



GENETICS AND CYTOLOGY

INTERNATIONAL JOURNAL DEVOTED TO GENETICAL
AND CYTOLOGICAL SCIENCES

Published by

THE EGYPTIAN SOCIETY OF GENETICS

Volume 54

July 2025

No. 2

Molecular Phylogeny of Six Species Belonging to Cactaceae and Aizoaceae Families using Start Codon Targeted (SCoT) Polymorphism and *in silico* matK Barcode Analysis

MOHAMED A. M. ATIA^{1,*}, TAGHREED KH. ABDELMONEIM¹, ALI M. EL-HALAWANY², AHMED R. HAMED³, RANA M. IBRAHIM², ARWA SALAH¹, AMIRA NASRALLAH⁴, MAHMOUD A. BASRY⁵, HOSSAM M. ZAKARIA⁶, SEHAM S. EL-HAWARY²

¹ Genome Mapping Department, Agricultural Genetic Engineering Research Institute (AGERI), Agricultural Research Center (ARC), 12619, Giza, Egypt

² Pharmacognosy Department, Faculty of Pharmacy, Cairo University, Kasr El Aini Street, 11562 Cairo, Egypt

³ Chemistry of Medicinal Plants Department, National Research Centre, 33 El-Bohouth St., Dokki, Giza 12622, Egypt

⁴ Department of Botany and Microbiology, Faculty of Science, Cairo University, Giza, Egypt

⁵ Plant Molecular Biology Department, Agricultural Genetic Engineering Research Institute (AGERI), Agricultural Research Center (ARC), 12619, Giza, Egypt

⁶ Plant Gene Transfer Department, Agricultural Genetic Engineering Research Institute (AGERI), Agricultural Research Center (ARC), 12619, Giza, Egypt

***corresponding author** Prof. Dr. Mohamed A. M. Atia Email: matia@matia.org

Abbreviations: PCA, Principal Component Analysis

eywords: Halophytes, Cactaceae, SCoT, DNA Barcoding, matK, *In Silico* Analysis

Cactaceae is the second-largest Neotropical family of semi-endemic dicotyledonous angiosperm plants (Hunt *et al.*, 2006). They can grow in different habitats around the world, including Australia, India, the Mediterranean Basin, the Middle East, and South Africa regions. Cacti can grow efficiently under various challenging environmental conditions (Hernández-Hernández *et al.*, 2014). *Cactaceae* family has many unique physiological and morphological characteristics such as: (1) contracted shoots modified into areoles, (2) leaves transformed into spines shoot apical meristems, and (3) sunken ovaries in the receptacles, which are covered with bracts and areoles (Gibson and Nobel, 1986; Boke, 1944 and Osuna-Martínez *et al.*, 2014).

For decades, Cacti have been fascinated by botanists interested in their taxonomy. *Cactaceae* family includes many varieties, many of which have commercial value as ornamental plants and are illegally harvested from their natural habitats (Goettsch *et al.*, 2015; Pérez-Molphe-Balch *et al.*, 2015; 1944 and Hernández-Hernández *et al.*, 2011). Cactus is primarily used as folk medicine in some rural areas to treat diseases such as measles, viruses, worms, and urinary diseases. Cacti have a therapeutic effect against cancer cells (Boshara, 2014). Cacti are one of the most important plant groups listed as threatened under national

and international agreements (Arias *et al.*, 2005 and Hernández-Oria *et al.*, 2007).

The *Cactaceae* family contains approximately 1850 species in 139 genera, under the Caryophyllales order (Throne, 2000). This plant order includes several families with commonly shared morphological features such as *Basellaceae*, *Aizoaceae*, *Portulacaceae*, and *Didiereaceae*, as *Cactaceae*. Based on plant and floral morphology, four main subfamilies were described family of *Cactaceae*: *Opuntioideae*, *Cactoideae*, *Maihuenioideae*, and *Pereskioideae* (Gibson and Nobel, 1986; Boke, 1944 and Osuna-Martínez *et al.*, 2014). During the last decade, the relationships between these families were distinguished using molecular phylogenetic studies (Cuénoud *et al.* 2002). The *Cactoideae* is the main subfamily of seven tribes: *Cacteae*, *Cereeae*, *Echinocereae*, *Hylocereeae*, *Notocacteae*, *Rhipsalideae*, and *Trichocereae* (Hunt *et al.*, 2006). Within the *Cacteae* tribe, *Ferocactus* is the third largest genus in terms of the number of species (Gibson and Nobel, 1986 and Boke, 1944). It contains 25-30 species. The genus is known as barrel-Cacti, usually, an unbranched stem carries a dense cover of spines and colorful flowers that are native to Mexico and the USA (Brinker, 2009). *Ferocactus* was assumed to take a crucial phylogenetic role in the *Cacteae* classifications (Hunt *et al.*, 2006).

Family *Aizoaceae* consists of 143 genera with approximately 2300 species (Said *et al.*, 2019). Most of the plant extracts of these species contain several secondary metabolites that exhibit many powerful biological activities such as antibacterial, hypolipidemic, anti-inflammatory, and antiplatelet agents (Osuna-Martínez *et al.*, 2014; Pirie *et al.*, 2014 and Said *et al.*, 2019). Previous studies showed that the *Aizoaceae* family is only distantly related to *Cactaceae* (Throne, 2000). *Cactaceae*'s rapid diversification is a significant phylogenetic issue (Ritz *et al.*, 2007 and Realini *et al.*, 2015).

Since the taxonomy of these two families remains in a somewhat confusing status due to several factors (Ritz *et al.*, 2007). Therefore, it is crucial to utilize the modern molecular markers systems to facilitate the taxonomy issues between and within these two families.

In recent years, novel molecular marker techniques have been developed and utilized in various plant genetic studies. Among them, a new trend of marker systems called gene-targeting markers or functional markers have been reported (Pocazi *et al.*, 2013).

Comparing to the other popular molecular marker techniques such as; Microsatellites, AFLP, CDDP, RAPD and ISSR; gene-targeting markers are positioned in or near of certain genes, therefore, it has become more effective and powerful in detecting the genetic variations in phylogenetics and diversity

studies (Mokhtar and Atia, 2019; Adawy and Atia, 2014; Ahmed *et al.*, 2016 and Pocazi *et al.*, 2013).

At the end of 20th century, Collard and his colleague Mackill successfully developed one of the most novel and effective functional markers in plant. This marker system called Start Codon Targeted (SCoT) Polymorphism which depends on the short-conserved flanking region around the ATG start codon in plant genes (Collard and Mackill, 2009). SCoT amplicons are found to be highly reproducible due to uses of single 18-mer primer with an annealing temperature of 50°C (Collard and Mackill, 2009). DNA analysis using SCoT markers has proved to be an efficient and inexpensive way to provide molecular data to assess genetic diversity, and it has been used successfully to determine genetic relationships for many plants (Atia *et al.*, 2021).

