Pak. J. Bot., 56(5): DOI: http://dx.doi.org/10.30848/PJB2024-5(11)

EFFECTIVENESS OF DNA BARCODES (*RBCL*, *MATK*, ITS2) IN IDENTIFYING GENERA AND SPECIES IN CACTACEAE

LIHUA WEI[1], FLOR CRISTINA PACHECO-REYES[2], JOSÉ ÁNGEL VILLARREAL-QUINTANILLA[3], VALENTÍN ROBLEDO-TORRES[4], JUAN ANTONIO ENCINA-DOMÍNGUEZ[5], EDGAR EDUARDO LARA-RAMÍREZ[1] AND MIGUEL ÁNGEL PÉREZ-RODRÍGUEZ[3*]

[1]*Centro de Biotecnología Genómica, Instituto Politécnico Nacional, Blvd. del Maestro S/N Esq. Elías Piña. Col. Narciso Mendoza, Reynosa, Tamaulipas, 88710, México*
[2]*Universidad Autónoma Agraria Antonio Narro, Calzada Antonio Narro No. 1923, Saltillo, Coahuila, 25315, México*
[3]*Departamento de Botánica, Universidad Autónoma Agraria Antonio Narro, Calzada Antonio Narro No. 1923, Saltillo, Coahuila, 25315, México*
[4]*Departamento de Horticultura, Universidad Autónoma Agraria Antonio Narro, Calzada Antonio Narro No. 1923, Saltillo, Coahuila, 25315, México*
[5]*Departamento de Recursos Naturales, Universidad Autónoma Agraria Antonio Narro, Calzada Antonio Narro No. 1923, Saltillo, Coahuila, 25315, México*
[*]*Corresponding author's email: miguel_cbg@hotmail.com*

Abstract

In this study, we acquired 234 barcoding sequences (78 *rbcL*, 78 *matK*, 78 ITS2) representing 49 species of 28 genera in the family Cactaceae. Then we evaluated the discriminatory power of three DNA barcode regions (*rbcL*, *matK*, and ITS2) in species of the family Cactaceae using three methods, BLAST, genetic distance, and neighbour-joining trees. In BLAST analysis, the *matK* barcode region had the highest identification rate at the generic and species levels (83.33% and 50%, respectively). With the closest genetic distance method, the barcode combination groups had higher discrimination rates than any individual barcode region, of which the three barcodes' combination (*rbcL* + *matK* + ITS2) obtained the highest identification rate (52.56%). In neighbour-joining analysis, the phylogenetic tree generated with the ITS2 region had the highest discrimination rate (92.31%), and the best barcode combination group was the *rbcL* + *matK* (87.18%). An overlap between the intra- and interspecific distances was observed from DNA barcoding gap analysis. Our results showed that the *matK* region had a better efficiency in identification at the generic and species levels in the family Cactaceae. Additionally, we strongly recommend using the ITS2 as a DNA barcode region in this family due to its high classification ability in constructing phylogenetic trees.

Key words: BlAST, PCR, Neighbour-joining, Barcoding gap, Cactaceae.

Introduction

The Cactaceae family (Caryophyllales, Angiosperms), is distributed on the American continents, from southern South America to northern North America, covering different habitats from arid and semiarid regions to deciduous forests (Guerrero *et al*., 2019). This family consists of approximately 1800 species, of which at least 100 genera exist; Mexico has the greatest diversification of cactus species, with approximately 52 genera and 850 species, with an endemism that is close to 84% of the total species (Tenorio-Escandón *et al*., 2022). These species are mainly distributed in the Chihuahua and Sonoran Deserts, where there are 141 endemic species (Villarreal-Quintanilla *et al*., 2017) alongside the Tehuacan-Cuicatlan Reserve (Hernández-Hernández *et al*., 2011). Cactaceae is a better-known cacti group, which exhibit hallmark features to survive under prolonged drought conditions such as Crassulacean acid metabolism which allows plants uptake CO_2 at night while reducing water loss, a lack or reduced number of leaves, succulent tissues to store water (Guerrero *et al*., 2019; Köhler *et al*., 2020), spines to protect stems from extreme temperatures, herbivores and parasites (Aliscioni *et al*., 2021). The cactus family plays an important ecological role in the ecosystem. Flowers, fruits and pollen are food and the plant structure provides shelter for the local fauna, prevents soil erosion and promotes moisture and organic material in soil (Orr *et al*., 2015; Delgado-Fernández *et al*., 2017; Guerrero *et al*., 2019). In addition, Cactaceae species also have other important uses such as ornamental, medicinal, food and forage purposes (Barrios *et al*., 2020; Estrada-Castillón *et al*., 2021), mainly in Mexico, as an important source of income. Approximately 84 cacti species have been cultivated (Estrada-Castillón *et al*., 2021). Certain species of cacti are endangered group due to its popularity for xeriscaping, private collectors and illegal commercialisation, especially in rare species with high ornamental value (Yesson *et al*., 2011), while some cacti species are considered highly invasive worldwide and most of them spread as ornamental plants, causing negative conflicts around biodiversity, the economy and resource availability (Novoa *et al*., 2017). Therefore, how to quickly and accurately identify those species has become an urgent problem to be solved.

Three subfamilies (Opuntioideae, Pereskioideae and Cactoideae) are currently recognised in the family Cactaceae. During evolution, cacti have undergone fast diversification processes (Guerrero *et al*., 2019), resulting in uncommon phenomena in plants (e.g. morphological convergence, homoplasy), which make taxonomic identification difficult (Breslin *et al*., 2021; Villalobos-Barrantes *et al*., 2022). Morphological convergence refers morphological similarities between phylogenetically

distant taxa. For example, columnar and spherical shapes have evolved repeatedly in Cactaceae (Villalobos-Barrantes *et al*., 2022). Homoplasy refers to independent evolutionary origins of organs with the same function in different species. For example, the giant columnar cacti of North and South America diverge in morphological and molecular phylogenetic classification (Copetti *et al*., 2017). The subfamily Opuntioideae is the most widespread. However, the demarcation of taxa and species within this subfamily remains controversial (Köhler *et al*., 2020). From traditional morphological analysis, this subfamily is divided into as many as 20 genera, while from molecular phylogenetic studies, mostly based on molecular markers *rpl16, trnL-trnF* and ITS sequences, more smaller genera were identified (Köhler *et al*., 2020). Therefore, more phylogenetic studies with different molecular markers are needed to better understand the Opuntioideae taxonomy (Majure & Puente, 2014; Köhler *et al*., 2020). The subfamily Cactoideae has the highest diversity, including 1,222 species characterised by succulent stems with ribs or tubercles and aroles without glochids (El Mokni *et al*., 2020; Morais da Silva *et al*., 2021). Likewise, demarcation boundaries within the Cactoideade subfamily are controversial. The division of Cactoideae into two subfamilies has been proposed, with insufficient molecular evidence (Hernández-Hernández *et al*., 2011). The subfamily Pereskioideae, considered a relict, includes cacti with ancestral morphology that are not similar to typical cacti but share similar physiological patterns to cacti without leaves and a succulent stem, holding shrubs with photosynthetic leaves and non-succulent stems. Two clades (Leuenbergeria and Pereskia) were proposed based on the rapid and slow (respectively) development of the stem bark (Moore *et al*., 2018; Walker *et al*., 2018; Guerrero *et al*., 2019).

Generally, there are still many drawbacks in identifying cactus specimens morphologically. First, morphological identification requires professional taxonomists, but experts also have geographical limitations. Therefore, it is often preferred to use locality rather than morphology as the primary factor in taxonomic identification (Helsen *et al*., 2009). Especially in the Cactaceae, the above-mentioned morphological convergence also poses challenges to taxonomists in species identification. Second, ephemeral characteristics of some cacti, such as flowers and fruits, as two important morphological classification characteristics, and their existence period is very short. This brings trouble to some species that have similar morphological characteristics and need to be identified by flowers and fruits (Anderson, 2001; Hunt *et al*., 2006; Yesson *et al*., 2011). Third, there is no general consensus on the morphological classification of species within certain genera, such as the *Opuntia* Mill. genus taxonomy, which suggests that the stability of morphological taxonomy still needs to be improved. (Labra, 2003). Therefore, it is necessary to develop a strategy that provides for the rapid identification of cactus taxa, enabling the conservation of priority species. Since 2003, DNA barcoding has become an important tool for taxonomy, especially when morphological features are lacking. DNA barcoding is a technique that uses standardized short fragments of an organism's genomic DNA to achieve species identification (Li *et al*., 2015). A DNA barcode must have three essential characteristics: high versatility, good quality sequences, and high discriminatory power (Hebert *et al*., 2003). One of the main characteristics that an ideal barcode should have is the separation within species variability versus between species variability. This concept is called the “gap barcode”, which was proposed by Meyer and Paulay (2005), and shows the effectiveness of DNA barcodes in distinguishing between species. When the minimum value of the interspecific divergence is greater than the maximum value of the intraspecific divergence, a gap will appear (Meyer & Paulay, 2005). As a consequence, a threshold can be established for the species delimitation. Below this threshold are identified as the same species (Wiemers & Fiedler, 2007). According to current molecular and morphological species identification, a large number of species have interspecific distances greater than 2% and intraspecific distances less than 1% (Phillips *et al*., 2022). Additionally, barcode gaps could even be used to identify new species (Hebert *et al*., 2004; Smith *et al*., 2006).

According to these standards mentioned above, the Consortium for the Barcode of Life (CBOL) recommended two chloroplast loci as the central DNA barcodes in plants: the region of the maturase K gene (*matK*) and the ribulose-bisphosphate carboxylase gene (*rbcL*) (CBOL Plant Working Group *et al*., 2009). Additionally, The ITS2 region was selected as a barcode candidate because the sequences generated with this region have been widely used for phylogenetic reconstructions at the genus and species levels (Schultz & Wolf, 2009; Keller *et al*., 2010; Marghali *et al*., 2015; Feng *et al*., 2016). Chen *et al*., (2010) validated and proposed the inclusion of the ITS2 region as a complementary region in DNA barcode studies as it seems to be essential in the identification of species. Compared to the plastid region, this DNA region has many advantages, such as high ubiquity, small size (200–400 bp), and high recognition ability in closely related plant groups (Chen *et al*., 2010).

