Plant Biotechnology Reports (2025) 19:55–66
https://doi.org/10.1007/s11816-024-00952-y

Online ISSN 1863-5474
Print ISSN 1863-5466

ORIGINAL ARTICLE

Genetic diversity of closely related *Calligounum* species collected from Saudi habitats by analyzing the *mat*K and *rpoC1* genes, and SCoT and IRAP markers

Amal Mohamed AlGarawi[1] · Magda Elsayed Abd-Elgawad[2]

Received: 9 March 2024 / Revised: 23 October 2024 / Accepted: 5 December 2024 / Published online: 13 January 2025
© The Author(s) 2025

Abstract

The genus *Calligonum* L. Her (*Polygonaceae*) comprises 158 species with widespread distribution in regions such as India, China, North Africa, Pakistan, Afghanistan, Saudi Arabia, and South Europe. *Calligonum* L. is a prominent shrubby psammophyte found in deserts, known for its traditional medicinal uses. The aim of this study is to use Start Codon Target (SCoT) and Inter-Retrotransposon Amplified Polymorphism (IRAP) markers to evaluate the genetic diversity of *Calligonum comosum* and *Calligonum tetrapterum* in Saudi Arabia. In addition, it uses the raxmlHPC algorithm and DNA barcoding techniques (matK and rpoC1) to examine phylogenetic relationships. SCoT and IRAP markers revealed distinct genetic patterns, clustering *Calligonum* species based on their genetic similarities. DNA barcoding of matK and rpoC1 genes provided valuable insights into the evolutionary relationships within *Calligonum* species. Phylogenetic analyses highlighted well-supported structures with outgroup sequences showing early divergence. Conserved motifs analysis confirmed the presence of shared motifs in both isolated and identified genes, validating the potential use of isolated genes as biomarkers. This comprehensive genetic analysis enhances our understanding of *Calligonum* species' genetic relationships, contributing valuable information for taxonomic classification and molecular marker validation. We are the first to add sequence of rpoC1 gene to *Calligonum comosum* and *Calligonum tetrapterum* in Gen-Bank. The identified conserved motifs and phylogenetic insights underscore the potential applications of *Calligonum* genes in various fields, including medicine and biodiversity conservation.

Keywords *Calligounum* · DNA barcoding · IRAP · matK · rpoC1 · SCoT

Introduction

The genus *Calligonum* L. (*Polygonaceae*) contains 158 species distributed across different geographic areas like such as India, China, North Africa, Pakistan, Afghanistan, Saudi Arabia, and Southern Europe (Purohit and Kumar 2020; Alaklabi 2022). The mobile dunes are the most common home to one important species: *Calligonum comosum* L'Hert (Purohit and Kumar 2020; Alaklabi 2022). Growing up to three meters tall, this evergreen shrub bears silvery white flowers, grayish white stems, and white branches (Fal'kovich 2023). In traditional folk medicine, *Calligonum comosum*, also called Escanbil, is used to treat menstrual cramps and weighty periods (Kiani et al. 2019). Its bioactive chemical components, such as tannins, anti-oxidative flavonoids, and glycoside derivatives, also give it anti-inflammatory, anticancer, and antioxidant qualities (Kiani et al. 2019; Atef et al. 2020). It can also be used to treat gynecological conditions that cause infertility (Tahmasebi et al. 2018; Barati and Movahedin 2021).

Short DNA fragments called barcode markers are used in DNA barcoding to identify species (Antil et al. 2023). DNA barcoding uses barcode markers to identify species, with a curated database needed for successful results. Global initiatives aim to make biodiversity data publicly available (Leray et al. 2019; Antil et al. 2023). Plant DNA barcoding is essential for conservation, ecology, product verification,

✉ Magda Elsayed Abd-Elgawad
Magda.Elsayed@rimp.bsu.edu.eg;
m_abdelgawad@outlook.com

Amal Mohamed AlGarawi
Aalgarawi@ksu.edu.sa

[1] Department of Botany and Microbiology, College of Science, King Saud University, 11451 Riyadh, Saudi Arabia