In plants, DNA barcoding is one of the modern and powerful approaches used for discrimination between species, specimen's identification, phylogenetic analyses and molecular taxonomy. Among the seven leading plastid regions (*atpF–atpH* spacer, *matK* gene, *rbcL* gene, *rpoB* gene, *rpoC1* gene, *psbK–psbI* spacer, and *trnH–psbA* space) regularly used in plant DNA barcoding, only two regions were reported as the standard barcode for land plants *matK* gene and *rbcL* gene. Based on the discrimination analysis of 397 plant species, The Consortium for the Barcode of Life

(CBOL) recommends the maturase K (*matK*) and ribulose-1,5-bisphosphate Carboxylase (*rbcL*) loci as the core 2-locus barcode set for plant routine work and identification of unknown specimens (Nyffeler, 2002 and Hernández-Hernández *et al.*, 2011).

matK is one of the most quickly evolving plastid barcodes due to its universality and, high levels of discrimination among close plant species. Moreover, recent studies revealed that *matK* region is highly preferred as a universal barcode in plants because of interspecific divergence, high evolutionary rate, and low transition/transversion rate (Hernandez-Oria *et al.*, 2007). Recently, data mining of DNA barcodes in plants is an outstanding step that directly aids in species identification. The tree-based phylogeny is a popular analysis to estimate the relationships among close species as well as the time of divergence between a group of plants. Finally, it is noteworthy that to generate signatures for species identification, it will be essential to find the nucleotide patterns and their precise positions in the *matK* barcode gene can discriminate between the close species (Nyffeler, 2002).

Therefore, this study aims to assess the genetic diversity among six halophyte plants belonging to families Aizoaceae (namely, *Aptenia cordifolia*, *Glottiphyllum linguiforme*, *Carpobrotus edulis*) and *Cactaceae* (namely, *Ferocactus glaucescens*, *F. pottsii*, *F.*

Herrerae) using SCoT-PCR system as gene-targeting marker technique in concomitant with *in silico* comparative analysis of *matK* barcode gene signatures. The combination of molecular markers analysis and *in silico* barcode analysis is crucial for the identification and authentication of these six halophytes.

MATERIAL AND METHODS

Plant materials

Three *Ferocactus* plants belonging to family *Cactaceae* (*Ferocactus pottsii* (Salm-Dyck) Backeb, *Ferocactus glaucescens* (DC.) Britton & Rose, and *F. Herrerae* J.G. Ortega) and three *Aizoaceae* plants (*Glottiphyllum linguiforme* (L.) N.E. Br., *Aptenia cordifolia* var. *variegata* (L.f.) Schwantes, and *Carpobrotus edulis* (L.) N.E. Br), were obtained from Helal Cactus Farm, Shibin El-Qanater, Qalyubia Governorate, Egypt. Botany specialist Prof. Dr. Mohamed El-Gebaly further confirmed the plant's identities. Voucher specimens of the plants were kept at the Herbarium of the Department of Pharmacognosy, Faculty of Pharmacy, Cairo University. Details about the collected samples' names and their *matK* gene and protein accession numbers were depicted in Table (1).

Molecular Analysis

Plant DNA extraction

Genomic DNA was extracted from fresh parts (100 mg) of the collected six

plants using a DNAeasy Plant Mini Kit (QIAGEN, Santa Clarita, CA) and according to the manufacturer's protocol. The DNA was quantified with NanoDrop Spectrophotometer (Thermo Fisher Scientific Inc.). The DNA concentrations were adjusted in all samples to 10 ng/μL for subsequent molecular analyses.

Start Codon Targeted (SCoT) Polymorphism Analysis

SCoT PCR amplification was carried out according to the procedure described by (Atia *et al.* 2017). A set of 10 SCoT primers were tested against the six species. PCR was carried out in 25 μL reaction mixtures containing 1X PCR buffer, 1.5 mM MgCl₂, 0.2 μM of each dNTPs, 1 μM of primer, 1U Go-Taq Flexi polymerase (Promega) and 25 ng genomic DNA.

Thermocycling amplification was performed in a Perkin-Elmer/ Gene Amp PCR system 9700 (PE Applied Biosystem). The amplification was programmed at 94°C for 5 min as an initial denaturation cycle, followed by 35 cycles, each cycle comprised of (94°C for 1min, 50°C for 1min, then 72°C for 90 s) with a final extension at 72°C for 7 min. The amplification products were resolved by electrophoresis in a 1.5% agarose gel containing ethidium bromide (0.5ug/mL) in 1X TBE buffer. A 100 bp plus DNA ladder was used as molecular size standards. PCR products were visualized on UV light and photographed using a Gel Doc™ XR+ System with Image Lab™ Software (Bio-Rad®).

Data analysis

For molecular data analysis, the generated/amplified bands were scored visually. To reduce errors, only the clear and distinguishable bands were scored. The bands were scored as present (1) or absent (0) to create the binary data set. Polymorphism percentage was calculated by dividing the number of amplified polymorphic bands by the total number of amplified bands by the same primer or primer combination. A similarity matrix was constructed to measure genetic distances between pairs of plants; these distances were estimated between all possible pairs. The pairwise comparisons were made between plants based on the Jaccard similarity coefficient (Ni Wattanakul *et al.*, 2013). The genetic similarity estimate (GS) between each pair of plants was calculated using the expression $GS = a/(n-d)$, in which *a* is the number of positive coincidences; *n* is the total number of fragments, and *d* is the number of negative coincidences. The genetic distances (GD) between pairs of plants were estimated by $GD = 1 - GS$.

A dendrogram was generated by cluster analysis using the unweighted pair group method of the arithmetic averages (UPGMA) for all different marker systems. Principal component analysis (PCA) was conducted according to Wold *et al.* (1987).

The efficiency of the SCoT primers was determined by calculating the following parameters: Expected

Heterozygosity ($H = 1 - \sum p_i^2$ according to Liu (2017)), Polymorphism Information Content ($PIC = 1 - \sum p_i^2 - \sum p_j^2$ according to Botstein *et al.*, 1980), Effective Multiplex Ratio ($E = n \beta$ according to Powell *et al.* (1996)), Marker Index ($MI = E H_{av}$ according to Powell *et al.* (1996)), Mean Heterozygosity ($H_{av} = \sum H_n / n_p$ according to Powell *et al.* (1996)), Discriminating Power ($D = 1 - C$ according to Tessier *et al.* (1999)), Resolving Power ($R = \sum I_b$ according to Prevost and Wilkinson (1999)) using the iMEC online tool (Amiryousef *et al.*, 2018).