To the best of our knowledge, most barcoding studies in Mexico have focused on clade (Yesson *et al*., 2011; García Aguilar *et al*., 2013; Majure & Puente, 2014) and species identification (Tapia *et al*., 2017; Aquino *et al*., 2019), with few reports on the effectiveness of DNA barcoding. The objective of our study was to analyse the *matK*, *rbcL*, and ITS2 barcode regions to test their universality and estimate their discriminatory power in species of Cactaceae using three evaluation criteria (BLAST, genetic distance, and neighbour-joining trees). In addition, the barcode gap was evaluated in the proposed barcodes.

Material and Methods

Plant sampling: Coahuila region is an arid zone, with 127 species of Cactaceae (Hernández *et al*., 2004) but their genetic diversity is unknown. Thus, we took a convenience sample in Saltillo, Coahuila that comprises a total of 49 samples were collected in the southeast of Coahuila,

Mexico (Supplementary Table 1), belonging to 17 species of the Cactaceae; 15 species were represented by at least 2 individuals. Approximately 1 g of each sample was collected and saved in 96% alcohol for DNA extraction. The herbarium vouchers were prepared, morphologically identified using morphological characters as described by Flores (2005), and deposited at the Antonio Narro Saltillo México (ANSM) Herbarium in Saltillo, Coahuila, Mexico. The BOLD Systems numbers of the newly obtained sequences are available in Supplementary Table 1.

DNA extraction, PCR amplification, and DNA sequencing: Total genomic DNA was extracted using a modified cetyltrimethyl ammonium bromide (CTAB) method (Sharma *et al*., 2003). The concentration and quality of the extracted DNA were checked with 0.8% agarose gel electrophoresis. DNA was stored at –20°C for later analysis. Polymerase chain reaction (PCR) amplification of the *matK*, *rbcL*, and ITS2 genes was performed with GoTaq Green Master Mix (Promega, USA) using the respective primers. The primers and PCR conditions are detailed in (Table 1). The amplified PCR fragments were verified by agarose gel electrophoresis (0.8%). PCR products were purified using the commercial Wizard SV Gel and PCR Clean Up SystemTM (Promega, USA), following the manufacturer's instructions, and sequenced by the Langebio sequencing service (Irapuato, Mexico).

Supplementary Table 1. List of collected samples. All specimens were identified based on the morphology of the species. The sample collection area, sample voucher numbers and BOLD.

Subfamily	Genus	Species	Collection area	Voucher number	BOLD Systems accession number		
					rbcL	*matK*	ITS2
Cactoideae	*Acharagma*	*roseanum*	Ramos Arizpe	103261	-	CACMK004-22	CACIT004-22
Cactoideae	*Acharagma*	*roseanum*	Ramos Arizpe	103263	CACRB007-22	CACMK006-22	CACIT005-22
Cactoideae	*Acharagma*	*roseanum*	Ramos Arizpe	103262	-	CACMK005-22	-
Cactoideae	*Ariocarpus*	*retusus*	Saltillo	103247	CACRB004-22	CACMK003-22	CACIT002-22
Cactoideae	*Ariocarpus*	*retusus*	Saltillo	103251	CACRB006-22	-	CACIT003-22
Cactoideae	*Ariocarpus*	*retusus*	Saltillo	103250	CACRB005-22	-	-
Cactoideae	*Astrophytum*	*capricorne*	General Cepeda	103237	CACRB002-22	CACMK001-22	CACIT001-22
Cactoideae	*Astrophytum*	*capricorne*	General Cepeda	103236	CACRB001-22	-	-
Cactoideae	*Astrophytum*	*capricorne*	General Cepeda	103238	CACRB003-22	-	-
Cactoideae	*Coryphantha*	*delaetiana*	General Cepeda	103241	CACRB008-22	CACMK007-22	CACIT006-22
Cactoideae	*Coryphantha*	*delaetiana*	General Cepeda	103242	CACRB009-22	-	CACIT007-22
Cactoideae	*Coryphantha*	*delaetiana*	General Cepeda	103243	CACRB010-22	CACMK008-22	CACIT008-22
Cactoideae	*Coryphantha*	*radians*	Ramos Arizpe	103256	-	-	-
Cactoideae	*Coryphantha*	*radians*	Ramos Arizpe	103252	CACRB011-22	-	-
Cactoideae	*Coryphantha*	*radians*	Ramos Arizpe	103257	-	-	-
Cactoideae	*Echinocactus*	*horizonthalonius*	General Cepeda	103253	CACRB015-22	CACMK010-22	CACIT009-22
Cactoideae	*Echinocactus*	*horizonthalonius*	General Cepeda	103255	CACRB016-22	CACMK011-22	CACIT010-22
Cactoideae	*Echinocactus*	*horizonthalonius*	General Cepeda	103254	-	CACMK009-22	-
Cactoideae	*Epithelantha*	*bokei*	Ramos Arizpe	103275	CACRB012-22	-	-
Cactoideae	*Epithelantha*	*bokei*	Ramos Arizpe	103276	-	-	-
Cactoideae	*Epithelantha*	*bokei*	Ramos Arizpe	103277	CACRB014-22	-	-
Cactoideae	*Ferocactus*	*hamatacanthus*	Ramos Arizpe	103215	CACRB019-22	-	CACIT012-22
Cactoideae	*Ferocactus*	*hamatacanthus*	Ramos Arizpe	103216	CACRB017-22	CACMK012-22	-
Cactoideae	*Ferocactus*	*hamatacanthus*	Saltillo	103217	CACRB018-22	CACMK013-22	CACIT011-22
Cactoideae	*Mammillaria*	*chionocephala*	Saltillo	103235	CACRB022-22	-	CACIT014-22
Cactoideae	*Mammillaria*	*chionocephala*	Saltillo	103202	CACRB020-22	-	CACIT013-22
Cactoideae	*Mammillaria*	*chionocephala*	Saltillo	103205	CACRB021-22	-	-
Cactoideae	*Mammillaria*	*heyderi*	General Cepeda	103221	CACRB023-22	CACMK015-22	CACIT015-22
Cactoideae	*Mammillaria*	*heyderi*	General Cepeda	103222	CACRB024-22	CACMK016-22	CACIT016-22
Cactoideae	*Mammillaria*	*heyderi*	General Cepeda	103223	CACRB025-22	CACMK017-22	-
Cactoideae	*Mammillaria*	*pottsii*	General Cepeda	103224	CACRB026-22	CACMK018-22	-
Cactoideae	*Mammillaria*	*pottsii*	General Cepeda	103225	-	-	-
Cactoideae	*Mammillaria*	*pottsii*	General Cepeda	103226	-	CACMK019-22	-
Cactoideae	*Sclerocactus*	*scheeri*	Saltillo	103248	CACRB031-22	CACMK021-22	-
Cactoideae	*Sclerocactus*	*scheeri*	Saltillo	103274	CACRB030-22	CACMK020-22	-
Cactoideae	*Sclerocactus*	*scheeri*	Saltillo	103249	CACRB032-22	CACMK022-22	-
Cactoideae	*Stenocactus*	*multicostatus*	Saltillo	103218	CACRB027-22	-	-
Cactoideae	*Stenocactus*	*multicostatus*	Saltillo	103219	CACRB028-22	-	-
Cactoideae	*Stenocactus*	*multicostatus*	Saltillo	103220	CACRB029-22	-	-
Cactoideae	*Thelocactus*	*bicolor*	Saltillo	103244	-	-	CACIT018-22
Cactoideae	*Thelocactus*	*bicolor*	Saltillo	103246	CACRB033-22	-	CACIT017-22
Cactoideae	*Thelocactus*	*bicolor*	Saltillo	103245	-	CACMK023-22	CACIT019-22
Cactoideae	*Thelocactus*	*phymatothele*	Saltillo	103258	CACRB036-22	-	CACIT020-22
Cactoideae	*Thelocactus*	*phymatothele*	Saltillo	103259	CACRB037-22	-	CACIT021-22
Cactoideae	*Thelocactus*	*phymatothele*	Saltillo	103260	CACRB038-22	-	CACIT022-22
Cactoideae	*Thelocactus*	*rinconensis var nidulans*	Saltillo	103272	CACRB039-22	-	CACIT023-22
Cactoideae	*Thelocactus*	*rinconensis var rinconensis*	Ramos Arizpe	103273	CACRB040-22	CACMK024-22	CACIT025-22
Cactoideae	*Thelocactus*	*rinconensis var rinconensis*	Ramos Arizpe	103240	CACRB041-22	CACMK026-22	CACIT024-22
Cactoideae	*Thelocactus*	*rinconensis var rinconensis*	Saltillo	103271	-	CACMK025-22	-

Data analysis

The sequences of species of the Cactaceae were downloaded corresponding to the three DNA barcode regions (*rbcL*, *matK*, and ITS2) from the databases of BOLD Systems and GenBank (Supplementary Table 2). Sequences were filtered according to the documented nucleotide length for each region and without ambiguous "N" bases.

The alignment of the query sequences (generated in the laboratory) and the reference sequences (downloaded from the database) was carried out for each region and then adjusted using the ClustalW tool (Thompson *et al*., 1994) implemented in MEGA X (Version 10.2.6). In this study, only the sequences of species represented by more than one individual were used. The viable sites were verified using MEGA X (Tamura *et al*., 2011). DNA barcode gap analysis was performed with the SPIDER package (Version 1.5.0) in the R environment (Version 4.3.2) (Meyer & Paulay, 2005) by calculating the relative distribution of inter- and intraspecific genetic variation under the Kimura 2-parameter (K2P) distance model which assumed equal transitions and different transversion rates in nucleotide substitution pattern. The effectiveness of the barcodes was evaluated using three methods widely used in DNA barcode technology (BLAST, genetic distance, and neighbour-joining trees).

The BLAST method was utilised to construct a query database with the sequences generated in this study with each barcode and perform a sequence query with the support of the NCBI website, the command "blastn," and the MegaBlast algorithm. If a value of $E < 1 \times 10{-}5$ and a maximum correct rate of 98 to 100% was obtained from individuals of the expected species, the species identification was considered successful. If the query sequence matched several species, including the expected species, it was treated as an ambiguous identification, and the sequences that did not match the expected species were considered failed (Chen *et al*., 2015).