[2] Biotechnology Department, Research Institute of Medicinal and Aromatic Plants, Beni-Suef University, Beni-Suef, Egypt

and reconstructing past vegetation and climate (Sheth and Thaker 2017; Tizard et al. 2019). It can identify species from entire plants or remnants like pollen, seeds, and roots (Kress 2017; Lennartz et al. 2021). Barcoding techniques are used in various contexts, such as verifying plant products like tea (Avarave and Thomas 2021), berries (Wu et al. 2018), olive oil (Nehal et al. 2021), spices (Negi et al. 2021), and medicinal plants(Yu et al. 2021). Building phylogenetic trees in phylogenetic community ecology requires DNA barcoding (DeSalle and Goldstein 2019). Chloroplast loci matK and rbcL are recommended for molecular identification in plants (Skuza et al. 2019). The Consortium Barcode of Life supports using both loci for accurate species identification (CBOL Plant Working Group, 2009). MatK's strong discriminatory capabilities across a variety of taxa make it especially useful for examining plant phylogenetic relationships (Liu et al. 2019).

Multilocus markers, such as IRAP markers, are commonly used in genetic research to validate DNA (Guan et al. 2020). IRAP markers arise from variations in retrotransposon insertion sites across genomic regions and generate informant DNA fingerprints that aid in kinship analysis and genetic identification (Amiteye 2021; Kalendar et al. 2021; Arvas et al. 2023). They have been used to assess genetic diversity in species such as bamboo and sakura, demonstrating their ability to evaluate samples (Stepanov et al. 2020; Li et al. 2020).

SCoT markers are useful tools for studying plant genetics because they target regions surrounding the ATG start codon (Mishra et al. 2020). Due to their low recombination rates, they provide advantages in marker-assisted breeding programs. Studies on date palms, mango, and other species have demonstrated the utility of SCoT for DNA fingerprinting and genetic analysis (Adawy et al. 2014) (Luo et al. 2010). SCoT assessed the genetic variations of 29 wild plants in Al Jubail, Saudi Arabia (Alotaibi and Abd-Elgawad 2022).

Numerous studies have been conducted on *Calligonum*'s genetic diversity to support taxonomy and conservation. While RAPD and ISSR analyses demonstrated broad variability in *C. comosum* populations in Tunisia (Gasmi et al. 2022), RAPD markers in Egypt demonstrated notable diversity (Mahdy and Rizk 2023). In China, cpDNA, ITS, and ETS sequences verified the link between *C. jeminaicum* and *C. mongolicum* (Shi et al. 2019), while full plastome sequencing elucidated taxonomic relationships within the genus (Song et al. 2020).

However, few studies particularly address the genetic diversity of *Calligonum* species in Saudi environments. This study seeks to close this gap by examining genetic diversity using matK and rpoC1 genes in conjunction with SCoT and IRAP markers. The purpose of this study is to assess the genetic diversity for Saudi *Calligonum comosum* and *Calligonum tetrapterum* using SCoT and IRAP markers. It also uses DNA barcoding methods and the raxmlHPC algorithm to analyze phylogenetic relationships. This study advances future conservation strategies and possible uses in agriculture, improving breeding programs and medicine by shedding light on the relationships between *Calligonum* species.

Methods

Collection of plant samples

Calligounum comosum plants were collected from two sites: Al-Kharj (C1) and Al-Qassim (C2), as well as *Calligonum tetrapterum*, which was collected from two sites: Najran (T1) and Al-Qassim (T2). The plant samples were identified in the King Saud University Herbarium (KSU) of the Department of Botany and Microbiology.

Amplification of SCoT and IRAP markers

Eleven SCoT and eleven IRAP primers were used (Table 1). PCR amplification was carried out in 25µL reaction volume containing 12.5-µL 2X PCR Master Mix (sigma), 2-µL primer (10pcmol/ µL), 3-µL genomic DNA (10 ng/ µL) and 7.5-µL dH_2O (Ibrahim et al. 2019). The amplification process was performed in a Perkin-Elmer/GeneAmp® PCR System 9700 (PE Applied Biosystems, Waltham, MA, USA) programmed to initial denaturation cycle for 3 min at 94 °C followed by 35 and 40 cycles for IRAP and SCoT, with each cycle consisting of a denaturation at 94 °C for 40 s., annealing at 50 °C for 50 s., and elongation at 72 °C for 60 s. The primer extension segment was extended to 7 min at 72 °C in the final cycle. The amplification products were resolved by electrophoresis in a 1.5% agarose gel containing ethidium bromide (0.5ug/ml) in 1X TBE buffer at 95 V. A 100 bp DNA ladder was used as a molecular size standard PCR products were visualized on UV light and photographed using a Gel Documentation System (BIO-RAD 2000).

