the same family confirmed by taxonomic observation too. Similarly *Herniaria hemistemon* and *Spargularia marina* are closely related in their evolutionary history and

belong to the same family Caryophyllaceae. In a double-locus (*rbcLa* and ITS2) phylogenetic trees, tree nodes and close branches supported the taxonomic clarity, relatedness and clade phylogeny. For each species, the higher interspecific distance obtained by ITS2 (33.17%) in compare to *rbcLa* (8.72%), so the value of closeness differed in phylogenetic tree of ITS2 (**Figure 2**) and *rbcLa* (**Figure 3**). Phylogenetic tree analysis of ITS2 and *rbcL* provided species resolution in a better way by using pairwise distance matrix value and indicated their closeness and distance with each other. Similarly, in the building of community parsimony, *rbcL* and ITS2 sequence alignment in constructing phylogenetic tress nodes showed support value in plant phylogeny [18] [29]. We noticed that the *matK* region of chloroplast genome is little problematic in sequence retrieving, phylogenetic data analysis and identification of plants in comparing to *rbcLa* and ITS2 barcodes, similar to other studies [17] [32] [42]. In the present study, *rbcLa* and ITS2 demonstrated excellent reliability for species authentication and less genetic diversity, due to the limited dispersal capacity in the plant communities of the arid and semiarid regions in comparison to tropics where it is frequent factor. So, the DNA barcode analysis of those plants is of a little bit challenging for species differentiation [18]. We noticed that the *matK* region of chloroplast genome is little problematic in sequence retrieving, phylogenetic data analysis and identification of plants in comparing to *rbcLa* and ITS2 barcodes, similar to other studies [17] [32] [42].

In conclusion, primarily, *rbcLa* and ITS2 are very useful barcode region and can be of wider application in the study of desert plant identification, ecology, and plant diversity study. Moreover, our sequences submitted to the GenBank will be of very helpful data for future study in various aspects. As, DNA barcoding has potentials to transfigure the systems of taxonomists work by giving considerable momentum during this era as a helpful tool for plant identification at molecular level.

Acknowledgements

The research was supported by the Research Grant # "TS-LS-BIOT.011_2015-2017" from College of Graduate studies, Agricultural Biotechnology Program, Department of Life Sciences, Arabian Gulf University. We like to Acknowledge CCDB for DNA analysis and sequencing work.

Conflicts of Interest

The authors declare no conflict of interest.

References

[1] Willcox, M.L. and Bodeker, G. (2004) Traditional Herbal Medicines for Malaria. *British Medical Journal*, **329**, 1156-1159. https://doi.org/10.1136/bmj.329.7475.1156

[2] Gurib-Fakim, A. (2006) Medicinal Plants: Traditions of Yesterday and Drugs of Tomorrow. *Molecular Aspects of Medicine*, **27**, 1-93.

https://doi.org/10.1016/j.mam.2005.07.008

[3] Atanasov, A.G., Waltenberger, B., Pferschy-Wenzig, E., Linder T., Wawrosch, C., Uhrin, P., Temml, V., Wang, L., Schwaiger, S., Heiss, E.H., Rollinger, J.M., Schuster, D., Breuss, J.M., Bochkov, V., Mihovilovic, M.D., Kopp B., Bauer, R., Dirsch, V.M. and Stuppner, H. (2015) Discovery and Resupply of Pharmacologically Active Plant-Derived Natural Products: A Review. *Biotechnology Advances*, **33**, 1582-1614. |https://doi.org/10.1016/j.biotechadv.2015.08.001

[4] World Health Organization (2011) The World Traditional Medicines Situation, in Traditional Medicines: Global Situation, Issues and Challenges. 3rd Edition. http://digicollection.org/hss/en/m/abstract/Js18063en/

[5] Brower, V. (2008) Back to Nature: Extinction of Medicinal Plants Threatens Drug Discovery. *Journal of the National Cancer Institute*, **100**, 838-839. https://doi.org/10.1093/jnci/djn199