To test the discrimination rates based on the distance analysis, the "best close match" function (Meier *et al*., 2008) of the SPIDER package (Brown *et al*., 2012) under the Kimura 2-parameter (K2P) method was used for each DNA region with all possible combinations. The function "best close match" was to find the best-matching reference barcode in the corresponding database below a threshold value. In this study, the query sequences that were only assigned to one correct taxon within a 95% threshold were identified as "correct." If the assignment included not only the correct taxon, it was considered "ambiguous." When the query sequences were assigned to the incorrect taxon, they were considered "incorrect." The sequences without any assignment were considered "unidentified" (Meier *et al*., 2006).

In this study, neighbour-joining trees were used to evaluate the clade recovery capabilities of barcodes The construction of phylogenetic tree was done with MEGA X (Kumar *et al*., 2018; Glen *et al*., 2020) using the neighbour-joining clustering method, which used the distance matrix derived from the multialigments and is consistent with the use of different models of evolution to recover the true tree (Saitou and Nei, 1987). We used K2P genetic distance model (Kimura, 1980) to evaluate discriminatory performance by calculating the proportion of species by taking the quotient of transition/transversion ratio (Saitou & Nei, 1987). All ambiguous positions were removed for each sequence pair (pairwise deletion option). Only all species of one genus forming a monophyletic clade in the tree were considered to have been successfully identified.

Results

Universality of the barcode regions and the characteristics of the sequences: Two candidate DNA regions (*rbcL* and ITS2) showed high amplification rates in 49 samples, with percentages of 97.95% and 89.79%, respectively. Regarding the *matK* region, the initially used primer pairs 427F and 1248R showed extremely low amplification rates (< 10%) (Yu *et al*., 2011). In addition, it was impossible to generate high-quality sequences. Therefore, this study used our newly designed specific *matK* primers for Cactaceae (MmaF and MmaR) (Table 1), with successful amplification achieved for 91.95% of the samples. The sequence recovery rate varied for each barcode. A relatively high rate of good quality sequences was generated with the *rbcL* region (81.25%), while for the *matK* (54.16%) and ITS2 regions (56.81%), recoverability was moderately low. This study only included sequences that lacked unambiguous nucleotides.

A total of 234 DNA barcode sequences were available, 39 of which were generated in this study (13 *rbcL*, 12 *matK*, and 14 ITS2 marker sequences), and 195 reference sequences were from the database. (Table 2) summarises the sequence characteristics of the three DNA barcodes. The aligned sequence length of *rbcL*, *matK*, and ITS2 regions were 587, 725, and 253 bp, respectively. The *matK* region showed the highest variability (57.10%), followed by the ITS2 region (45.84%) and the *rbcL* region (41.56%). The *rbcL* and *matK* regions had a similar percentage of informative sites, 54.51% and 54.75%, respectively. The ITS2 region showed the lowest percentage of informative sites (39.92%). The genetic distances of these three barcode regions were also calculated based on the K2P model. The *rbcL* region showed the highest mean of intraspecific and interspecific distances (12.66% and 27.30%, respectively), followed by the *matK* (1.50% and 17.48%, respectively) and ITS2 regions (0.45% and 6%, respectively).

Barcode gap assessment: Ideally, the genetic variation of a barcode should show a "gap" between the intra- and interspecific distances rather than an overlap of both. In this study, the distributions of intra- and interspecific divergences were examined in the three individual regions (*rbcL*, *matK*, and ITS2) and in the four possible combinations (*rbcL* + *matK*; *rbcL* + ITS2; *matK* + ITS2; and *rbcL* + *matK* + ITS2). Based on the analysis, the frequency distribution of intra- and interspecific divergence showed that there were no gaps in the DNA barcodes due to intraspecific divergences overlapping with interspecific divergences. However, both regions' individual and multi-region combinations showed a different pattern in the divergence distribution (Fig. 1).

Table 1. PCR primers and programmes used for DNA amplification.

Barcode region	Primers (5´- 3´)	PCR conditions	Reference
rbcL	*rbcL*a-F: ATGTCACCACAAACAGAGACTAAAGC *rbcL*a-R: GTAAAATCAAGTCCACCRCG	94°C, 5 min; 35 × (94°C, 30 s; 50°C, 40 s; 72°C, 40 s); 72°C, 10 min; ∞4°C.	(Kress *et al.*, 2009)
matK	Mma-F: TAGTTTTCGGTCTAATTTGAAA Mma-R: ATAGATTCTTCTTGGTTGAGA	94°C, 5 min; 35 × (94°C, 15 s; 56°C, 20 s; 72°C, 50 s); 72°C, 10 min; ∞4°C.	(Designed in this study)
ITS2	ITS-2F: ATGCGATACTTGGTGTGAAT ITS-3R: GACGCTTCTCCAGACTACAAT	9 5°C, 4 min; 35 × (94°C, 45 s; 56°C, 1 min; 72°C, 1 min); 72°C, 10 min; ∞4°C.	(Chen *et al.*, 2010b)

Table 2. Summary of the variability of three DNA barcode regions (*rbcL*, *matK*, and ITS2).

Barcode region	Number of accessions	Sequence length	Length of aligned sequence	Content of G + C (%)	No. of variable sites (%)	No. of informative sites (%)	*Intraspecific distance range (mean) %	*Interspecific distance range (mean)%
rbcL	78	552-592	587	0.4390	331 (41.56)	320 (54.51)	0-56.14 (12.66)	0-56.75 (27.30)
matK	78	423-1041	725	0.3281	414 (57.10)	397 (54.75)	0-50 (1.50)	0-55.35 (17.48)
ITS2	78	203-680	253	0.7325	116 (45.84)	101 (39.92)	0-7.46 (0.45)	0-33.33 (6.00)

NOTE: *The mean intra- and interspecific distances was calculated for the species, which were represented by > 1 individual

Table 3. Identification success rates of seven datasets using the "best close match" function of the SPIDER package from R.

Barcode region	Best close match (%)				Threshold (%)
	Ambiguous	Correct	Incorrect	Unidentified	
rbcL	23 (29.49)	26 (33.33)	28 (35.90)	1 (1.28)	0.51
matK	35 (44.87)	22 (28.21)	16 (20.51)	5 (6.41)	0.45
ITS2	34 (43.59)	23 (29.49)	20 (25.64)	1 (1.28)	0.06
matK+ITS2	18 (23.08)	30 (38.46)	29 (37.18)	1 (1.28)	0.29
rbcL+ITS2	15 (19.23)	31 (39.74)	28 (35.90)	4 (5.13)	0.27
rbcL+*matK*	16 (20.51)	33 (42.31)	24 (30.77)	5 (6.41)	0.15
rbcL+*matK*+ITS2	4 (5.13)	41 (52.56)	29 (37.18)	4 (5.13)	0.13

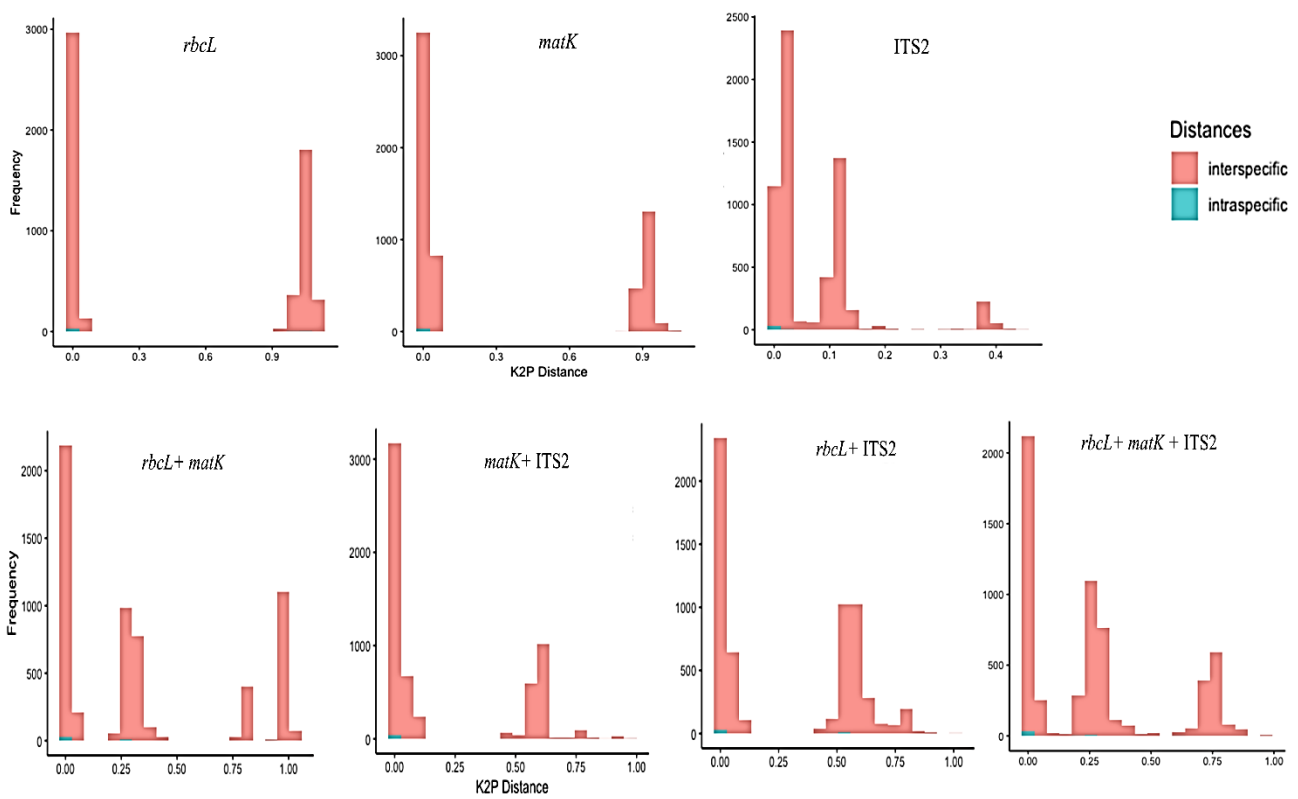


Fig. 1. Relative distribution of intra- and interspecific distances for the three DNA regions (*rbcL*, *matK*, and ITS2) and multi-region combinations (*rbcL* + *matK*, *matK* + ITS2, *rbcL* + ITS2, *rbcL* + *matK* + ITS2). As the intraspecific divergences overlapped with interspecific divergences no barcode gaps was confirmed.