matK and rpoC1 barcoding analysis

DNA barcoding analysis was performed with the plastidial *mat*K *and rpoc1* region (Table 1). For PCR amplification and sequencing of *mat*K and *rpoc1*, the reaction mixture consisted of 1 × buffer (Promega), 15 mM $MgCl_2$, 0.2-mM dNTPs, 20pcoml of each primers, 1u of Taq DNA polymerase (GoTaq, Promega), 40-ng DNA and ultra-pure water to a final volume of 50 uL. PCR amplification was performed in a Perkin-Elmer/GeneAmp® PCR System 9700 (PE Applied Biosystems) programmed to fulfill 40 cycles after an initial denaturation cycle for 5 min at 94 °C. Each cycle consisted of a denaturation step at 94 °C for 30 s., an annealing step at 47 °C and 50 °C for 30 s. and an elongation step at 72 °C for

Table 1 List of IRAP primers, matK primers, SCoT primers and rpoC1 primers sequence used to analyze *Calligounum* species

No	IRAP	Primer sequence	SCoT	Primer sequence
1	IRAP-01	5'-TTAGACCCGGAACCGCCGTG-3'	SCoT-1	5'-ACGACATGGCGACCACGC-3'
2	IRAP-02	5'-ATCCTTCGCGTAGATCAAGCGCCA-3'	SCoT-2	5'-ACCATGGCTACCACCGGC-3'
3	IRAP-03	5'-GAAGTACCGATTTACTTCCGTGTA-3'	SCoT-3	5'-ACGACATGGCGACCCACA-3'
4	IRAP-04	5'-ATGTGACAGTCGACTAACCAC-3'	SCoT-4	5'-ACCATGGCTACCACCGCA-3'
5	IRAP-05	5'-TGGCGCTTGATCTACGCGAAGGA-3'	SCoT-5	5'-CAATGGCTACCACTAGCG-3'
6	IRAP-06	5'-AACTTGATCCAGATCATCTCC-3'	SCoT-6	5'-CAATGGCTACCACTACAG-3'
7	IRAP-07	5'-TGATCAGAGAAGAAAGGGGA-3'	SCoT-7	5'-ACAATGGCTACCACTGAC-3'
8	IRAP-08	5'-GCGCTTAGCGTTAGGCTAACT-3'	SCoT-9	5'-ACAATGGCTACCACTGCC-3'
9	IRAP-09	5'-CGTGTAGCCTCTTTGGCCCTA-3'	SCoT-10	5'-ACAATGGCTACCACCAGC-3'
10	IRAP-10	5'-CGAGTAGTAGGAAGGAACCGG-3'	SCoT-11	5'-ACAATGGCTACCACTACC-3'
11	IRAP-11	5'-GAACGATTACCTCACAGACA-3	SCoT-12	5'-CAACAATGGCTACCACCG-3'
	matK	Primer sequence	rpoC1	Primer sequence
	matK-F	5'-CGATCTATTCATTCAATATTTC-3'	rpoC1-F	5'-GGCAAAGAGGGAAGATTTCG-3'
	matK-R	5'-TCTAGCACACGAAAGTCGAAGT-3'	rpoC1-R	5'-CCATAAGCATATCTTGAGTTGG-3'

