Nucleotide based validation of the endangered plant *Diospyros mespiliformis* (Ebenaceae) by evaluating short sequence region of plastid *rbcL* gene

Abdullah Alaklabi1, Ibrahim A. Arif2,3, Sameera O. Bafeel4, Ahmad H. Alfarhan2,3, Anis Ahamed2,3, Jacob Thomas5 and Mohammad A. Bakir2,3*

1Department of Biology, College of Arts and Science, Al-Baha University (BU), Baljurashi, Saudi Arabia  
2Department of Botany and Microbiology, College of Science, King Saud University (KSU), Riyadh, Saudi Arabia  
3Saudi Biological Society and Prince Sultan Research Chair for Environment and Wildlife, King Saud University  
4Department of Biology, King Abdulaziz University (KAU), Jeddah, Saudi Arabia

*Corresponding author. Email: mbakir@ksu.edu.sa

Abstract

*Diospyros mespiliformis* (Hochst. ex A.DC.; Ebenaceae) is a large deciduous medicinal plant. This plant species is currently listed as endangered in Saudi Arabia. Molecular identification of this plant species based on short sequence regions (571 and 664 bp) of plastid *rbcL* (ribulose-1, 5-bisphosphate carboxylase) gene was investigated in this study. The endangered plant specimens were collected from Al-Baha, Saudi Arabia (GPS coordinate: 19.8543987, 41.3059349). Phylogenetic tree inferred from the *rbcL* gene sequences showed that this species is very closely related with *D. brandisiana*. Close relationship was also observed among *D. bejaudii*, *D. Philippinensis* and *D. releyi* (≥99.7% sequence homology). The partial *rbcL* gene sequence region (571 bp) that was amplified by *rbcL* primer-pair *rbcLaF*-*rbcLaR* failed to discriminate *D. mespiliformis* from the closely related plant species, *D. brandisiana*. In contrast, primer-pair *rbcL1F*-*rbcL724R* yielded longer amplicon, discriminated the species from *D. brandisiana* and demonstrated nucleotide variations in 3 different sites (645G>T; 663A>C; 710C>G). Although *D. mespiliformis* (EU980712) and *D. brandisiana* (EU980656) are very closely related species (99.4%); however, studied specimen showed 100% sequence homology with *D. mespiliformis* and 99.6% with *D. brandisiana*. The present findings showed that *rbcL* short sequence region (664 bp) of plastid *rbcL* gene, amplified by primer-pair *rbcL1F*-*rbcL724R*, can be used for authenticating samples of *D. mespiliformis* and may provide help in authentic identification and management process of this medicinally valuable endangered plant species.

Keywords: *Diospyros mespiliformis*; endangered plant; identification; partial *rbcL*.  
Abbreviations: BLAST_Basic Local Alignment Search Tool; BOLD_Barcode of Life Data; nt_nucleotide; R1_primer pair *rbcLaF*-*rbcLaR*; R2_primer pair *rbcL1F*-*rbcL724R*; *rbcL*_Ribulose-1,5-bisphosphate carboxylase/oxygenase

Received: 6 Oct 2013; Revised: 8 Jan 2014; Accepted: 16 Jan 2014.