Supplementary Table 2. Sequences downloaded from BOLD Systems and GenBank. For each species of the list, the database accession numbers for *rbcL*, *matK*, and ITS2 were provided.

Subfamily	Genus	Species	Barcode region		
			rbcL	***matK***	**ITS2**
Cactoideae	*Ariocarpus*	*retusus*	-	MK449033.1	-
Cactoideae	*Ariocarpus*	*retusus*	-	GBVR3489-13*	-
Cactoideae	*Acharagma*	*roseanum*	MK449086.1	-	-
Cactoideae	*Hylocereus*	*costaricensis*	JQ590992.1	JQ587174.1	JX289629.1
Cactoideae	*Hylocereus*	*costaricensis*	JQ590993.1	JQ587175.1	JX289631.1
Pereskioideae	*Pereskia*	*aculeata*	JX905965.1	JX905944.1	L78035.1
Pereskioideae	*Pereskia*	*aculeata*	LN871112.1	AY042626.1	HQ872561.1
Opuntioideae	*Maihueniopsis*	*ovata*	LN555681.1	JF786723.1	JF786887.1
Opuntioideae	*Maihueniopsis*	*ovata*	LN555676.1	JF786724.1	JF786888.1
Opuntioideae	*Nopalea*	*cochenillifera*	JF787166.1	JF786727.1	EU559652.1
Opuntioideae	*Nopalea*	*cochenillifera*	JF787167.1	JF786728.1	EU559654.1
Opuntioideae	*Nopalea*	*hondurensis*	JF787172.1	JF786732.1	JF786896.1
Opuntioideae	*Nopalea*	*hondurensis*	JF787171.1	JF786733.1	JF786897.1
Opuntioideae	*Consolea*	*corallicola*	JF787157.1	F786713.1	JF786877.1
Opuntioideae	*Consolea*	*corallicola*	JF787156.1	JF786714.1	JF786878.1
Opuntioideae	*Consolea*	*nashii*	JF787159.1	GBVS4591-13*	JF786881.1
Opuntioideae	*Consolea*	*nashii*	JF787160.1	GBVS4592-13*	JF786882.1
Opuntioideae	*Consolea*	*rubescens*	JF787161.1	JF786719.1	JF786883.1
Opuntioideae	*Consolea*	*rubescens*	JF787162.1	JF786720.1	JF786884.1
Opuntioideae	*Opuntia*	*bravoana*	JF787193.1	JF786754.1	JF787038.1
Opuntioideae	*Opuntia*	*bravoana*	F787194.1	JF786755.1	JF787039.1
Opuntioideae	*Opuntia*	*camanchica*	JF787195.1	JF786816.1	JF786917.1
Opuntioideae	*Opuntia*	*camanchica*	JF787253.1	JF786756.1	JF786973.1
Opuntioideae	*Opuntia*	*cespitosa*	JF787197.1	JF786759.1	JF786920.1
Opuntioideae	*Opuntia*	*cespitosa*	JF787198.1	JF786760.1	JF786921.1
Opuntioideae	*Opuntia*	*chlorotica*	SDH784-14*	JF786763.1	HQ872601.1
Opuntioideae	*Opuntia*	*chlorotica*	JF787201.1	FN997530.1	JF786924.1
Opuntioideae	*Opuntia*	*dillenii*	JF787289.1	JF786854.1	JF787011.1
Opuntioideae	*Opuntia*	*dillenii*	JF787290.1	JF786855.1	JF787010.1
Opuntioideae	*Opuntia*	*ellisiana*	JF787212.1	JF786775.1	JF786935.1
Opuntioideae	*Opuntia*	*ellisiana*	JF787213.1	JF786776.1	JF786936.1
Opuntioideae	*Opuntia*	*engelmannii*	MT254721.1	GBVR3506-13*	HQ872523.1
Opuntioideae	*Opuntia*	*engelmannii*	MT254722.1	GBVR3697-13*	HQ872543.1
Opuntioideae	*Opuntia*	*ficus-indica*	GBVE4445-11*	GBVR3686-13*	EU428845.1
Opuntioideae	*Opuntia*	*ficus-indica*	SDH786-14*	GBVS4658-13*	EU930379.1
Opuntioideae	*Opuntia*	*humifusa*	EF590552.1	KJ772965.1	JQ245718.1
Opuntioideae	*Opuntia*	*humifusa*	GQ248660.1	MH621623.1	JF786949.1
Opuntioideae	*Opuntia*	*leucotricha*	JF787231.1	JF786795.1	HQ872539.1
Opuntioideae	*Opuntia*	*leucotricha*	F787232.1	JF786796.1	HQ872585.1
Opuntioideae	*Opuntia*	*macbridei*	JF787236.1	JF786799.1	JF786957.1
Opuntioideae	*Opuntia*	*macbridei*	JF787269.1	JF786800.1	JF786958.1
Opuntioideae	*Opuntia*	*macrorhiza*	MK526261.1	JF786802.1	JF786960.1
Opuntioideae	*Opuntia*	*macrorhiza*	JF787239.1	JF786803.1	JQ245719.1
Opuntioideae	*Opuntia*	*microdasys*	SDH789-14*	AY042622.1	HQ872501.1
Opuntioideae	*Opuntia*	*microdasys*	JF787246.1	GBVS4683-13*	HQ872548.1
Opuntioideae	*Opuntia*	*oricola*	SDH791-14*	GBVS4686-13*	HQ872569.1
Opuntioideae	*Opuntia*	*oricola*	JF787249.1	JF786812.1	JF786969.1
Opuntioideae	*Opuntia*	*pachyrrhiza*	JF787250.1	JF786813.1	JF786970.1
Opuntioideae	*Opuntia*	*pachyrrhiza*	JF787251.1	JF786814.1	JF786971.1
Opuntioideae	*Opuntia*	*phaeacantha*	SDH792-14*	FN997327.1	HQ872550.1
Opuntioideae	*Opuntia*	*phaeacantha*	JF787254.1	GBVS4691-13*	HQ872527.1
Opuntioideae	*Opuntia*	*polyacantha*	MG245909.1	JF786822.1	MG236336.1
Opuntioideae	*Opuntia*	*polyacantha*	MG248094.1	JF786823.1	HQ872573.1
Opuntioideae	*Opuntia*	*pusilla*	KJ773704.1	KJ772966.1	JQ245721.1
Opuntioideae	*Opuntia*	*pusilla*	JF787263.1	JF786829.1	JF786984.1
Opuntioideae	*Opuntia*	*repens*	JF787272.1	JF786835.1	JF786993.1
Opuntioideae	*Opuntia*	*repens*	F787273.1	JF786836.1	JF786992.1
Opuntioideae	*Opuntia*	*rufida*	JF787276.1	JF786840.1	HQ872591.1
Opuntioideae	*Opuntia*	*rufida*	JF787277.1	JF786841.1	JF786997.1
Opuntioideae	*Opuntia*	*stricta*	KU877485.1	KJ772967.1	HQ872534.1
Opuntioideae	*Opuntia*	*stricta*	JQ412402.1	JQ412275.1	HQ872558.1
Opuntioideae	*Opuntia*	*strigil*	JF787291.1	JF786857.1	JF787014.1
Opuntioideae	*Opuntia*	*strigil*	JF787292.1	JF786858.1	JF787012.1
Opuntioideae	*Opuntia*	*tomentosa*	JF787296.1	JF786862.1	HQ872546.1
Opuntioideae	*Opuntia*	*tomentosa*	JF787297.1	JF786863.1	HQ872582.1
Opuntioideae	*Opuntia*	*salmiana*	JF787278.1	JF786842.1	JF786998.1
Opuntioideae	*Opuntia*	*salmiana*	JF787279.1	JF786843.1	JF786999.1

Species discrimination based on different analytical methods: In the BLAST analytic results of our query sequences, the *matK* region obtained the highest identification rate at the generic and species levels (83.33% and 50%, respectively), followed by the *rbcL* region (71.43% and 14.29%, respectively). Both *matK* and *rbcL* regions had similar ambiguous identification rates (14.29% and 16.67%, respectively). However, due to the lack of ITS2 reference sequences in the database for the collected species in this study, the identification rate of the ITS2 region was not obtained.

Based on the "best close match" analysis (Table 3), we found that, of the three regions analysed individually, the *rbcL* region had the highest correct identification rate (33.33%), followed by the ITS2 (29.49%) and *matK* regions (28.21%). However, the incorrect identification rate of the *rbcL* region was also high (35.90%). For these regions' combination analysis, the three regions group (*rbcL* + *matK* + ITS2) showed the highest identification rate (52.56%), and the highest incorrect identification rate (37.18%). The second highest combination group was *rbcL* + *matK* (42.31%), with the lowest incorrect identification rate (30.77%) of combination groups. Another two combination groups (*rbcL* + ITS2 and *rbcL* + ITS2) also increased the identification rate (38.46% and 39.74%, respectively) compared with the use of any single region, but with a high incorrect identification rate (37.18% and 35.90%, respectively).

In the neighbour-joining analysis, only the species that formed monophyletic clades according to their genus were considered successfully identified. Of the three regions' individual analysed results, the tree generated with the ITS2 region had the highest discrimination rate (92.31%), followed by *matK* (88.46%) and *rbcL* (82.05%). Of the regions' analysed combination results, the best combination was *rbcL* + *matK*, with a discrimination percentage of 87.18%, followed by *rbcL* + *matK* + ITS2 (85.90%), *matK* + ITS2 (83.33%), and *rbcL* + ITS2 (76.92%).