1 min for *mat*K and *rpoc1*, respectively. The primer extension segment was extended to 7 min at 72 °C in the final cycle. The amplification products were resolved by electrophoresis in a 1.5% agarose gel containing ethidium bromide (0.5ug/ml) in 1X TBE buffer at 95 V. A 100 bp DNA ladder was used as a molecular size standard. PCR products were visualized on UV light and photographed using a Gel Documentation System (BIO-RAD 2000). Amplified products for all PCR were purified using EZ-10 spin column PCR products purification PCR reaction mixture was transferred to 1.5-ml microfuge tube and three volumes was added of binding buffer 1 after that the mixture solution was transferred to the EZ-10 column and let it stand at room temperature for 2 min after that centrifuge, 750 ul of wash solution was added to the column and centrifuge at 10.000 rpm for 2 min, repeated washing, 10.000 rpm was spine for an additional minute to remove any residual wash solution. The column was transferred into a clean 1.5-ml microfuge tube and adds 50 ul of elution buffer, incubated at room temperature for 2 min and when store purified DNA at – 20 °C.

matK and rpoC1 sequencing

The sequencing of the product PCR was carried through in an automatic sequencer ABI PRISM 3730XL Analyzer using Big Dye TM Terminator Cycle Sequencing Kits following the protocols supplied by the manufacturer. Single-pass sequencing was performed on each template using *mat*K and *rpoC1* Forward primer. The fluorescent-labeled fragments were purified from the unincorporated terminators with an ethanol precipitation protocol. The samples were resuspended in distilled water and subjected to electrophoresis in an ABI 3730xl sequencer (The American University in Cairo School of Sciences and Engineering).

Data analysis

For SCoT and IRAP analysis, only clear and unambiguous bands were visually scored as either present (1) or absent (0) for all samples and final data sets included both polymorphic and monomorphic bands. Then, a binary statistic matrix was constructed. Dice's similarity matrix coefficients were then calculated between genotypes using the unweighted pair group method with arithmetic averages (UPGMA). This matrix was used to construct a phylogenetic tree was performed according to Euclidean similarity index using the PAST3 software Version 1.91 (Hammer et al. 2001).

matK and rpoC1 comparative analysis

In this study, our primary goal was to validate our experimental findings through a comparative analysis of the matK and rpoC1 genes. To achieve this, we retrieved matK and rpoC1 gene sequences from all *Calligonum* species available on the NCBI database. Specifically, we obtained 22 matK sequences from 22 *Calligonum* species, and included 2 additional matK sequences from the nucleotide database on NCBI (MK228108.1, KY293577.1), as presented in Table 3. In addition, we retrieved 22 rpoC1 sequences from 22 *Calligonum* species, as detailed in Table 4.

This comprehensive approach enables us to conduct a meaningful comparative analysis between the genes isolated in our experiment and those identified in the database. Our objective is to explore similarities between the two sets of genes and derive insights from these comparisons, thereby validating our genes as molecular markers.

To facilitate this analysis, we utilized the MAFFT algorithm (Katoh and Standley 2013), a well-established tool for multiple sequence alignment. This step ensures the accurate

alignment of our sequences, providing a robust foundation for subsequent comparisons and interpretations.

Phylogenetic relationships

To deduce the evolutionary relationships and formulate a phylogenetic tree based on the aligned sequences, we utilized the raxmlHPC algorithm (maximum likelihood) (Rokas 2011). Recognized for its precision and dependability in constructing phylogenetic trees, raxmlHPC was selected to unveil the evolutionary context of our sequences. The resulting phylogenetic tree file produced by raxmlHPC was then uploaded and visualized through the iTOL (Interactive Tree of Life) software platform (Bailey et al. 2015). This procedure enabled us to generate a visually informative and scientifically robust representation of the evolutionary relationships among the sequences.

Conserved motifs analysis

For the identification of conserved motifs shared between our isolated genes and those identified in the NCBI-supported genes, we employed the MEME Suite algorithm (Bailey et al. 2015). This motif analysis encompassed both MATK and rpoC1 genes, previously identified and supported on NCBI, from all Calligonum species, as well as our isolated genes. The analysis was configured to detect a maximum of five motifs. The resulting conserved motifs were visualized using the TBtools software (Chen et al. 2020). Notably, these findings were integrated with the outcomes of the phylogenetic tree analysis to offer a comprehensive understanding of motif conservation across diverse evolutionary lineages.