[6] Newman, D.J. and Cragg, G.M. (2012) Natural Products as Sources of New Drugs over the 30 Years from 1981 to 2010. *Journal of Natural Products*, **75**, 311-335. https://doi.org/10.1021/np200906s

[7] De Luca, V., Salim, V., Atsumi, S.M. and Yu, F. (2012) Mining the Biodiversity of Plants: A Revolution in the Making. *Science*, **336**, 1658-1661. https://doi.org/10.1126/science.1217410

[8] Public Commission for the Protection of Marine Resources, Environment and Wildlife (2006) Bahrain First National Report to the Convention on Biological Diversity. 1-17. http://www.biodiv.be/bahrain/implementation/national-reports-cbd/bahrain-first-national-report

[9] Jameel, A., Mandeel, Q. and Al-Laith, A. (2010a) Folk Medicine and the Use of Herbal Plants in the Kingdom of Bahrain. *Arab Gulf Journal of Scientific Research*, **28**, 91-104.

[10] Abbas, J.A., El-Oqlah, A.A. and Mahasneh, A.M. (1992) Herbal Plants in the Traditional Medicine of Bahrain. *Economic Botany*, **46**, 158-163. https://doi.org/10.1007/BF02930630

[11] Jameel, A., Al-Laith, A. and Qaher, M. (2010b) Knowledge, Attitudes and Practices regarding Medicinal Plants Used by the Indigenous People of Bahrain. *Arab Gulf Journal of Scientific Research*, **28**, 105-117

[12] Manal, A.S., Malabika, R.P., Ahmed, A.S., Mohammed, A. and Asma A. (2014) Somatic Embryogenesis and Organogenesis for Regeneration of Endangered Multipurpose Desert Plant *Leptadenia pyrotechnica* Forsk. Decne in the Kingdom of Bahrain. *American Journal of Plant Sciences*, **5**, 2342-2353. https://doi.org/10.4236/ajps.2014.515248

[13] Malabika, R.P. and Mohammad, S.A. (2014) The Role of Biotechnology in the Conservation of Biodiversity. *Journal of Experimental Biology and Agriculture Sciences*, **2**, 352-363.

[14] Gonzalez, M.A., Baraloto, C., Engel, J., Mori, S.A., Pétronelli, P., Riéra, B., Roger, A., Thébaud, C. and Chave, J. (2009) Identification of Amazonian Trees with DNA barcodes. *PLoS ONE*, **4**, e7483. https://doi.org/10.1371/journal.pone.0007483

[15] Hebert, P.D.N., Cywinska, A., Ball, S.L. and deWaard, J.R. (2003) Biological Identifications through DNA Barcodes. *Proceedings of the Royal Society B: Biological Sciences*, **270**, 313-321. https://doi.org/10.1098/rspb.2002.2218

[16] Techen, N., Parveen, I., Pan, Z. and Khan, I.A. (2014) DNA Barcoding of Medicinal Plant Material for Identification. *Current Opinion in Biotechnology*, **25**, 103-110.

https://doi.org/10.1016/j.copbio.2013.09.010

[17] Hollingsworth, P.M., Graham, S.W. and Little, D.P. (2011) Choosing and Using a Plant DNA Barcode. *PLoS ONE*, **6**, e19254. https://doi.org/10.1371/journal.pone.0019254

[18] 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. https://doi.org/10.1371/journal.pone.0054921

[19] Kress, W.J., García-Robledo, C., Uriarte, M. and Erickson, D.L. (2015) DNA Barcodes for Ecology, Evolution, and Conservation. *Trends in Ecology & Evolution*, **30**, 25-35. https://doi.org/10.1016/j.tree.2014.10.008

[20] Fazekas, A.J., Burgess, K.S., Kesanakurti, P.R., Graham, S.W., Newmaster, S.G., Husband, B.C., Percy, D.M., Hajibabaei, M. and Barrett, S.C.H. (2008) Multiple Multilocus DNA Barcodes from the Plastid Genome Discriminate Plant Species Equally Well. *PLoS ONE*, **3**, e2802. https://doi.org/10.1371/journal.pone.0002802