Based on the morphological characteristics of the genera and species analysis, the group of data analysed here involved the three subfamilies, Opuntioideae, Cactoideae, and Pereskioideae. In this study, the subfamilies were represented by at least one species. According to this information, it could be expected that the formation of monophyletic clades would achieve the separation of species by subfamilies. Except for the ITS2 region, the phylogenetic trees generated by regional individuals and regional combinations have achieved the separation and identification of the two monophyletic groups, the Cactoideae and Opuntioideae subfamilies (Fig. 2 and Supplementary Figs. 1–6). The ITS2 region achieved the clear separation of four mainly monophyletic groups (Cactoideae, Opuntioideae, Pereskioideae, and the genus *Hylocereus*). This ITS2 barcode region separated the only species of the Pereskioideae subfamily (*Pereskia aculeata*) into a monophyletic clade, which was not observed in any other tree with the barcode region individually or in multi-region analysis (Fig. 2).

The phylogenetic tree generated with *rbcL* region did not group complete *Opuntia* species into Opuntideae subfamily (Supplementary Fig. 1). Similarly, the phylogenetic tree generated with *matK* did not assign three *Opuntia* species into Opuntideae subfamily, and nor did group *Ariocarpus* and *Hylocereus* species into Cactoideae subfamily (Supplementary Fig. 2). In the phylogenetic tree generated with ITS2 region, only *Hylocereus* species were not correctly assigned to Cactoideae subfamily (Fig. 2). In the phylogenetic trees generated with regions' combination, the phylogenetic tree obtained with *rbcL*+ ITS2 did not group all the *Opuntia* species to Opuntioideae subfamilies (Supplementary Fig. 5), while for the phylogenetic trees generated with *rbcL* + *matK* and *matK* + ITS2 groups, *Ariocarpus* and *Hylocereus* species were not assigned correctly to Cactoideae subfamily; three Opuntia species were not classified as Opuntioideae subfamilies (Supplementary Figs. 3-4). The phylogenetic tree generated with three regions' combination (*rbcL* + *matK* + ITS2), did not group all the *Opuntia* species into Opuntideae subfamily (Supplementary Fig. 6).

Discussion

In the development of DNA barcode technology, primers effectiveness in the analysis of biological samples is crucial. Therefore, molecular markers must have a high degree of universality to achieve successful amplification and sequencing of large sample populations. In addition, they must possess a sufficient nucleotide substitution rate to allow for discrimination between species (Hollingsworth *et al*., 2011; Naciri *et al*., 2012). The three candidate DNA barcode regions selected in this study showed relatively high universal amplification coverage in the 17 species of Cactaceae. Despite this, the recovery rate of the *matK* and ITS2 sequences was low, making it impossible to use all query sequences generated in this study with these regions in the bioinformatic analysis. There are varied reports on the success of the universality of these primers in different groups of plants (Lahaye *et al*., 2008; Maloukh *et al*., 2017). Amplification with the designed *matK* primers was more successful. However, the sequence recovery rate was still low. Similarly, other authors reported a low recovery rate of sequences using *matK* primers, which corresponded to our results (Kress *et al*., 2009; Kang *et al*., 2017). The elimination of *matK* sequences in the data analysis due to poor quality has been reported in other studies Korotkova *et al*., (2011) mentioned successful amplification with *matK* primers in Cactaceae. They also pointed out that sequencing requires greater effort than the use of other regions. Variable results in amplification and sequencing with *matK* primers may be due to the diverse evolutionary rates in different taxa. In addition, the high rate of nucleotide substitutions may affect the conservation of loci in different plant groups (Gillman *et al*., 2010).

The high universality of the *rbcL* region in terms of sequence amplification and recoverability was similar to those obtained in other research work (CBOL Plant Working Group *et al*., 2009; Kang *et al*., 2017; Carneiro de Melo Moura *et al*., 2019; Ismail *et al*., 2020). Good recoverability of the *rbcL* region has been mentioned (Roy *et al*., 2010; Cabelin & Alejandro, 2016; Amandita *et al*., 2019). Several authors have demonstrated the high universality of *rbcL* primers and showed that the sequence generated by this locus was sufficient to function as the central region of DNA barcoding, as it provided sufficient variation for distinguishing species among different plant groups (Kress & Erickson, 2007; Burgess *et al*., 2011; Gu *et al*., 2011).

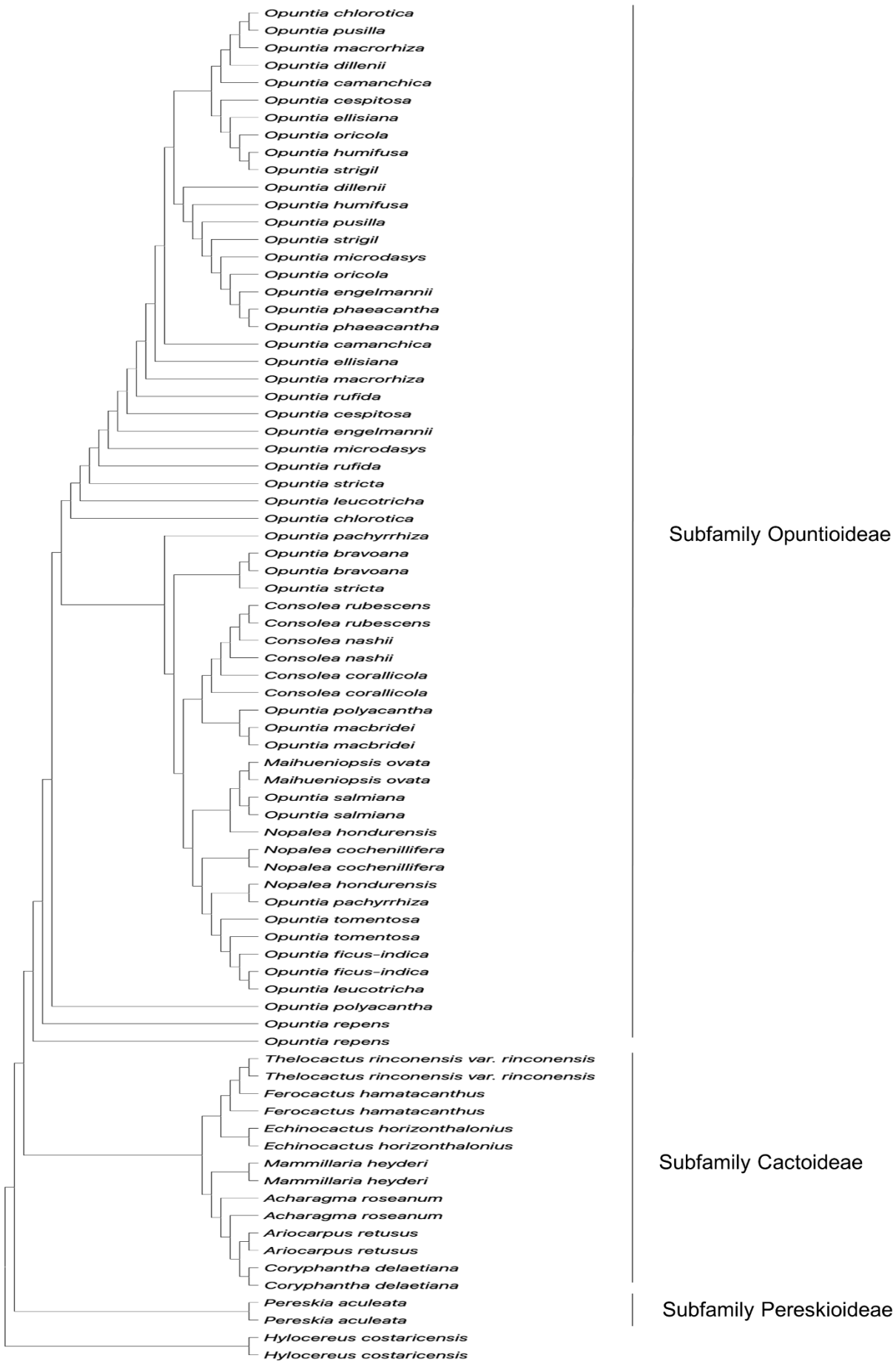


Fig. 2. Neighbour-joining tree for species of the Cactaceae family using ITS2 barcode region based on K2P genetic distance.

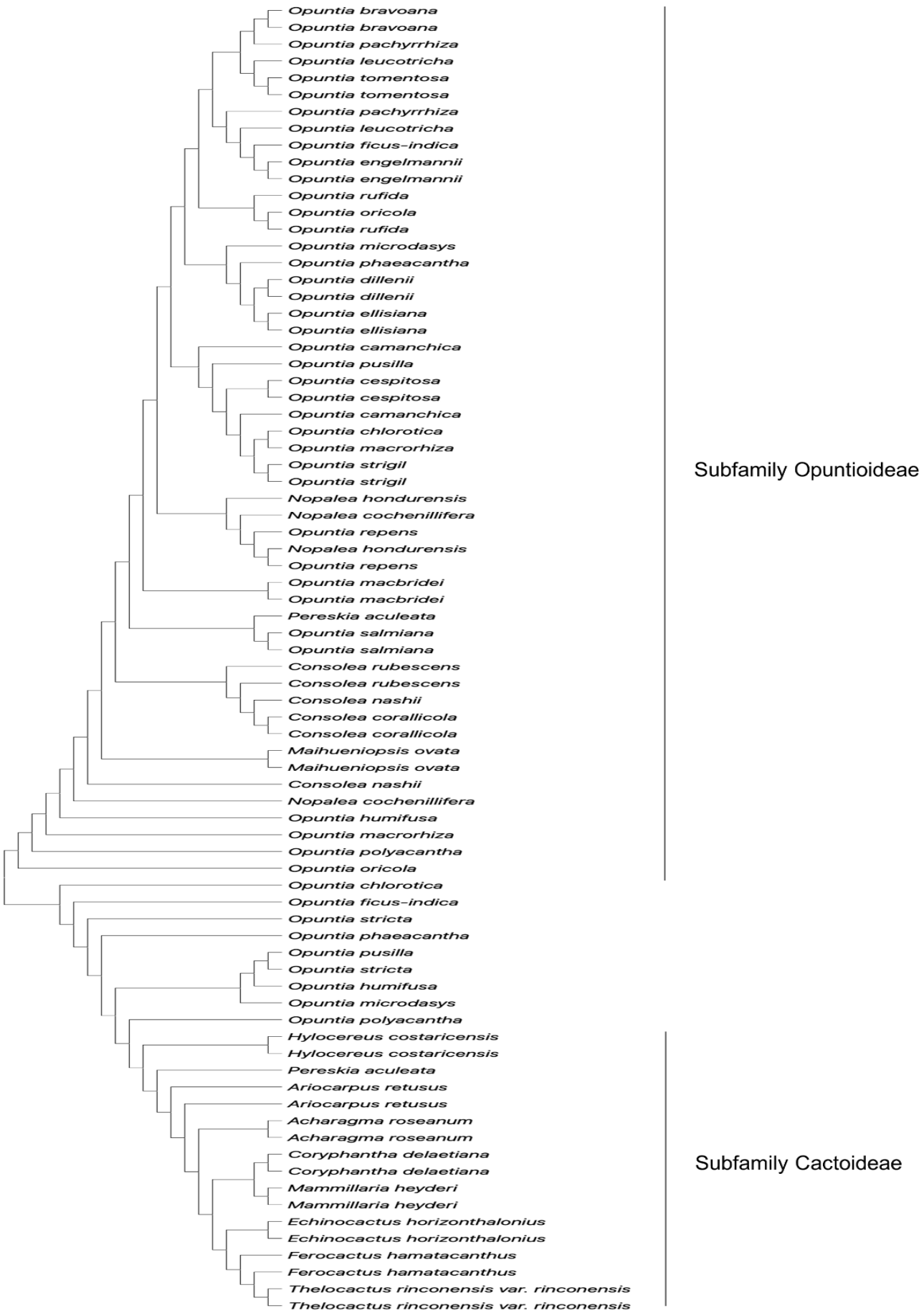
Supplementary Fig. 1. Neighbour-joining tree for species of the Cactaceae family using *rbcL* barcode region based on K2P genetic distance.