***In silico* matK Barcode Analysis**

We retrieved the available *matK* gene sequences as well as *matK* protein sequences from the NCBI database. Out of the six species, only *matK* sequences of five species were found, while we didn't find any published DNA or protein sequences for the *Glottiphyllum linguiform* species. The accession numbers of downloaded *matK* DNA and protein sequences were listed in Table (1). According to our survey, we noticed that the NCBI only contains one barcode sequence to two sequences at a maximum of each one of the five species.

The *matK* DNA barcode gene sequences were aligned with the Clustal W multiple sequence alignment program. The evolutionary history was inferred by using the Maximum Likelihood method based on the Tamura 3-parameter model?. Statistical support for each constructed

tree was provided by two statistical data analysis as bootstrapping (1000 replications) and pairwise distance. And the maximum likelihood of transition/transversion bias was calculated by MEGA Software version 6 (Tamura *et al.*, 2013).

The protein sequences of the *matK* were aligned with MUSCLE (<https://www.ebi.ac.uk/Tools/msa/muscle>) and edited manually in BioEdit v7.1.3.038. The analyses of genetic pairwise distances were computed in MEGA X with Kimura-2-parameter (K2P) model. Differences between intra- and inter-specific distances with four single barcodes were evaluated using pairwise distance matrix in MEGA X software.

The species discriminatory efficacy of each *matK* gene and protein barcode candidates was assessed with four tree-based methods, which include the the Minimum Evolution method (ME), the Maximum Likelihood method (ML), the Neighbor-Joining method (NJ), and the Maximum Parsimony method (MP). The NJ methods of all accessions were conducted using MEGA X. The reliability of node was supported by a bootstrap test of 1,000 pseudo-replicates with K2P distance parameter. For ML analysis, the phylogenetic trees were constructed in MEGA X. The clade support was assessed using ML with thorough bootstrap analyses, run 10 times starting from random seeds under GTRGAMMA model, and 1,000 non-

parametric bootstrap values. The maximum parsimony (MP) methods from the MEGA X software with 1000 bootstrap replicates (Florida State University).

RESULTS AND DISCUSSION

SCoT Analysis

The DNA of the six species were amplified using 10 SCoT primers to investigate the genetic diversity and screen the degree of polymorphism between them. The ten SCoT primers yielded a total of 144 scorable bands with an average of 14.4 band/primer, out of which 86 (59.7%) were found to be polymorphic with an average of 8.6 bands/primer (Table 2). The number of amplified DNA fragments per primer ranged from 11 bands (primer SCoT-9) to 18 bands (primer SCoT-8) and the percentage of polymorphism ranged from 53.8% (primers SCoT-3) to 73.3% (primer SCoT-6) (Table 2).

A wide range of the expected heterozygosity and polymorphism information content values were observed; the values were ranged between 0.2 to 0.5 with an average of 0.3 per primer. The primers SCoT-7 and SCoT-8 showed the highest PIC and H among all used primers. The effective multiplex ratio values were very high with an average of 10.18. On the other hand, the marker index values were very low with an average of 0.043. The discriminating power values were ranged from 0.2 to 0.7 with an average of 0.48. The primers

SCoT-7 and SCoT-8 D values are considered moderately high; 0.694 and 0.761, respectively. Also, the resolving power values were ranged between 0.6 to 6 with an average value of 0.296. The primer SCoT-6 showed the best-resolving power among all the used SCoT primers as summarized in Table (2).

The correlation coefficient analysis for the measured efficiency parameters revealed that there was a robust significant linear relationship between (PIC and H), (D and H) and (D and PIC) using the SCoT primer system ($r = 1$). Similarly, significant relationships were found between (E and H), (E and PIC), (H.av and H), (H.av and PIC), (H.av and E), (MI and H), (MI and PIC), (MI and H.av), (D and E), (D and H.av) and (D and MI) ($r = -0.7173, -0.7173, 0.79394, 0.79394, -0.8997, 0.63636, 0.63636, 0.84242, -0.7173, 0.79394, 0.63636$, respectively) as shown in Table (3).

Molecular Phylogeny and PCA Analysis

To estimate the genetic distances between the six species, Jaccard's similarity matrix was used (Table 4) and a dendrogram was generated based on UPGMA analysis of SCoT data (Fig. 1a). The dendrogram comprised three clusters, the first cluster grouped the species *F. pottsii*, *F. herrerae*, and *F. glaucescens* with the first two species being more genetically similar. The second cluster grouped two species *C. edulis* and *A. cordifolia* var. *variegata*. Meanwhile, the

third cluster comprised only one species *G. linguiforme* which was more diverged among all tested species.

On the other hand, the PCA analysis of SCoT data exhibited a similar consistent result as compared to the cluster analysis (dendrogram). The PCA results observed that the grouping remains similar as obtained by the cluster analysis (Fig. 1b). Although we have employed only ten SCoT primers, the total obtained loci were 144 with 59.7% polymorphic loci. Our clustering and PCA results successfully reflected a discrimination power within and between the *Aizoaceae* and *Ferocactus* species with different levels of divergence.

In silico matK Barcode Analysis

Based on the phylogenetic analysis of the *matK* gene sequences, the five species *F. glaucescens*, *F. herrerae*, *F. pottsii*, *A. cordifolia* var. *variegata*, and *C. edulis* were clustered at two major clades. An overview of the *matK* gene accessions alignment was represented in Fig. (2) with fewer instances of nucleotide differences between members belonging to even the family Cactaceae (*F. glaucescens*, *F. pottsii*, and *F. Herrerae*) as well as family Aizoaceae (*Aptenia cordifolia*, and *Carpobrotus edulis*) than members of different genera. Four different tree-based methods; the Minimum Evolution method (ME), the Maximum Likelihood method (ML), the Neighbor-Joining method (NJ), and the Maximum Parsimony method (MP) were used to reconstruct the topology of the

tree for the five species represented in Fig. (3). The ME optimal tree with the sum of branch length = 0.15791219, the ML tree with the highest log likelihood (-1554.33), the NJ tree with the sum of branch length = 0.15791219, and the MP tree with length = 109 are shown. All the methods resulted in two major clades; clade-I included *F. glaucescens*, *F. herrerae*, and *F. pottsii*. Meanwhile, clade-II included *A. cordifolia* var. *variegata* and *C. edulis*. For the ML and MP trees, the species *F. pottsii* and *F. herrerae* were more close to each other than the *F. glaucescens* species with a high bootstrap value (99%). On the other hand, for the ME and NJ trees the species; *F. pottsii* and *F. glaucescens* were more close to each other than the *F. herrerae* species with a moderate bootstrap value (86%). The distance matrices score values are summarized in Table 6. Notably, the small values of the distance matrices score among the Cactaceae family species; *F. glaucescens*, *F. herrerae*, and *F. pottsii* (ranged from 0.00320 to 0.01562).

Rates of different transitional substitutions and those of transversional substitutions were also estimated (Table 5). The relative values of instantaneous rates were considered as the sum of the values made equal to 100 and the nucleotide frequencies were A=26.91%, T/U=39.16%, C = 18.66% and G = 14.66%.