[21] Bidak, L.M., Kamal, S.A., Halmy, M.W.A. and Heneidy, S.Z. (2015) Goods and Services Provided by Native Plants in Desert Ecosystems: Examples from the Northwestern Coastal Desert of Egypt. *Global Ecology and Conservation*, **3**, 433-447. https://doi.org/10.1016/j.gecco.2015.02.001

[22] Liu, J., Möller, M., Gao, L., Zhang, D. and Li, D. (2011) DNA Barcoding for the Discrimination of Eurasian Yews (*Taxus L.*, Taxaceae) and the Discovery of Cryptic Species. *Molecular Ecology Resources*, **11**, 89-100. https://doi.org/10.1111/j.1755-0998.2010.02907.x

[23] Hollingsworth, M.L., Andra, Clark, A., 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**, 439-457. https://doi.org/10.1111/j.1755-0998.2008.02439.x

[24] CBOL Plant Working Group (2009) A DNA Barcode for Land Plants. *Proceedings of the National Academy of Sciences* USA, **106**, 12794-12797. https://doi.org/10.1073/pnas.0905845106

[25] Fazekas, A.J., Kuzmina, M.L., Newmaster, S.G. and Hollingsworth, P.M. (2012) DNA Barcoding Methods for Land Plants. *Methods in Molecular Biology*, **858**, 223-252. https://doi.org/10.1007/978-1-61779-591-6_11

[26] Kuzmina, M.L. and Ivanova, N.V. (2016) PCR Amplification for Plants and Fungi. CCDB Protocols. http://ccdb.ca/site/wpcontent/uploads/2016/09/CCDB_Amplification-Plants.pdf

[27] Ivanova, N.V. and Kuzmina, M.L. (2013) Protocols for Dry DNA Storage and Shipment at Room Temperature. *Molecular Ecology Resources*, **13**, 890-898. https://doi.org/10.1111/1755-0998.12134

[28] Levin, R.A., Wagner, W.L., Hoch, P.C., Nepokroeff, M., Pires, J.C., Zimmer, E.A. and Sytsma, K.J. (2003) Family-Level Relationships of Onagraceae Based on Chloroplast rbcL and ndhF Data. *American Journal of Botany*, **90**, 107-115. https://doi.org/10.3732/ajb.90.1.107

[29] Kress, W.J., Erickson, D.L., Jones, F.A., Swenson, N.G., Perez, R., Sanjur, O. and Bermingham, E. (2009) Plant DNA Barcodes and a Community Phylogeny of a Tropical Forest Dynamics Plot in Panama. *Proceedings of the National Academy of Sciences* USA, **106**, 18621-18626. https://doi.org/10.1073/pnas.0909820106

[30] Chen, S., Yao, H., Han, J., Liu, C., Song, J., Shi, L., Zhu, Y., Ma, X., Gao, T., Pang X.,

Luo, K., Li, Y., Li, X., Jia, X., Lin, Y. and Leon, C. (2010) Validation of the ITS2 Region as a Novel DNA Barcode for Identifying Medicinal Plant Species. *PLoS ONE*, **5**, e8613. https://doi.org/10.1371/journal.pone.0008613

[31] Dunning, L.T. and Savolainen, V. (2010) Broad-Scale Amplification of *matK* for DNA Barcoding Plants, a Technical Note. *Botanical Journal of the Linnean Society*, **164**, 1-9. https://doi.org/10.1111/j.1095-8339.2010.01071.x

[32] Ford, C.S., Ayres, K.L., Toomey, N., Haider, N., Van, Alphen, Stahl, J., Kelly, L.J., Wikström, N., Hollingsworth, P.M., Duff, R.J., Hoot, S.B., Cowan, R.S., Chase, M.W. and Wilkinson, M.J. (2009) Selection of Candidate Coding DNA Barcoding Regions for Use on Land Plants. *Botanical Journal of the Linnean Society*, **159**, 1-11. https://doi.org/10.1111/j.1095-8339.2008.00938.x