Supplementary Fig. 2. Neighbour-joining tree for species of the Cactaceae family using *matK* barcode region based on K2P genetic distance.

Supplementary Fig. 3. Neighbour-joining tree for species of the Cactaceae family using *rbcL* + *matK* barcode region based on K2P genetic distance.

Supplementary Fig. 4. Neighbour-joining tree for species of the Cactaceae family using *matK* + ITS2 barcode region based on K2P genetic distance.

Supplementary Fig. 5. Neighbour-joining tree for species of the Cactaceae family using *rbcL* + ITS2 barcode region based on K2P genetic distance.



Supplementary Fig 6. Neighbour-joining tree for species of the Cactaceae family using *rbcL* + *matK* + ITS2 barcode region based on K2P genetic distance.

In 2011, CBOL proposed the incorporation of ITS and ITS2 as a central DNA barcode along with *rbcL* and *matK* for plants. In this study, the ITS2 region had an acceptable amplification success rate but a low recovery rate. Despite this, the usable sequences had a good sequence length, suitable for bioinformatic analysis. However, a good recovery rate of the ITS2 barcode has been reported in different plant families (Chen *et al*., 2010; Chen *et al*., 2015). Due to amplification and sequencing problems in some taxa, some studies still question the use of the ITS2 region (Fu *et al*., 2011; Zhang *et al*., 2016).

For the evaluation of primer effectiveness in Cactaceae, our study still had some limitations. This family has around 1800 species, and more samples are necessary to further assess the universality of these primers, which will be useful for the barcoding project in this family.

In this study, none of the three proposed regions showed a gap due to an overlap of the intra- and interspecific distances of the sequences. Although the DNA barcode gap has been reported in some specific taxa, such as *Terastigma*, *Abies*, and *Cupressus* (Fu *et al*., 2011; Armenise *et al*., 2012). However, other authors reported the lack or non-existence of a barcode gap in some groups of land plants (Meier *et al*., 2006; Lahaye *et al*., 2008; Hollingsworth *et al*., 2009; Parmentier *et al*., 2013). However, despite this absence, the barcodes generated in this study showed relatively high discriminatory power with other applied analytical methods.

According to the analytical methods used here, the DNA barcodes selected in this study (two chloroplasts and one nuclear) showed variable identification power in the 49 cactus species analysed. In this study, *rbcL* and *matK* showed good identification rates at the genus and species levels using the BLAST method. Different studies have reported good identification at the generic and species levels with the BLAST method with these barcode regions (Chen *et al*., 2015; Cabelin & Alejandro, 2016; Yang *et al*., 2017). Gu *et al*., (2011) reported high percentages of identification in *Ligustrum* spp. at the genus level with the three regions studied here and proposed the ITS region as the most promising barcode in the identification of species at the generic level. An important factor limiting the success of species identification with DNA barcode technology is the availability of sequences of the corresponding taxa in databases. In this study, the high identification rates of *rbcL* and *matK* regions are not only because of the discriminatory power of the barcode itself but also because of the rich sequence resources in the database. However, the identification rate with the ITS2 region was influenced by the lack of sufficient reference sequences for the taxon of interest. Other authors have mentioned that samples belonging to species that have not been registered in the reference databases lead to an increase in the rates of unidentified samples, which is mainly due to incomplete molecular datasets rather than the type of data analysis (Cowan & Fay, 2012; Parmentier *et al*., 2013; Amandita *et al*., 2019).

When the samples were analysed using the genetic (closest) distance method, the individual barcodes *rbcL*, *matK*, and ITS2 indicated low identification rates (33.33%, 28.21%, and 29.49%, respectively). The highest identification rate was only achieved at 52.21% with the *rbcL* + *matK* + ITS2 combination. Similarly, Yan *et al*., (2015) described low identification percentages with this analysis method in species of *Rhododendron*. The highest correct identification rate was obtained with four regions, *rbcL* + *matK* + ITS2 + *psbA-trnH* (50.1%); individually, the *matK* region achieved a percentage of 23.51%. Low identification rates have also been described by other authors who attributed these results to the sensitivity of this method to divergence processes and the type of sampling performed (Gu *et al*., 2011; Piredda *et al*., 2011; Chen *et al*., 2015). In this part of the analysis, we did not get much genetic distance and it is possible that Cactaceae is a "young" family with number of very recently divergent lineages.

The neighbour-joining method is based on a grouping algorithm that creates a phylogenetic tree; however, its creation is influenced by the parameters used, mainly by the distance matrix and index, and by the effects of lineage classification between species with very recent divergences (Saitou & Nei, 1987; van Velzen *et al*., 2012; Simeone *et al*., 2013; Yan *et al*., 2015). The tree-based method evaluated the discrimination efficiency at the species level, so if the species of a genus formed a monophyletic clade, it was considered a successful identification. In this study, the phylogenetic tree generated with ITS2 showed the highest discrimination success rate of species in the individual barcode regions analysis. Jiménez-Barron *et al*., (2020) have also reported ITS had a better performance in phylogenetic analysis in Asparagaceae. Han *et al*., (2016) supported the use of ITS fragment as a barcode region in plant due to its better species discrimination in Aceraceae. The combination of two or more DNA regions did not improve the species discrimination rate; the most acceptable combination was with barcode regions *rbcL* + *matK*. In previous studies, multi-region analysis of two or more loci did not provide a significant gain in species discrimination (Lahaye *et al*., 2008; Shneyer & Rodionov, 2019).

Conclusion

In summary, it was observed that barcode *matK* represented better identification power in Cactaceae. However, the universality of primers of this gene is still an urgent problem to be solved. In addition, the ITS2 region demonstrated its strong species discrimination power in phylogenetic tree construction but failed in the BLAST method due to the lack of reference data in the database. Therefore, enriching data on this barcode should be conducted in the future. Furthermore, whether the ITS region can be used as a barcode for species identification in the Cactaceae is still controversial, as ITS has been shown to have non-concerted evolution in Cactaceae (Mammillaria) (Harpke *et al*., 2006). Our findings showed the species limits in Cactaceae, divergence times, the connection between diagnostic morphological characters and barcodes. Thus, further studies using machine learning algorithms that use the whole molecular, morphology, biogeography, and barcode Cactaceae data available in open databases could significantly improve the correct Cactaceae species classification.

References

Aliscioni, N.L., N. Delbón and D.E. Gurvich. 2021. Spine function in Cactaceae, A Review. *J. Prof. Assoc. Cactus Develop*., 23: 1-11.

Amandita, F.Y., K. Rembold, B. Vornam, S. Rahayu, I.Z. Siregar, H. Kreft and R. Finkeldey. 2019. DNA barcoding of flowering plants in Sumatra, Indonesia. *Ecol. Evol.*, 9(4): 1858-1868.

Anderson, E.F. 2001. *The cactus family*. Timber Press, Portland, Oregon.

Aquino, D., R.C. Cervantes, D.S. Gernandt and S. Arias. 2019. Species delimitation and phylogeny of *Epithelantha* (Cactaceae). *Sys. Bot*., 44(3): 600-615.

Armenise, L., M.C. Simeone, R. Piredda and B. Schirone. 2012. Validation of DNA barcoding as an efficient tool for taxon identification and detection of species diversity in Italian conifers. *Europ. J. For. Res*., 131(5): 1337-1353.

Barrios, D., J.A. Sánchez, J. Flores and E. Jurado. 2020. Seed traits and germination in the Cactaceae family: A review across the Americas. *Bot. Sci.*, 98(3): 417-440.

Breslin, P.B., M.F. Wojciechowski and L.C. Majure. 2021. Molecular phylogeny of the Mammilloid clade (Cactaceae) resolves the monophyly of Mammillaria. *Taxon*, 70(2): 308-323.

Brown, S.D. J., R.A. Collins, S. Boyer, M. Lefort, J. Malumnbres-Olarte, C.J. Vink and R.H. Cruickshank. 2012. Spider: an R package for the analysis of species identity and evolution, with particular reference to DNA barcoding. *Mol. Ecol. Resour*., 12(3): 562-565.

Burgess, K.S., A.J. Fazekas, P.R. Kesanakurti, S.W. Graham, B.C. Husband, S.G. Newmaster, D.M. Percy, M. Hajibabaei and S.C.H. Barrett. 2011. Discriminating plant species in a local temperate flora using the *rbcL+ matK* DNA barcode. *Methods Ecol. Evol*., 2(4): 333-340.