Results

SCoT and IRAP markers

Figure 1 displays photographs of agarose gel electrophoresis for PCR fingerprinting in 4 *Calligounum* species. Eleven SCoT markers yielded 121 amplicons, 57 of which (47.11%) were polymorphic (Table 1). The amplicons per primer ranged from 7 for SCoT-7 and SCoT-9 to 15 for SCoT-1. The number of polymorphic markers varied, ranging from 1 for SCoT-9 to 11 for SCoT-2. The average number of polymorphic amplicons per SCoT primer was 5.18. The lowest polymorphism, 14.3%, was detected in SCoT-9. On the other hand, the highest polymorphism 78.6% was recorded in SCoT-2 (Table 2).

Four species were analyzed using SCoT markers, revealing significant genetic variation. *C. comosum*2 had 22 polymorphic bands, *C. comosum*1 and *C. tetrapterum*1 had 37, and *C. tetrapterum*2 had 29. These findings imply complex genomic structures or increased genetic diversity. A Euclidean distance tree was created for species using SCoT markers, scoring 1 for presence and 0 for absence (Fig. 2; Table 3). The species are divided into two major groups: *C. tetrapterum* and *C. comosum*. The two *C. comosum* had the highest similarity of 87.7, while *C. comosum*1 and *C. tetrapterum*1 had the lowest of 75.3. These numbers show how similar the two most similar species were.

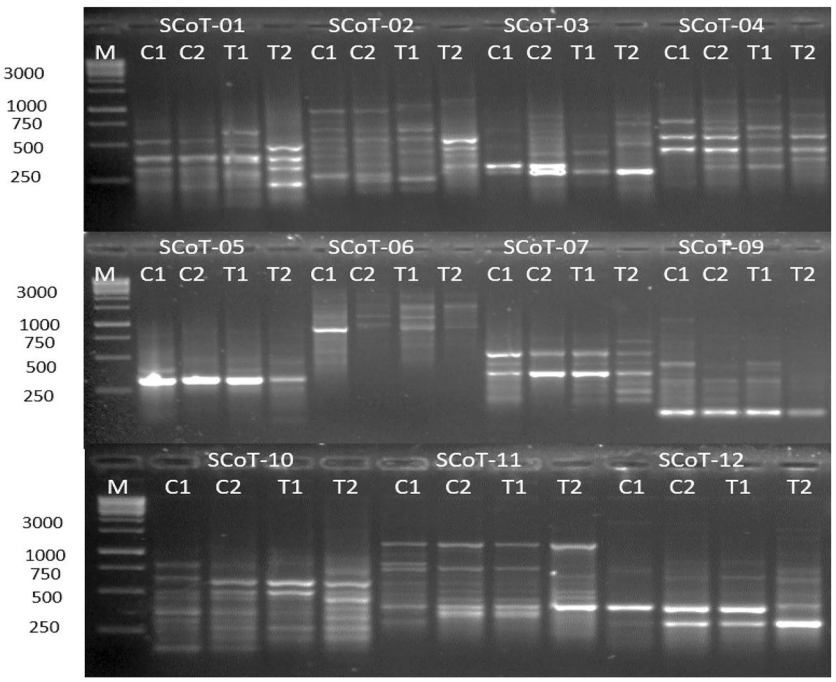


Fig. 1 Banding patterns of four samples *Calligounum* species (two *C. comosum*; C1, C2 and two *C. tetrapterum*; T1 and T2) amplified with the eleven SCoT primers, M: 1 kb DNA ladder

Table 2 The list of primers, total number of amplicons (TNAs), Number of polymorphic amplicons (NPAs), and percentage of polymorphism (P %), as revealed by SCoT and IRAP markers

SCoT	TNAs	NPAs	P %	IRAP	TNAs	NPAs	P %
SCoT-1	15	10	66.7	IRAP-01	15	7	45.7
SCoT-2	14	11	**78.6**	IRAP-02	20	13	**65**
SCoT-3	10	6	60	IRAP-03	8	3	37.5
SCoT-4	12	5	41.7	IRAP-04	11	5	45.5
SCoT-5	12	6	50	IRAP-05	8	4	50
SCoT-6	8	4	50	IRAP-06	13	4	30.7
SCoT-7	7	3	42.9	IRAP-07	16	6	37.5
SCoT-9	7	1	**14.3**	IRAP-08	6	2	33.3
SCoT-10	14	3	21.4	IRAP-09	11	3	27.3
SCoT-11	12	5	41.7	IRAP-10	9	3	33.3
SCoT-12	10	3	30	IRAP-11	11	0	**0**
Total	**121**	**57**	**47.11**	**Total**	**128**	**50**	**39.1**
Average	**11**	**5.18**	**45.2**	**Average**	**11.6**	**4.55**	**32.78**