A phylogenetic analysis of the *matK* proteins data set also yielded two

main clades; clade-I included *F. glaucescens*, *F. herrerae*, and *F. pottsii*, and clade-II contained *A. cordifolia* var. *variegata* and *C. edulis* (Fig. 4). The evolutionary history was inferred using the Neighbor-Joining method (NJ) and Jones-Taylor-Thornton model (JTT). The NJ optimal tree with the sum of branch length = 0.30065583 is shown in Fig. 4a. Meanwhile, the JTT optimal tree with the highest log likelihood (-2010.51) is shown in Fig. 4b. For the NJ tree the species: *F. herrerae* and *F. glaucescens* were more close to each other than the *F. pottsii* with a low bootstrap value (57%). Furthermore, the JTT tree inferred that the species: *F. pottsii* and *F. herrerae* were more close to each other than the *F. glaucescens* with a moderate bootstrap value (77%).

The computed distance matrices score values are summarized in Table (7). Remarkably the small values of the distance matrices score among the *Cactaceae* family species; *F. glaucescens*, *F. herrerae*, and *F. pottsii* (ranged from 0.00452 to 0.05884). Also among the *Aizoaceae* family species; *A. cordifolia* var. *variegata* and *C. edulis* (0.02001 and 0.09453). While the distances among the species of the *Aizoaceae* and *Cactaceae* families showed relatively high values (ranged from 0.01853 to 0.32542).

Genetic diversity analysis

The *Aizoaceae* and *Cactaceae* are two huge plant groups that involve a massive number of species. The taxonomy of these two families remains

in a somewhat confusing status due to several factors (Ritz *et al.*, 2007). During the last decades, several molecular taxonomy studies based on nuclear and plastid DNAs markers have been reported (Griffith *et al.*, 2009; Edwards *et al.*, 2005; Powell *et al.*, 2018 and Thiede *et al.*, 2007) the other hand, the applications of classical molecular markers in fingerprinting and genetic diversity studies on these two families were very limited and completely achieved with aid of random marker techniques such as RAPD, ISSR, microsatellite and AFLP (Clark-Tapia *et al.*, 2005; Bentley *et al.*, 2015; Castro-Felix *et al.*, 2014 and Buys *et al.*, 2008). Indeed, the SCoT PCR system was employed in the current study for many reasons includes its gene-targeting mechanism and simplicity. Comparing to previous studies in which different arbitrary marker techniques such as RAPD, DAF...etc. were employed, utilization of the SCoT system in this study represents the first report to investigate and analyze the genetic variations within and between six halophytes belonging to *Aizoaceae* and *Cactaceae* families using SCoT-PCR as modern gene targeting marker system.

The families *Cactaceae* and *Aizoaceae* encompass a large number of species that manifest a wide diversity. *Cactaceae* family approximately comprises 130 genera and 2000 species that were originally native to the New World (Goettsch *et al.*, 2015 and Pérez-Molphe-Balch *et al.*, 2015). Also, the family *Aizoaceae* includes approximately

143 genera with about 2300 species. There is a gap in the molecular phylogeny of these two families, and still in a disconcerting status. This lack of knowledge enabled our study to bridge such gaps with the aid of molecular phylogeny of these families. Recently, the molecular phylogeny using novel molecular marker techniques have been developed and utilized in various plant genetic studies such as; Microsatellites, AFLP, CDDP, RAPD, and ISSR (Mokhtar and Atia, 2019; Adawy and Atia, 2014; Ahmed *et al.*, 2016 and Pocazi *et al.*, 2013). One of the most efficient and novel marker systems is Start Codon Targeted (SCoT) Polymorphism, which is weighed as a reproducible and cost-effective way to evaluate the genetic diversity and genetic relationships for many plants (Collard and Mackill, 2009).

The molecular analysis of the six species was amplified using 10 SCoT primers (Table (2) to investigate the genetic diversity and screen the degree of polymorphism between them. Amplification products of the six species with the primers yielded a total of 144 scorable bands, out of which 86 (59.7%) were found to be polymorphic with an average 8.6 bands/Primer (Table 2). The Calculations of the efficiency parameters (H, PIC, EMR, MI, H.av, D, and R) for the primers used revealed engrossing results. The maximum value of H and PIC for binary data is 0.5 because two alleles per locus are presumed, and both are affected by the number and frequency of

the alleles. The high values of the PIC and H for the SCoT-7 and SCoT-8 primers indicate the advanced discriminatory capacity of the SCoT marker system for the identification of Cacti genotypes. A closer inspection of the MI generated highlights the low distinguishing power which is due to a higher effective multiplex ratio component. The combined R-value of the primers also provides a measure of their collective performance for identification determinations. The primer SCoT-6 alone could identify 28 Cacti genotypes, according to the equation of Prevost and Wilkinson (1999; $0.15x + 1.78 = R$, where x is the number of genotypes identified). The D parameter described by Tessier *et al.* (1999) determines the efficiency of the primers in the identification of Cacti accessions. The D parameter describes the probability that two randomly chosen Cacti individuals have different patterns. A higher D parameter (closest to 1) implies a lower probability of confusion between Cacti accessions. The primers SCoT-7 and SCoT-8 D values are considered moderately polymorphic; 0.694 and 0.761, respectively. The analysis of the marker indices revealed that the primers SCoT-6, SCoT-7, and SCoT-8 are more informative than the other SCoT primers.

Molecular Phylogeny and PCA Analysis

It is worth noting that the 10 SCoT primers worked well in clustering the six species of Cacti during the phylogenetic analysis process. Three *species F. pottsii*,

F. herrerae and *F. glaucescens* were clustered into a sub-clade with the first two species being more genetically similar. The reason could be that *F. pottsii* and *F. herrerae* species may have a closer affiliation to each other than to *F. glaucescens*. In addition, the second cluster grouped two species (*C. edulis* and *A. cordifolia* var. *variegata*) forming a separate clade.

In silico *matK* Barcode Analysis

In this investigation, the relationship between five medicinal Cacti species based on phylogenetic tree analyses constructed from the *matK* genes and proteins was clarified. Unfortunately, the *Glottiphyllum linguiforme* species has no available sequence data. The phylogenetic analysis for the *matK* gene sequences resulted in two major clades; clade-I included *F. glaucescens*, *F. herrerae*, and *F. pottsii*, and clade-II contained *A. cordifolia* var. *variegata* and *C. edulis* (Fig. 3). By comparing the four trees using the ME, ML, NJ, and MP methods, the ML and MP trees were identical to each other with high supportive bootstrap values. Besides, the ME and NJ trees were identical to each other with high supportive bootstrap values. In a close inspection, the ML and MP trees have a higher bootstrap value (99%) than the ME and NJ trees, for classifying the *F. pottsii* and *F. herrerae* together. In general, the differences in the topologies between the four reconstructed trees were slightly insignificant,

empowering the classification of the Cacti species.