Cabelin, V.L.D. and G.J.D. Alejandro. 2016. Efficiency of *matK*, *rbcL*, *trnH-psbA*, and *trnL-F* (cpDNA) to molecularly authenticate Philippine ethnomedicinal Apocynaceae through DNA barcoding. *Pharm. Mag*., 12(Suppl 3): S384-S384.

Carneiro de Melo Moura, C., F. Brambach, K. Jair Hernandez Bado, K. V Krutovsky, H. Kreft, S.S. Tjitrosoedirdjo, I.Z. Siregar and O. Gailing. 2019. Integrating DNA barcoding and traditional taxonomy for the identification of dipterocarps in remnant lowland forests of Sumatra. *Plants*, 8(11): 461.

CBOL Plant Working Group, P.M. Hollingsworth, L.L. Forrest, J.L. Spouge, M. Hajibabaei, S. Ratnasingham, M. van der Bank, M.W. Chase, R.S. Cowan and D.L. Erickson. 2009. A DNA barcode for land plants. *Proceedings of the National Academy of Science*s., 106(31): 12794-12797.

Chen, J., J. Zhao, D.L. Erickson, N. Xia and W.J. Kress. 2015. Testing DNA barcodes in closely related species of *Curcuma* (Zingiberaceae) from Myanmar and China. *Mol. Ecol. Resour*., 15(2): 337-348.

Chen, S., H. Yao, J. Han, C. Liu, J. Song, L. Shi, Y. Zhu, X. Ma, T. Gao and X. Pang. 2010. Validation of the ITS2 region as a novel DNA barcode for identifying medicinal plant species. *PloS One*., 5(1): e8613-e8613.

Copetti D., A. Búrquez, E.J. L.M. Bustamante, Charboneau, K.L. Childs, L.E. Eguiarte, S. Lee, T.L. Liu, M.M. McMahon, N.K. Whiteman, R.A. Wing, M.F. Wojciechowski and M.J. Sanderson. 2017. Extensive gene tree discordance and hemiplasy shaped the genomes of North American columnar cacti. *Proceedings of the National Academy of Sciences*., 114(45): 12003-12008.

Cowan, R.S. and M.F. Fay. 2012. Challenges in the DNA barcoding of plant material. In *Plant DNA Fingerprinting and Barcoding* (pp. 23–33). Springer.

Delgado-Fernández, M., J.G. Escobar-Flores and K. Franklin. 2017. El cardón gigante (Pachycereus pringlei) y sus interacciones con la fauna en la península de Baja California, México. *Acta Universitaria*., 27(5): 11-18.

El-Mokni, R., F. Verloove, A. Guiggi and M.H. El Aouni. 2020. New records of cacti (Opuntioideae & Cactoideae, Cactaceae) from Tunisia. *Bradleya*., 2020(38): 35-50.

Estrada-Castillón, E., J.Á. Villarreal-Quintanilla, J.A. Encina-Domínguez, E. Jurado-Ybarra, L.G. Cuéllar-Rodríguez, P. Garza-Zambrano, J.R. Arévalo-Sierra, C.M. Cantú-Ayala, W. Himmelsbach and M.M. Salinas-Rodríguez. 2021. Ethnobotanical biocultural diversity by rural communities in the Cuatrociénegas Valley, Coahuila; Mexico. *J. Ethnobiol. Ethnomed*., 17(1): 1-22.

Feng, S., M. Jiang, Y. Shi, K. Jiao, C. Shen, J. Lu, Q. Ying and H. Wang. 2016. Application of the ribosomal DNA ITS2 region of *Physalis* (Solanaceae): DNA barcoding and phylogenetic study. *Front. Plant Sci*., 7: 1047.

Flores, A. 2005. Guía de cactáceas del estado de Coahuila. *SEP-Instituto de Ecología de Coahuila Jardines Para La Humanidad y Gobierno de Coahuila. Coahuila, México*. 196 pp.

Fu, Y., W. Jiang and C. Fu. 2011. Identification of species within *Tetrastigma* (Miq.) Planch.(Vitaceae) based on DNA barcoding techniques. *J. Sys. Evol*., 49(3): 237-245.

García Aguilar, M. de los A., T. Terrazas, O. Segura León, S. Arias, H. Vibrans and L. López-Mata. 2013. Caracterización molecular de tres especies de *Hylocereus* (Cactaceae) presentes en México. *Revista Fitotecnia Mexicana*., 36(1): 13-22.

Gillman, L.N., D.J. Keeling, R.C. Gardner and S.D. Wright. 2010. Faster evolution of highly conserved DNA in tropical plants. *J. Evol. Biol*., 23(6): 1327-1330.

Glen, S., T. Koichiro and K. Sudhir. 2020. Molecular evolutionary genetics analysis (MEGA) for macOS. *Mol. Biol. Evol*., 37(4): 1237-1239.

Gu, J., J. Su, R. Lin, R. Li and P. Xiao. 2011. Testing four proposed barcoding markers for the identification of species within *Ligustrum* L. (Oleaceae). *J. Sys. Evol*., 49(3): 213-224.

Guerrero, P.C., L.C. Majure, A. Cornejo-Romero and T. Hernández-Hernández. 2019. Phylogenetic relationships and evolutionary trends in the Cactus family. *J. Hered*., 110(1): 4-21.

Han, Y.W., D. Duan, X.F. Ma, Y. Jia, Z.L. Liu, G.F. Zhao and Z.H. Li. 2016. Efficient identification of the forest tree species in Aceraceae using DNA barcodes. *Front. Plant Sci*., 7: 1707.

Harpke, D., and A. Peterson. 2006. Non-concerted ITS evolution in Mammillaria (Cactaceae). *Mol. Phylogen. Evol*., 41(3): 579-593.

Hebert, P.D., E.H. Penton, J.M. Burns, D.H. Janzen and W. Hallwachs. 2004. Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. *Proceedings of the National Academy of Sciences*., 101(41): 14812-14817.

Hebert, P.D.N., A. Cywinska, S.L. Ball and J.R. deWaard. 2003. Biological identifications through DNA barcodes. *Proceedings. Biological Sciences*., 270(1512): 313-321.

Helsen, P., R.A. Browne, D.J. Anderson, P. Verdyck and S.Van Dongen. 2009. Galapagos' *Opuntia* (prickly pear) cacti: extensive morphological diversity, low genetic variability. *Biol. J. Linn. Soc*., 96(2): 451-461.

Hernández, H.M., C. Gómez-Hinostrosa and B. Goettsch. 2004. Checklist of Chihuahuan Desert Cactaceae. *Harvard Pap. Bot*., 9: 51-68.

Hernández-Hernández, T., H.M. Hernández, J.A. De-Nova, R. Puente, L.E. Eguiarte and S. Magallón. 2011. Phylogenetic relationships and evolution of growth form in Cactaceae (Caryophyllales, Eudicotyledoneae). *Amer. J. Bot.*, 98(1): 44-61.

Hollingsworth, M.L., A. Andra Clark, L.L. Forrest, J. Richardson, R.T. Pennington, D.G. Long, R. Cowan, M.W. Chase, M. Gaudeul and P.M. Hollingsworth. 2009. Selecting barcoding loci for plants: evaluation of seven candidate loci with species-level sampling in three divergent groups of land plants. *Mol. Ecol. Resour*., 9(2): 439-457.

Hollingsworth, P.M., S.W. Graham and D.P. Little. 2011. Choosing and using a plant DNA barcode. *PloS One*., 6(5): e19254-e19254.

Hunt, D., N. Taylor and G. Charles. 2006. *The New Cactus Lexicon, Volumes I and II*. DH Books, Milborne Port.

Ismail, M., A. Ahmad, M. Nadeem, M.A. Javed, S.H. Khan, I. Khawaish, A.A. Sthanadar, S.H. Qari, S.M. Alghanem and K.A. Khan. 2020. Development of DNA barcodes for selected Acacia species by using *rbcL* and *matK* DNA markers. *Saudi J. Biol. Sci*., 27(12): 3735-3742.

Jiménez-Barron, O., R. García-Sandoval, S. Magallón, A. García-Mendoza, J. Nieto-Sotelo, E. Aguirre-Planter and L.E. Eguiarte. 2020. Phylogeny, diversification rate, and divergence time of *Agave sensu lato* (Asparagaceae), a group of recent origin in the process of diversification. *Front. Plant Sci*., 1651.

Kang, Y., Z. Deng, R. Zang and W. Long. 2017. DNA barcoding analysis and phylogenetic relationships of tree species in tropical cloud forests. *Sci. Rep*., 7(1): 1-9.

Keller, A., F. Förster, T. Müller, T. Dandekar, J. Schultz and M. Wolf. 2010. Including RNA secondary structures improves accuracy and robustness in reconstruction of phylogenetic trees. *Biol. Direct*., 5(1): 1-12.

Kimura, M. 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol*., 16(2): 111-120.

Köhler, M., M. Reginato, T.T. Souza-Chies and L.C. Majure. 2020. Insights into chloroplast genome evolution across Opuntioideae (Cactaceae) reveals robust yet sometimes conflicting phylogenetic topologies. *Front. Plant Sci*., 11: 729.

Korotkova, N., T. Borsch, D. Quandt, N.P. Taylor, K.F. Müller and W. Barthlott. 2011. What does it take to resolve relationships and to identify species with molecular markers? An example from the epiphytic Rhipsalideae (Cactaceae). *Amer. J. Bot*., 98(9): 1549-1572.

Kress, W.J., and D.L. Erickson. 2007. A two-locus global DNA barcode for land plants: the coding *rbcL* gene complements the non-coding *trnH-psbA* spacer region. *PLoS One*., 2(6): e508-e508.

Kress, W.J., D.L. Erickson, F.A. Jones, N.G. Swenson, R. Perez, O. Sanjur and E. Bermingham. 2009. Plant DNA barcodes and a community phylogeny of a tropical forest dynamics plot in Panama. *Proceedings of the National Academy of Sciences*., 106(44): 18621-18626.