[33] White, T.J., Bruns, T.D., Lee, S.B. and Taylor, J.W. (1990) Amplification and Direct Sequencing of Fungal Ribosomal RNA Genes for Phylogenetics. *PCR Protocols: A Guide to Methods and Applications*, 315-322. https://doi.org/10.1016/B978-0-12-372180-8.50042-1

[34] Edgar, R.C. (2004) MUSCLE: Multiple Sequence Alignment with High Accuracy and High Throughput. *Nucleic Acids Research*, **32**, 1792-1797. https://doi.org/10.1093/nar/gkh340

[35] Kumar, S., Stecher, G. and Tamura, K. (2016) MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Molecular Biology and Evolution*, **33**, 1870-1874. https://doi.org/10.1093/molbev/msw054

[36] Kimura, M. (1980) A Simple Method for Estimating Evolutionary Rates of Base Substitutions through Comparative Studies of Nucleotide Sequences. *Journal of Molecular Evolution*, **16**, 111-120. https://doi.org/10.1007/BF01731581

[37] Kress, W.J., Erickson, D.L., Swenson, N.G., Thompson, J., Uriarte, M. and Zimmerman, J.K. (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**, e15409. https://doi.org/10.1371/journal.pone.0015409

[38] Kuzmina, M.L., Johnson, K.L., Barron, H.R. and Hebert, P.D. (2012) Identification of the Vascular Plants of Churchill, Manitoba, Using a DNA Barcode Library. *BMC Ecology*, **12**, 25. https://doi.org/10.1186/1472-6785-12-25

[39] Costion, C., Ford, A., Cross, H., Crayn, D. Harrington, M. and Lowe, A. (2011) Plant DNA Barcodes can Accurately Estimate Species Richness in Poorly Known Floras. *PLoS ONE*, **6**, e26841. https://doi.org/10.1371/journal.pone.0026841

[40] Taberlet, P., Coissac, E., Pompanon, F., Gielly, L., Miquel, C., Valentini, A., Vermat, T., Corthier, G., Brochmann, C. and Willerslev, E. (2007) Power and Limitations of the Chloroplast trnL (UAA) Intron for Plant DNA Barcoding. *Nucleic Acids Research*, **35**, e14. https://doi.org/10.1093/nar/gkl938

[41] Jurado-Rivera, J.A., Vogler, A.P., Reid, C.A.M., Petitpierre, E. and Gómez-Zurita, J. (2009) DNA Barcoding Insect-Host Plant Associations. *Proceedings Biological Sciences*, **276**, 639-648. https://doi.org/10.1098/rspb.2008.1264

[42] Kuzmina, M.L., Braukmann, T.W.A., Fazekas, A.J., Graham, S.W., Dewaard, S.L., Rodrigues, A., Bennett, B.A., Dickinson, T.A., Saarela, J.M., Catling, P.M., Newmaster, S.G., Percy, D.M., Fenneman, E., Lauron-Moreau, A., Ford, B., Gillespie, L., Subramanyam, R., Whitton, J., Jennings, L., Metsger, D., Warne, C.P., Brown, A., Sears, E., Dewaard, J.R., Zakharov, E.V. and Hebert, P.D.N. (2017) Using Herbarium-Derived DNAs to Assemble a Large-Scale DNA Barcode Library for the Vascular Plants of Canada. *Applications in Plant Sciences*, **5**, 12. https://doi.org/10.3732/apps.1700079

[43] Yao, H., Song, J., Liu, C., Luo, K., Han, J., Li, Y., Pang, X., Xu, H., Zhu, Y., Xiao, P. and Chen, S. (2010) Use of ITS2 Region as the Universal DNA Barcode for Plants and Animals. *PLoS ONE*, **5**, e13102. https://doi.org/10.1371/journal.pone.0013102

[44] Saitou, N. and Nei, M. (1987) The Neighbor-Joining Method: A New Method for Reconstructing Phylogenetic Trees. *Molecular Biology and Evolution*, **4**, 406-425.