The distance matrices values showed in Table (6) and Table (7), the small values indicate more closely related sequences. The values of distance matrices support the close relationship among the *Cactaceae* family species; *F. glaucescens*, *F. herrerae*, and *F. pottsii* showed low distance matrices values. Also, between the *Aizoaceae* family species; *A. cordifolia* var. *variegata* and *C. edulis* showed low distance matrices values, indicating their close relationships. While a high degree of divergence between the members of the *Cactaceae* and *Aizoaceae* families is very clear in the high values of the distance matrices.

By utilizing the Maximum Likelihood Estimate (MLE) method to calculate the nucleotide substitution matrix of the *matK* gene sequences, the nucleotide frequencies were lower for all nucleotides except nucleotide T/U=39.16%.

Besides, the phylogenetic analysis for the *matK* protein sequences inferred the same major clades as the *matK* gene sequences (Fig. (4)). By comparing the two methods NJ and JTT, the JTT tree was more supportive in clustering *F. pottsii* and *F. herrerae* with higher bootstrap value (77%) than the NJ tree.

In conclusion, our results are in agreement with the previous classification

of both the *Cactaceae* and *Aizoaceae* families, studies showed that the *Aizoaceae* family is only distantly related to *Cactaceae* (Gibson and Nobel, 1986; Boke, 1944 and Osuna-Martínez *et al.*, 2014). As the members of *Ferocactus* clustered together, and the members of the *Aizoaceae* family clustered together.

Authors' Contributions : M. A., T. K., A. S., A. N. collected and analyzed the data, WE and M.A. reviewed and edited this manuscript, M. A., M. B. and R. I. designed this research and drafted this manuscript. All authors contributed to the analysis and evaluation of the results. All authors contributed to the article and approved the submitted version.

Ethical approval: for researches involving animals or humans) Not applicable

Conflicts of Interest: The authors declare no conflict of interest related to this article.

SUMMARY

Halophytes have been used throughout the centuries for their therapeutic properties, while there is a recent increase in global demand for herbal medicines and food supplements. The *Aizoaceae* and *Cactaceae* are two huge plant groups belonging to the halophytes that involve a massive number of species. But, there are no sufficient studies that manipulate the molecular phylogeny of these two families, and they are still in a disconcerting status. Our

study contributes fundamental information for the phylogenetic analysis of the *Aizoaceae* and *Cactaceae* families. Ten SCoT primers were used to study the genetic diversity of six halophyte species; *F. glaucescens*, *F. herrerae*, *F. pottsii*, *G. linguiforme*, *A. cordifolia* var. *variegata*, and *C. edulis*. Based on the analysis of markers indices the primers SCoT-6, SCoT-7 and SCoT-8 have a significant discrimination power and are suitable for the detection of the genetic diversity in the studied species and it may be concluded on the basis of this study that the genetic diversity analysis with SCoT markers was more effective. Besides, our results showed consistency with the previous studies for the classification and clustering of halophytes species. The *in silico* analysis of the *matK* genes and proteins also manifests a similar pattern to the SCoT markers analysis. In conclusion, the *F. glaucescens*, *F. herrerae*, and *F. pottsii* were clustered together as they are members of the *Cactaceae* family, while the *G. linguiforme*, *A. cordifolia* var. *variegata*, and *C. edulis* were clustered together as they are members of the *Aizoaceae* family.

ACKNOWLEDGMENTS

The authors would like to sincerely thank the administration of the Agricultural Genetic Engineering Research Institute (AGERI), as well as the administration of the Agricultural Research Center (ARC), Egypt, for their continued support.

REFERENCES

- Adawy S. S. and Atia M. A., (2014). A multidisciplinary molecular marker approaches to assess the genetic diversity in Egyptian date palm. *J. of Biotechnology and Research*, 4: 1-12.
- Ahmed S. S., Atia M. A., Abd El-Aziz G. H. and Fahmy A. H., (2016). American Journal of Agricultural Science. American J. of Agricultural Science, 3(3): 48-58.
- Arias S., Guzmán U., Mandujano M.C., Soto M. and Golubov J., (2005). Las especies mexicanas de cactáceas en riesgo de extinción. I. Una comparación entre los listados NOM-059-ECOL-2001 (México), La lista Roja (UICN) y CITES. *Cact Suc Mex* 50(4):100-125
- Atia M. A. M., El-Moneim D. A., Abdelmoneim T. K., Reda E. H., Shakour Z. T. A., El-Halawany A. M., El-Kashoury E. S. A., Shams K. A., Abdel-Azim N. S. and Hegazy M. E. F., (2021). Evaluation of genetic variability and relatedness among eight *Centaurea* species through CAAT-box derived polymorphism (CBDP) and start codon targeted polymorphism (SCoT) markers. *Biotechnology & Biotechnological Equipment*, 35(1): 1230-1237.
- Bentley L., Barker N. P. and Dold A. P., (2015). Genetic diversity of the endangered *Faucaria tigrina* (*Aizoaceae*) through ISSR “fingerprinting” using automated fragment detection. *Biochemical Systematics and Ecology*, 58:156-161.
- Boke N. H., (1944). Histogenesis of the leaf and areole in *Opuntia cylindrica*. *American Journal of Botany*, 31(6): 299-316.
- Boshara O. A. A., (2014). Phytochemical Screening for Leaves, Cortex and Pith of the Cactus *Euphorbia trigona* L (Doctoral dissertation, University of Gezira).
- Brinker F., (2009). Prickly pear as food and medicine. *Journal of dietary supplements*, 6(4): 362-376.
- Buys M. H., van Rensburg L. L. J., Mienie C. M. S., Barker N., Burgoyne P. M., Mills L., Van Rensburg L. and Hartmann H., (2008). Applying AFLPs in *Aizoaceae*: The *Delosperma herbeum* complex as a case study. *Biochemical Systematics and Ecology*, 36(2): 92-100.
- Castro-Felix P., Rosas-Espinoza V. C., Díaz-Cárdenas B., Pérez-Valencia L.I., Huerta-Martínez F. M. and Santerre A., (2014). Genetic diversity within a declining natural population of *Ferocactus histrix* (DC) L. *insay. Plant Species Biology*, 29(3): E21-E30.
- Clark-Tapia R., Alfonso-Corrado C., Eguiarte L. E. and Molina-Freaner F., (2005). Clonal diversity and