Kumar, S., G. Stecher, M. Li, C. Knyaz and K. Tamura. 2018. MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol*., 35(6): 1547-1549.

Labra, M., F. Grassi, M. Bardini, S. Imazio, A. Guiggi, S. Citterio, E. Banfi and S. Sgorbati. 2003. Genetic relationships in *Opuntia* Mill. genus (Cactaceae) detected by molecular marker. *Plant Sci*., 165(5): 1129-1136.

Lahaye, R., M. Van der Bank, D. Bogarin, J. Warner, F. Pupulin, G. Gigot, O. Maurin, S. Duthoit, T. G. Barraclough and V. Savolainen. 2008. DNA barcoding the floras of biodiversity hotspots. *Proceedings of the National Academy of Sciences*., 105(8): 2923-2928.

Li, X., Y. Yang, R.J. Henry, M. Rossetto, Y. Wang and S. Chen. 2015. Plant DNA barcoding: from gene to genome. *Biol. Rev*., 90(1): 157-166.

Majure, L.C., and R. Puente. 2014. Phylogenetic relationships and morphological evolution in *Opuntia* s. str. and closely related members of tribe Opuntieae. *Succ. Plant Res*., 8: 9-30.

Maloukh, L., A. Kumarappan, M. Jarrar, J. Salehi, H. El-Wakil and T.V.R. Lakshmi. 2017. Discriminatory power of *rbcL* barcode locus for authentication of some of United Arab Emirates (UAE) native plants. 3 *Biotech*., 7(2): 1-7.

Marghali, S., I. Fadhlaoui, M. Gharbi, N. Zitouna and N. Trifi-Farah. 2015. Utility of ITS2 sequence data of nuclear ribosomal DNA: Molecular evolution and phylogenetic reconstruction of *Lathyrus* spp. *Sci. Hort*., 194: 313-319.

Meier, R., G. Zhang and F. Ali. 2008. The use of mean instead of smallest interspecific distances exaggerates the size of the "barcoding gap" and leads to misidentification. *Sys. Biol*., 57(5): 809-813.

Meier, R., K. Shiyang, G. Vaidya and P.K.L. Ng. 2006. DNA barcoding and taxonomy in Diptera: a tale of high intraspecific variability and low identification success. *Sys. Biol*., 55(5): 715-728.

Meyer, C.P., and G. Paulay. 2005. DNA barcoding: error rates based on comprehensive sampling. *PLoS Biol*., 3(12): e422-e422.

Moore, A.J., J.M. De Vos, L.P. Hancock, E. Goolsby and E.J. Edwards. 2018. Targeted enrichment of large gene families for phylogenetic inference: phylogeny and molecular evolution of photosynthesis genes in the Portullugo clade (Caryophyllales). *Sys. Biol*., 67(3): 367-383.

Morais da Silva, G., A. de Santana Lopes, T. Gomes Pacheco, K. Lima de Godoy Machado, M.C. Silva, J.D. de Oliveira, V.A. de Baura, E. Balsanelli, E. Maltempi de Souza and F. de Oliveira Pedrosa. 2021. Genetic and evolutionary analyses of plastomes of the subfamily Cactoideae (Cactaceae) indicate relaxed protein biosynthesis and tRNA import from cytosol. *Braz. J. Bot*., 44(1): 97-116.

Naciri, Y., S. Caetano and N. Salamin. 2012. Plant DNA barcodes and the influence of gene flow. *Mol. Ecol. Resour*., 12(4): 575-580.

Novoa, A., J.J. Le Roux, D.M. Richardson and J.R.U. Wilson. 2017. Level of environmental threat posed by horticultural trade in Cactaceae. *Conser. Biol*., 31(5): 1066-1075.

Orr, T.J., S.D. Newsome, and B.O. Wolf. 2015. Cacti supply limited nutrients to a desert rodent community. *Oecologia*., 178(4): 1045-1062.

Parmentier, I., J. Duminil, M. Kuzmina, M. Philippe, D.W. Thomas, D. Kenfack, G.B. Chuyong, C. Cruaud and O.J. Hardy. 2013. How effective are DNA barcodes in the identification of African rainforest trees? *PLoS One*., 8(4): e54921-e54921.

Phillips, J.D., D.J. Gillis and R.H. Hanner. 2022. Lack of Statistical Rigor in DNA Barcoding Likely Invalidates the Presence of a True Species' Barcode Gap. *Front. Ecol. Evol*., 10: 859099.

Piredda, R., M.C. Simeone, M. Attimonelli, R. Bellarosa and B. Schirone. 2011. Prospects of barcoding the Italian wild dendroflora: oaks reveal severe limitations to tracking species identity. *Mol. Ecol. Resour*., 11(1): 72-83.

Roy, S., A. Tyagi, V. Shukla, A. Kumar, U.M. Singh, L.B. Chaudhary, B. Datt, S.K. Bag, P.K. Singh and N.K. Nair. 2010. Universal plant DNA barcode loci may not work in complex groups: a case study with Indian *Berberis* species. *PloS One*., 5(10): e13674-e13674.

Saitou, N., and M. Nei. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol*., 4(4): 406-425.

Schultz, J., and M. Wolf. 2009. ITS2 sequence–structure analysis in phylogenetics: a how-to manual for molecular systematics. *Mol. Phylogen. Evol*., 52(2): 520-523.

Sharma, R., H.R. Mahla, T. Mohapatra, S.C. Bhargava and M.M. Sharma. 2003. Isolating plant genomic DNA without liquid nitrogen. *Plant Mol. Biol. Report*., 21(1): 43-50.

Shneyer, V. S., and A.V. Rodionov. 2019. Plant DNA Barcodes. *Biol. Bull. Rev.*, 9(4): 295-300.

Simeone, M.C., R. Piredda, A. Papini, F. Vessella and B. Schirone. 2013. Application of plastid and nuclear markers to DNA barcoding of Euro-Mediterranean oaks (*Quercus*, Fagaceae): problems, prospects and phylogenetic implications. *Bot. J. Linn. Soc*., 172(4): 478-499.

Smith, M.A., N.E. Woodley, D.H. Janzen, W. Hallwachs and P.D. Hebert. 2006. DNA barcodes reveal cryptic host-specificity within the presumed polyphagous members of a genus of parasitoid flies (Diptera: Tachinidae). *Proceedings of the National Academy of Sciences*., 103(10): 3657-3662.

Tamura, K., D. Peterson, N. Peterson, G. Stecher, M. Nei and S. Kumar. 2011. MEGA5: Molecular Evolutionary Genetics Analysis Using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. *Mol. Biol. Evol*., 28(10): 2731-2739.

Tapia, H.J., M.L. Bárcenas-Argüello, T. Terrazas and S. Arias. 2017. Phylogeny and circumscription of *Cephalocereus* (Cactaceae) based on molecular and morphological evidence. *Sys. Bot*., 42(4): 709-723.

Tenorio-Escandón, P., A. Ramírez-Hernández, J. Flores, J. Juan-Vicedo and A.P. Martínez-Falcón. 2022. A systematic review on *Opuntia* (Cactaceae; Opuntioideae) flower-visiting insects in the world with emphasis on Mexico: Implications for biodiversity conservation. *Plants*, 11(131): 1-17.

Thompson, J.D., D.G. Higgins, and T.J. Gibson. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucl. Acids Res*., 22(22): 4673-4680.

van Velzen, R., E. Weitschek, G. Felici and F.T. Bakker. 2012. DNA barcoding of recently diverged species: relative performance of matching methods. *PloS One*., 7(1): e30490-e30490.

Villalobos-Barrantes, H.M., B.M. Meriño, H.E. Walter and P.C. Guerrero. 2022. Independent Evolutionary Lineages in a Globular Cactus Species Complex Reveals Hidden Diversity in a Central Chile Biodiversity Hotspot. *Genes*., 13(2): 240.

Villarreal-Quintanilla, J.A., J.A. Bartolomé-Hernández, E. Estrada-Castillón, H. Ramírez-Rodríguez, and S.J. Martínez-Amador. 2017. El elemento endémico de la flora vascular del Desierto Chihuahuense. *Acta Botanica Mexicana*., 2017(118): 65-96.

Walker, J.F., Y. Yang, T. Feng, A. Timoneda, J. Mikenas, V. Hutchison, C. Edwards, N. Wang, S. Ahluwalia and J. Olivieri. 2018. From cacti to carnivores: Improved phylotranscriptomic sampling and hierarchical homology inference provide further insight into the evolution of Caryophyllales. *Amer. J. Bot*., 105(3): 446-462.

Wiemers, M., and K. Fiedler. 2007. Does the DNA barcoding gap exist?–a case study in blue butterflies (Lepidoptera: Lycaenidae). *Front. Zool*., 4: 1-16.

Yan, L., J. Liu, M. Möller, L. Zhang, X. Zhang, D. Li and L. Gao. 2015. DNA barcoding of *Rhododendron* (Ericaceae), the largest Chinese plant genus in biodiversity hotspots of the Himalaya–Hengduan Mountains. *Mol. Ecol. Resour*., 15(4): 932-944.

Yang, J., L. Vázquez, X. Chen, H. Li, H. Zhang, Z. Liu and G. Zhao. 2017. Development of chloroplast and nuclear DNA markers for Chinese oaks (*Quercus* subgenus *Quercus*) and assessment of their utility as DNA barcodes. *Front. Plant Sci*., 8: 816.

Yesson, C., R.T. Barcenas, H.M. Hernandez, M. De La LUZ RUIZ-MAQUEDA, A. Prado, V.M. Rodriguez and J.A. Hawkins. 2011. DNA barcodes for Mexican Cactaceae, plants under pressure from wild collecting. *Mol. Ecol. Resour*., 11(5): 775-783.

Yu, J., J. Xue and S. Zhou. 2011. New universal *matK* primers for DNA barcoding angiosperms. *J. Sys. Evol*., 49(3): 176-181.

Zhang, Y., W. Zhang, Y. Liu, J. Wang, G. Wang and Y. Liu. 2016. Development of monoclonal antibody-based sensitive ELISA for the determination of Cry1Ie protein in transgenic plant. *Anal. Bioanal. Chem*., 408(28): 8231-8239.

(Received for publication 20 March 2023)