Introduction

*Diospyros mespiliformis* Hochst. ex A. DC. (=*D. senegalensis* Perrot. ex A. DC.; known as Jackalberry or African Ebony) is a large deciduous tree. The plant *D. mespiliformis* belongs to the monophyletic family Ebenaceae which consists of two subfamilies Ebenoideae and Lissocarpoideae and four genera; *Euclea* *Diospyros*, *Lissocarpa* and *Royena* (Duangjai et al., 2006). Crude extract of the leaves of this plant contains alkaloid, tannins, saponins, glycosides, steroids, flavonoids and terpenoids. The leaf-extract of *D. mespiliformis* has significant activity against *Escherichia coli*, *Pseudomonas aeruginosa*, *Shigella* spp. and *Staphylococcus aureus* (Dangoggo et al., 2012). Diosquinone and plumbagin present in the root of *D. mespiliformis* have antibacterial activity against *E. coli*, *P. aeruginosa* and *S. aureus* (Lajubutu et al., 1995). However, this medicinally valued plant species is enlisted as endangered in Saudi Arabia (Colenette, 1998; Colenette, 1999). Authentic identification is crucial for sustainable management of endangered plants. The plastid-encoded *rbcL* gene sequence is the most frequently used for plant phylogenetic analyses. In a major attempt, 499 species of seed plants, representing all major taxonomic groups, were studied using *rbcL* gene sequences (Chase et al., 1993), phylogenetic relationships of plants under Dipsacales (Donoghue et al., 1993) and closely related genera of *Aegilops*, *Hordeum* and *Triticum* (Gielly and Taberlet, 1994) were determined using *rbcL* gene sequences. Recently, effectiveness of the core plant barcode regions (*rbcL* and *matK*) and a supplemental ribosomal DNA (ITS2) were examined using 900 specimens, representing 321 species of vascular plants. The sequencing success for *rbcL* gene is reportedly high for most plant species (Kuzmina et al., 2012). Geologically ancient DNA of *Hymenaea protera* (Fabaceae) from Miocene fossils was suspected as contaminant from *Arabidopsis thaliana*. Plastid *rbcL* gene sequences were used
Table 1. Pairwise rbcL gene sequence homology (%) of the specimen and closely related species.

<table>
<thead>
<tr>
<th>No.</th>
<th>Plant specimen/Species (GenBank accession numbers)</th>
<th>Nucleotide (nt) identities (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>R1 (HF568788)</td>
<td>520/520 nt (100%)</td>
</tr>
<tr>
<td>2.</td>
<td>R2 (HF568789)</td>
<td>568/568 nt (100%)</td>
</tr>
<tr>
<td>3.</td>
<td>D. mespiliformis (EU980712)</td>
<td>668/668 nt (100%)</td>
</tr>
<tr>
<td>4.</td>
<td>D. brandisiana (EU980656)</td>
<td>661/664 nt (99.6%)</td>
</tr>
</tbody>
</table>

**Fig 1.** Morphology of D. mespiliformis: (a) Tree in the sampling site. (b) Leaves; (c) Fruit.

Morphological features of D. mespiliformis are illustrated in Fig. 1. D. mespiliformis is a tall tree that can reach a height of 25m with a trunk circumference of more than 5m. It has a dense evergreen canopy. The bark is black to grey with rough texture. The fresh inner skin of the bark is reddish. Leaves are simple, alternate, leathery and dark green. Margin is smooth and new leaves are red in young plants. Flowers are creamy-coloured and bell-shaped. The male flowers are arranged in stalked bunches and female flowers are solitary. The fruit is a fleshy berry with an enlarged calyx, yellow to orange when ripe (SANBI, 2012).

**Phylogeny of D. mespiliformis and closely related species**

Application of primer-pairs R1 and R2 were successful for the PCR-amplification of the chloroplast rbcL gene region. We obtained shorter sequence length (571 bp) for the primer set R1 compared with that of the R2 (664 bp). Sequence obtained using the primer-pair R1 failed to distinguish D. mespiliformis from D. brandisiana (100% sequence homology (Table 1). In contrast, sequence region obtained by the primer-pair R2 showed 100% sequence similarity with D. mespiliformis and 99.6% with D. brandisiana, respectively (Table 1). Maximum likelihood tree showed both individually determined sequences (R1 and R2) grouped in the same clade with D. mespiliformis and D. brandisiana (62% bootstrap value) (Fig. 2). Other plant species under the genus Diospyros, for example D. argentea, D. bejaudii, D. philippinensis and D. ridleyi, seem to be very closely related as well and grouped in the same clade (83% bootstrap value) (Fig. 2).