- distribution in *Stenocereus eruca* (Cactaceae), a narrow endemic cactus of the Sonoran Desert. *American Journal of Botany*, 92(2): 272-278.
- Collard B.C. and Mackill D. J., (2009). Start codon targeted (SCoT) polymorphism: a simple, novel DNA marker technique for generating gene-targeted markers in plants. *Plant molecular biology reporter*, 27(1): 86-93.
- Edwards E. J., Nyffeler R. and Donoghue M. J., (2005). Basal cactus phylogeny: implications of *Pereskia* (Cactaceae) paraphyly for the transition to the cactus life form. *American Journal of Botany*, 92(7): 1177-1188.
- Gibson A. C. and Nobel P. S., (1986). *The cactus primer*. Harvard University Press.
- Goettsch B., Hilton-Taylor C., Cruz-Piñón G., Duffy J. P., Frances A., Hernández, H.M., Inger, R., Pollock C., Schipper J., Superina M. and Taylor N. P., (2015). High proportion of cactus species threatened with extinction. *Nature plants*, 1(10):1-7.
- Griffith M. P. and Porter J. M., (2009). Phylogeny of Opuntioideae (Cactaceae). *International Journal of Plant Sciences*, 170(1):107-116.
- Hernández-Hernández T., Brown J. W., Schlumberger B. O., Eguiarte L. E. and Magallón S., (2014). Beyond aridification: multiple explanations for the elevated diversification of cacti in the New World Succulent Biome. *New Phytologist*, 202(4): 1382-1397.
- Hernández-Hernández T., Hernández H. M., De-Nova J. A., Puente R., Eguiarte L. E. and Magallón S., (2011). Phylogenetic relationships and evolution of growth form in Cactaceae (Caryophyllales, Eudicotyledoneae). *American Journal of Botany*, 98(1): 44-61.
- Hernandez-Oria J. G., Chávez-Martínez R. J. and Sanchez-Martinez E., (2007). Factors for endangered cactaceae in a semi-arid region of the Southern Chihuahuan Desert, Mexico. *Interciencia*, 32(11): 728-734.
- Hunt D. R., Taylor N. P. and Charles G., 2006. *New cactus lexicon*. dh books.
- Liu B. H., (2017). *Statistical genomics: linkage, mapping, and QTL analysis*. CRC press.
- Mokhtar M. M. and Atia M. A. M., (2019). SSRome: an integrated database and pipelines for exploring microsatellites in all organisms. *Nucleic acids research*, 47(D1): D244-D252.
- Niwattanakul S., Singthongchai J., Naenudorn E. and Wanapu S., (2013). March. Using of Jaccard coefficient for keywords similarity. In *Proceedings of the international multiconference of engineers and*

- computer scientists (Vol. 1, No. 6: 380-384).
- Nyffeler R., (2002). Phylogenetic relationships in the cactus family (Cactaceae) based on evidence from *trnK/matK* and *trnL-trnF* sequences. *American Journal of Botany*, 89(2): 312-326.
- Osuna-Martínez U., Reyes-Esparza J. and Rodríguez-Fragoso L., (2014). Cactus (*Opuntia ficus-indica*): A Review on its Antioxidants Properties and Potential Pharmacological Use in Chronic Diseases. *Nat Prod Chem Res* 2: 153. doi: 10.4172/2329-6836.1000153 Page 2 of 8 and protein (0.5-1%); other compounds are only partly known and have not been quantitatively determined [9]. The sugar moiety includes mucilaginous components containing polymers, such as chains of (1-4)-linked β -D-galacturonic acid and R (1-2)-linked L-rhamnose residues [10, 11]. The physiological role of the *Opuntia ficus-indica*.
- Pérez-Molphe-Balch E., Santos-Díaz M. D. S., Ramírez-Malagón R. and Ochoa-Alejo N., (2015). Tissue culture of ornamental cacti. *Scientia Agricola*, 72: 540-561.
- Pirie A. D., Davies N. W., Ahuja K. D., Adams M. J., Shing C. M., Narkowicz C., Jacobson G. A. and Geraghty D. P., (2014). Hypolipidaemic effect of crude extract from *Carpobrotus rossii* (pigface) in healthy rats. *Food and Chemical Toxicology*, 66: 134-139.
- Powell R. F., Magee A. R. and Boatwright J. S., (2018). Decoding ice plants: challenges associated with barcoding and phylogenetics in the diverse succulent family *Aizoaceae*. *Genome*, 61(11): 815-821.
- Powell W., M. Morgante C. Andre M. Hanafey J., Vogel S. Tingey and Rafalski A., (1996). The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. *Molecular Breeding* 2: 225-238.
- Prevost A., and Wilkinson M. J., (1999). A new system of comparing PCR primers applied to ISSR fingerprinting of potato cultivars. *Theoretic and Applied Genetics* 98: 107-112.
- Realini M. F., González G. E., Font F., Picca P. I., Poggio L. and Gottlieb A. M., (2015). Phylogenetic relationships in *Opuntia* (*Cactaceae*, *Opuntioideae*) from southern South America. *Plant Systematics and Evolution*, 301(4): pp.1123-1134.
- Ritz C. M., Martins L., Mecklenburg R., Goremykin V. and Hellwig F. H., (2007). The molecular phylogeny of *Rebutia* (*Cactaceae*) and its allies demonstrates the influence of paleogeography on the evolution

of South American mountain cacti. American Journal of Botany, 94(8): 1321-1332.

Said A. A. E., Attia E. Z., Abdelmohsen U. R. and A Fouad M. A., (2019). Natural products potential of the genus *Aptenia*. Journal of advanced Biomedical and Pharmaceutical Sciences, 2(2): 59-62.

Tamura K., Stecher G., Peterson D., Filipski A. and Kumar S., (2013). MEGA6: molecular evolutionary genetics analysis version 6.0. Molecular biology and evolution, 30(12): 2725-2729.

Tessier C., David J., This P., Boursiquot J. M. and Charrier A., (1999). Optimization of the choice of molecular markers for varietal identification in *Vitis vinifera* L. Theoretical and Applied Genetics, 98(1): 171-177.

Thiede J., Schmidt S. A. and Rudolph B., (2007). Phylogenetic implication Thiede J., Schmidt S. A. and Rudolph B., (2007). Phylogenetic implication of the chloroplast rpoC1 intron loss in the *Aizoaceae* (Caryophyllales). Biochemical Systematics and Ecology, 35(6): 372-380.

Wold S., Esbensen K. and Geladi P., (1987). Principal component analysis. Chemometrics and intelligent laboratory systems, 2(1-3): 37-52.

Atia M. A., Sakr M. M. and Adawy S. S., 2017. Assessing date palm genetic diversity using different molecular markers. In Date Palm Biotechnology Protocols Volume II (pp. 125- 252 142). Humana Press, New York, NY.

Table (1): The scientific names of the studied plants and their code.

In Vitro Data		In Silico Data	
Name of species	Code	matK Gene Acc.	matK Protein Acc.
<i>Ferocactus glaucescens</i>	FG	FN997274	CBV65523.1
<i>Ferocactus pottsii</i>	FP	FN997043	CBV65294
<i>Ferocactus herrerae</i>	FH	FN997116	CBV65367.1
<i>Aptenia cordifolia</i> var. <i>variegata</i>	AP	HM850876	QBC68029.1
<i>Glottiphyllum linguiforme</i>	GL	NA	NA
<i>Carpobrotus edulis</i>	CE	JQ024942	AEX68601.1

Table (2) Primer Code, primer sequences, number of total bands, polymorphic bands and percentage of polymorphism of SCoT primers.

Primer Code	Primer Sequence	Number of Bands		% of polymorphism	H%	PIC	EMR	MI	H.av	D	R
		Total	Polymorphic								
SCoT-1	CAACAATGGCTACCACC A	16	9	56.3	0.2777 8	0.23920	13.333 33	0.03858	0.00289	0.3070 2	3.33333
SCoT-2	CAACAATGGCTACCACC C	15	9	60.0	0.3200 0	0.26880	12.000 00	0.04267	0.00356	0.3618 0	4.66667
SCoT-3	CAACAATGGCTACCACC G	13	7	53.8	0.2603 6	0.22646	11.000 00	0.03672	0.00334	0.2857 1	0.00000
SCoT-4	CAACAATGGCTACCACC T	15	9	60.0	0.3064 2	0.25947	12.166 67	0.04142	0.00340	0.3438 2	3.66667
SCoT-5	CAACAATGGCTACCAGC A	14	8	57.1	0.2933 7	0.25034	11.500 00	0.04016	0.00349	0.3270 2	1.00000
SCoT-6	CAACAATGGCTACCAGC C	15	11	73.3	0.4582 7	0.35327	9.6666 7	0.04922	0.00509	0.5872 7	6.00000
SCoT-7	CAACAATGGCTACCAGC G	15	9	60.0	0.4938 3	0.37189	8.3333 3	0.04572	0.00549	0.6941 3	2.00000
SCoT-8	CAACAATGGCTACCAGC T	18	11	61.1	0.4998 3	0.37491	8.8333 3	0.04088	0.00463	0.7615 1	5.66667
SCoT-9	CAACAATGGCTACCAGC A	11	6	54.5	0.4775 0	0.36350	6.6666 7	0.04823	0.00723	0.6363 6	0.66667
SCoT-10	CAACAATGGCTACCAGC C	12	7	58.3	0.4243 8	0.33433	8.3333 3	0.04912	0.00589	0.5207 4	2.66667
Total		144	86	59.7							
Average		14.4	8.6		0.3811	0.30422	10.183 33	0.04327	0.00450	0.4825 4	2.96667

H%= Expected Heterozygosity; PIC = Polymorphism Information Content; EMR = Effective Multiplex Ratio; MI = Marker Index; Hav = Mean Heterozygosity; D = Discriminating Power; R = Resolving Power.

Table (3) Spearman rank correlation coefficient among the efficiency parameters of SCoT-primers.

	<i>H</i>	<i>PIC</i>	<i>E</i>	<i>Hav</i>	<i>MI</i>	<i>D</i>	<i>R</i>
<i>H</i>		1.00000	-0.71730	0.79394 **	0.63636 *	1.00000 *	0.35758
<i>PIC</i>			-0.71730	0.79394 **	0.63636 *	1.00000 *	0.35758
<i>E</i>				-0.89970	-0.58967	-0.71730 *	0.26748
<i>Hav</i>					0.84242 **	0.79394 **	-0.00610
<i>MI</i>						0.63636 *	0.33333
<i>D</i>							0.35758
<i>R</i>							

* Significance at 5% level; ** Significance at 1% level.

Table (4) Genetic distances between the six halophyte species as calculated Jaccard’s similarity coefficient based on SCoT data.

	FP	FH	FG	CA	GL	AP
FP	100%					
FH	99%	100%				
FG	79%	78%	100%			
CA	74%	75%	75%	100%		
GL	61%	62%	62%	69%	100%	
AP	73%	74%	76%	78%	70%	100%

Table (5): Maximum likelihood estimates of nucleotide substitution matrix of the *matK* gene sequences among the five tested species.

	A	T/U	C	G
A	-	7.81	3.66	8.26
T/U	5.88	-	11.34	3.20
C	5.88	24.17	-	3.20
G	15.15	7.81	3.66	-
Total	26.91	39.79	18.66	14.66

The matrix was calculated based on the Tamura-Nei model (1993). Rates of different transitional substitutions are shown in bold and those of transversional substitutions are shown in *italics*.

Table (6) Pairwise distance matrix of the *matK* gene sequences among the five tested species.

	<i>F. potsii</i>	<i>F. herrerae</i>	<i>F. glaucescens</i>	<i>A. cordifolia</i>	<i>C. edulis</i>
<i>F. potsii</i>	-	0.00320	0.00419	0.01626	0.02238
<i>F. herrerae</i>	0.00676	-	0.00530	0.01640	0.02241
<i>F. glaucescens</i>	0.01088	0.01562	-	0.01779	0.02420
<i>A. cordifolia</i>	0.09460	0.09314	0.09761	-	0.01117
<i>C. edulis</i>	0.14003	0.13852	0.14672	0.05819	-

Table (7): Pairwise distance matrix of the *matK* Protein sequences among the five tested species.

	<i>F. potsii</i>	<i>F. herrerae</i>	<i>F. glaucescens</i>	<i>C. edulis</i>	<i>A. cordifolia</i>
<i>F. potsii</i>	-	0.00452	0.01508	0.03033	0.01853
<i>F. herrerae</i>	0.00768	-	0.01577	0.03550	0.01982
<i>F. glaucescens</i>	0.05452	0.05884	-	0.07386	0.02508
<i>C. edulis</i>	0.20373	0.23823	0.32542	-	0.02001
<i>A. cordifolia</i>	0.12548	0.13279	0.14726	0.09453	-

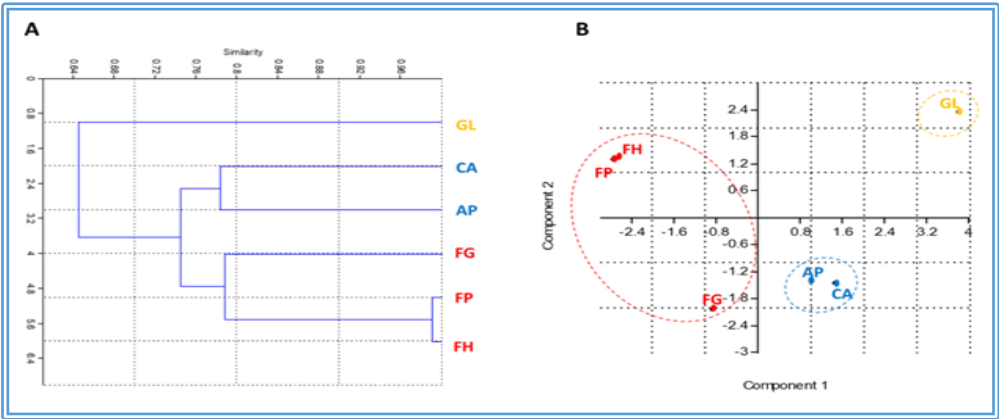


Fig. (1). (A) UPGMA cluster analysis based on Jaccard's similarity coefficient of SCoT analysis of the six halophyte species, (B) Principal Component Analysis (PCA) of the SCoT-PCR data showing the two-dimensional (Component 1 and Component 2) plot, each species denoted by a symbol. AP; *A. cordifolia* var. *variegata*, FP; *F. pottsii*, GL; *G. linguiforme*, CA; *C. edulis*, FH; *F. herrerae* and FG; *F. glaucescens*.