**Discussion**

Methods of sequence comparison (BLAST, genetic distance and tree topology) have been used for a wide range of molecular
identification purposes (Ross et al., 2008; Kazmina et al., 2012; Syme et al., 2013). The chloroplast rbcL gene sequence (NC_000932) region in model plant Arabidopsis thaliana is 1440 bp. The section of rbcL which is considered as the barcoding region at the 5' end of the gene in the A. thaliana sequence is coordinated between bp 1-599 (27-579=553; excluding primer sequences) for the primer set R1 and bp 1-743 (21-723=703; excluding primer sequences) for the primer set R2. Both of these primers have shown good universality in barcoding studies (CBOL Plant Working Group, 2013). However, the second primer set R2 supposed to provide ±150 bp of extra sequences compared with that of the R1. Our attempt using the sequence section (571 bp) obtained by rbcL primer-set R1 failed to distinguish D. mespiliformis from D. brandisiana. However, sequence section (664 bp) obtained by rbcL primers-set R2 discriminated the species from D. brandisiana and demonstrated nucleotide variations in 3 different sites (645G>T; 663A>C; 710C>G) (Fig. 3). These findings were supported by the pairwise comparison of the nearly complete previously published genBank sequences of D. mespiliformis (EU980712) and D. brandisiana (EU980656) (Fig. 4). The plant species D. mespiliformis (EU980712) and D. brandisiana (EU980656) were observed very closely related (1449/1458 nt; 99.4%) species (Table 1) and difficult to discriminate based on rbcL gene sequence (Fig. 2). The studied specimen showed identical sequence homology with D. mespiliformis (Table 1; Fig. 5). Similar to our findings and apart from rbcL gene, molecular systematics of the family Ebenaceae based on six plastid regions (tpB, matK, ndhF, trnK intron, trnL intron and trnL-trnF spacer) also demonstrated very close relationship among Diospyros species (D. bejaudii, D. brandisiana, D. cauliflora, D. celebica, D. curranii, D. ferruginescens, D. mespiliformis, D. oblonga, D. philippinensis and D. ridleyi). A unique morphological feature like ruminate-endosperm may be the explanation for such close relationship and is a unifying character for D. mespiliformis and other five closely related species (D. bejaudii, D. celebica, D. ferruginescens, D. philippinensis and D. ridleyi; Duan gjai et al., 2006). Plastid rbcL is the most commonly sequenced gene for phylogenetic studies of plants (Schuettpelz et al., 2006) because of the other barcoding plant marker such as matK showed problematic for the PCR-amplification (CBOL Plant Working Group, 2009; Bafeel et al., 2011) and obtaining clean sequence (Hollingsworth, 2011; Yu et al., 2011). As shown in this study, despite the fact that application of plastid rbcL gene in plants is universally accepted, placement of taxa
Fig 4. Alignment of previously published NCBI D. brandisiana (EU980656) and D. mespiliformis (EU980712) rbcL gene sequences. Arrows and black dots indicate the nucleotide variation between the two sequences (645G>T; 663A>C; 710C>G).

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Codon</th>
<th>Amino Acid Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>D. brandisiana</td>
<td>650</td>
<td>G</td>
</tr>
<tr>
<td>D. mespiliformis</td>
<td>650</td>
<td>G</td>
</tr>
<tr>
<td>R2</td>
<td>650</td>
<td>G</td>
</tr>
</tbody>
</table>

Fig 5. Alignment of D. mespiliformis (EU980712) and R2 (this study; HF568789) rbcL gene sequences.

Fig 6. Map showing the location of sample collection site (GPS coordinate: 19.8543987, 41.3059349), Al-Baha, Saudi Arabia (Adapted from https://maps.google.com).
under a single species is difficult for some plants with very close relationships (Gadik and Quinn, 1993; Les et al., 1997; Chen et al., 1999; Zuccarello and Lokhorst, 2005). Similar to these findings, the short region of rbcL gene sequence (534 bp) between Morinda umbellata and M. reticulata (Rubiaceae) revealed three SNPs at sites 1, 22 and 51 and suggested to be used for authenticating of these plant species (Nair et al., 2013). Three nucleotides in the rbcL gene sequence were identified as DNA markers for the genetic identity of Coffea canephora and C. congestis (Nandhini et al., 2013).

Materials and Methods

Plant material

The specimens obtained from the trees located in Al-Baha, Saudi Arabia (GPS coordinate: 19.8543987, 41.3059349; Fig. 6). The plant species was identified on the basis of morphological characters by the curator, herbarium, KSU (KSU herbarium voucher specimen no.11729 and 21537). The leaf samples of the collected specimens were individually placed in plastic pouches and transported to the laboratory, where the specimens were stored at -80°C until processed for DNA extraction.