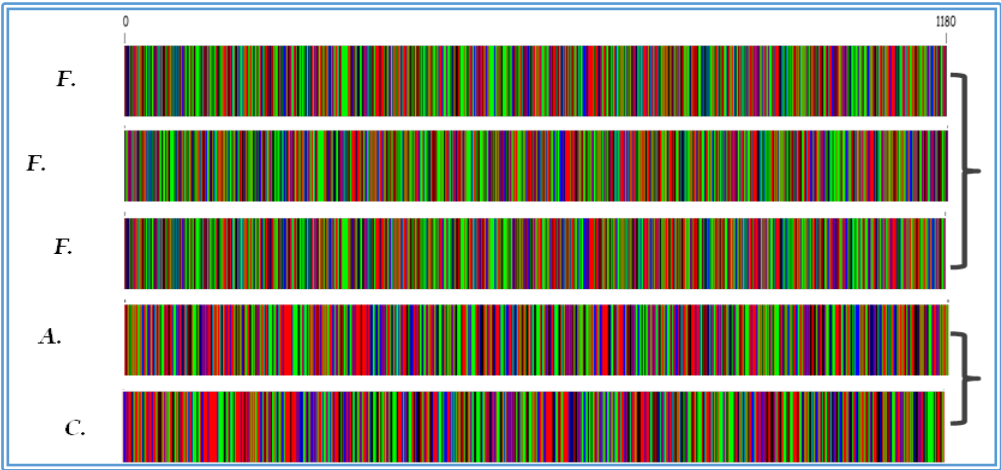


Fig. (2). Layout Comparison of *matK* DNA barcodes. Color-coded DNA sequences for the five species (*Ferocactus pottsii*, *F. glaucescens*, *F. herrerae*, *Aptenia cordifolia* and *Carpobrotus edulis*). Note that there are fewer instances of nucleotide differences between members plants belonging to even the family Cactaceae (*F. glaucescens*, *F. pottsii*, *F. Herrerae*) as well as family Aizoaceae (*Aptenia cordifolia*, *Carpobrotus edulis*) than members of different genera.

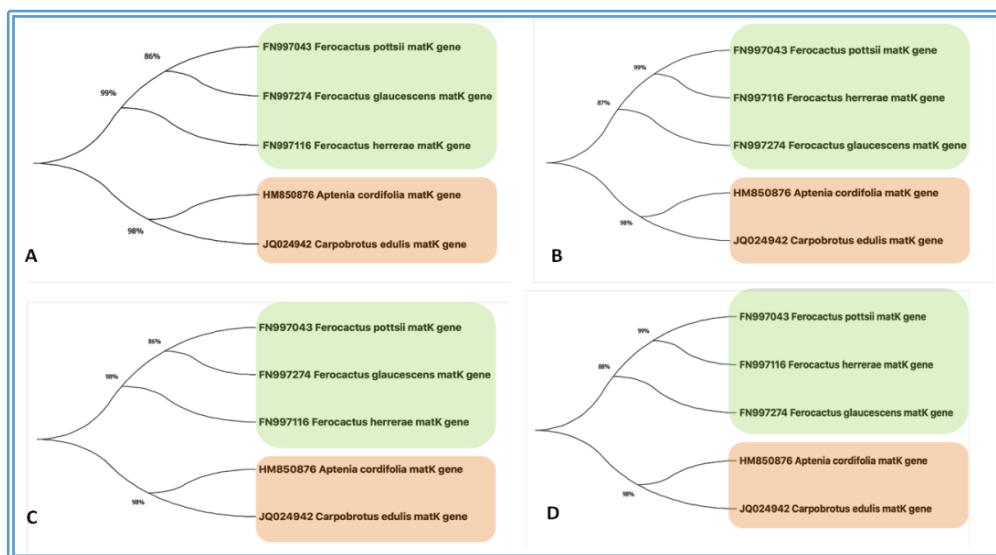
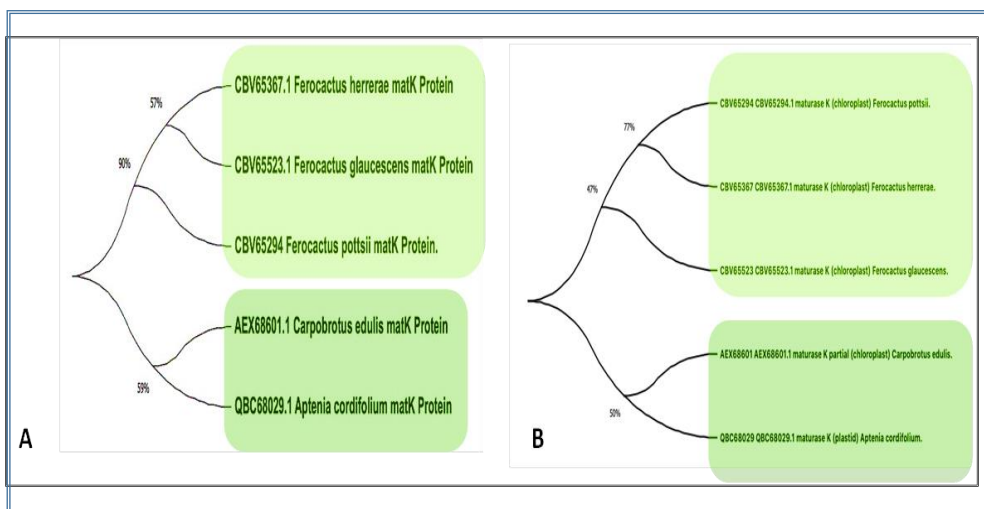


Fig. (3). A phylogenetic trees constructed based on the *matK* DNA barcode gene sequence analysis for the five species using the MEGA-X software. The evolutionary distances were calculated based on base substitutions per site using A) the Minimum Evolution method, B) the Maximum Likelihood method, C) the Neighbor-Joining method and D) the Maximum Parsimony method.

Fig. (4). A phylogenetic tree constructed based on the *matK* protein sequence



analysis for the five species using the MEGAX software. The evolutionary history was inferred using the (a) Neighbor-Joining method and (b) Jones-Taylor-Thornton model.