DNA extraction

The leaf specimen was crushed in liquid nitrogen using sterile mortar and pestle to get fine powder. DNA was extracted using DNeasy plant mini kit (Qiagen) and a QiAcube (Qiagen) DNA extraction instrument. Quality of the extracted DNA was determined using gel electrophoresis and Nanodrop 8000 Spectrophotometer (Thermo Scientific, Wilmington, USA). Isolated plant genomic DNA was preserved at -80°C.

PCR

Primers for the amplification of the chloroplast rbcL gene region included: R1 [rbcLaF (ATGTCACTACCAAGAGACGAGATTCAACCGGCTC) (Levin, 2003); rbcLaR (GTAAAATCAAGTCCACCRCG)] (Kress and Erickson, 2007)] and R2 [rbcLJF (ATGTCACTACCAAGACAGAAAC): rbcL724R (TCGCAATGATCCAGTAGC) (Fay et al. 1999)]. PCR reaction mixture (30 µL) contained the following: 15 µL of FideliTaq PCR Master Mix (USB Corporation, Cleveland, OH), giving a final concentration of 200 µM each deoxynucleotide and 1.5 mM MgCl₂, 1 µM (1.2 µL) each primer (Eurofins MWG Operon, Germany), 2 µL (±50 ng) of genomic DNA and the remaining volume was adjusted with sterile distilled water. PCR amplification was performed with a thermal cycler (Veriti, Applied Biosystems) as follows: 95°C for 1 min, followed by 35 cycles of 95°C for 30 s, 51°C for 30 s for primer pair R1 (48°C for 30 s for R2) and 68°C for 1 min, followed by an elongation step at 68°C for 5 min.

Agarose gel electrophoresis

Agarose gel (20 x 14 cm, 1% of 1x TAE buffer containing 0.5 µg/mL ethidium bromide) was used for electrophoresis of PCR-products. Gel images were obtained using Proxima C16 Phiv+(Isogen Life Science) UV transilluminator and Opticom (version 3.2.5; OptiGo) imaging system. The amplified PCR products were determined on gel for the presence or absence of bands. The size of PCR products were determined using a standard DNA ladder (Amersham 100 bp, GE Healthcare) in the gel and a band size detection software (TotalLab TL100 1D; version 2008.01).

Sequencing

Sequencing were conducted using the dideoxynucleotide chain-termination method with a DNA-sequencer (ABI PRISM 3130xl; Applied Biosystemsis/Hitachi) and a BigDye Terminator version 3.1 cycle sequencing RR-100 kit (Applied Biosystems), according to the manufacturer's instructions. Determined sequences were submitted to DDBJ/EMBL/GenBank database (Accession no. HF568788 and HF568789).

Assignment of taxa

BLAST (Basic Local Alignment Search Tool) and BOLD (Barcode of Life Data) searches were applied to the produced sequence using the available online databases. Relevant sequences with the specimen were retrieved from NCBI nucleotide database and aligned with ClustalX (Thompson et al., 1997). Phylogenetic analyses were conducted using the Maximum Likelihood (ML) method based on Tamura-Nei model (Tamura and Nei 1993) in MEGA5 (Tamura et al. 2011). Estimates of evolutionary divergence between sequences were conducted using the Maximum Composite Likelihood model (Tamura et al., 2004). The topologies of the phylogenetic trees were evaluated by using the bootstrap re-sampling method of Felsenstein (1985) with 1000 replicates. Pair-wise sequence comparisons of the closely related plant species were conducted using BLAST 2 Sequences (Tatusova and Madden, 1999).

Conclusion

This study demonstrated that very close relationship of D. mespiliformis with D. brandisiana and assignment of taxa based on only rbcL gene sequence seems to be difficult due to nearly identical sequences between these species. Use of the rbcL short sequence region (primer pair rbcLF- rbcL724R) and the variable nucleotides in the amplified region of D. mespiliformis can be used as DNA markers for authentic identification and sustainable management of this endangered plant.

References


Syme AE, Udovic F, Stajic V, Murphy DJ (2013) A test of sequence-matching algorithms for a DNA barcode database of invasive grasses. DNA Barcodes. 1: 19–26