The lowest and the highest value are in bold and underline

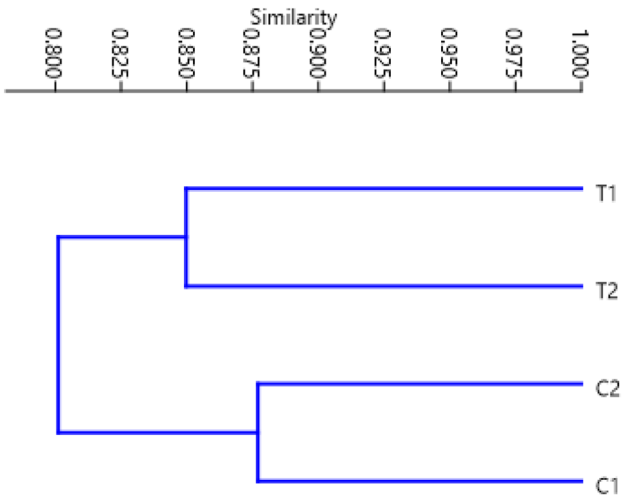


Fig. 2 Dendrograme for *Calligounum* species constructed from SCoT marker data using UPGMA and similarity matrix computed according to Dice coefficient

Table 3 Similarity matrix among the *Calligounum* species according to the coefficient of dice as shown by SCoT marker

	C1	C2	T1	T2
C1	100			
C2	87.7	100		
T1	75.3	81.9	100	
T2	79.1	84.2	85.0	100

The lowest and the highest value are in underline.

Figure 3 shows images of agarose gel electrophoresis used in four different *Calligounum* species for PCR fingerprinting. Eleven IRAP markers produced 128 amplicons, of which 50 (39.1%) were polymorphic (Table 1). For IRAP-08, and IRAP-02, the amplicons per primer varied from 6 to 20. There were different numbers of polymorphism markers: 0 for IRAP-11 and 13 for IRAP-02, respectively. 4.55 polymorphic amplicons on average were found for each IRAP primer. IRAP-11 exhibited the lowest polymorphism, 0%. However, Table 2 shows that IRAP-02 had the highest polymorphism of 65%.

Four species were analyzed using IRAP markers, revealing significant genetic variation. *C. comosum*1 had 26 polymorphic bands, *C. comosum*2 had 24, *C. tetrapterum*1 had 32, and *C. tetrapterum*2 had 23. Despite greater polymorphism in *C. tetrapterum1*, the identical band counts of the first three *Calligounum* indicated a closer genetic relationship. A Euclidean distance tree was created using binary data from IRAP markers, which were scored as 1 for presence and 0 for absence for each species (Fig. 4; Table 4). The tree divided the species into two main groups, with two species, *C. comosum*2 and *C. tetrapterum*2, creating the first cluster. *C. comosum*2 and *C. tetrapterum*2 had the highest similarity score of 91.5, while *C. comosum*1 and *C. tetrapterum*1 had the lowest score of 82.0.These numbers show how similar the two species with the highest degree of similarity were.

Molecular identification of matK and rpoC1 gene

Four DNA samples of *C. comosum* and *C. tetrapterum* were used as a template for PCR amplification of matK gene region identification of *C. comosum* and *C. tetraterum*. Results showed that partial nucleotide sequences of matK gene 786, 786, 785, and 785 nts were amplified for the PCR of C1, C2, T1, and T2 respectively, the sequences were documented in Gen-Bank under the accession numbers: OR855414.1, OR855415.1, OR855416.1, and OR855417.1 respectively (Table 5). According to rpoC1, results showed that partial nucleotide sequences of rpoC1 gene 555, 555,

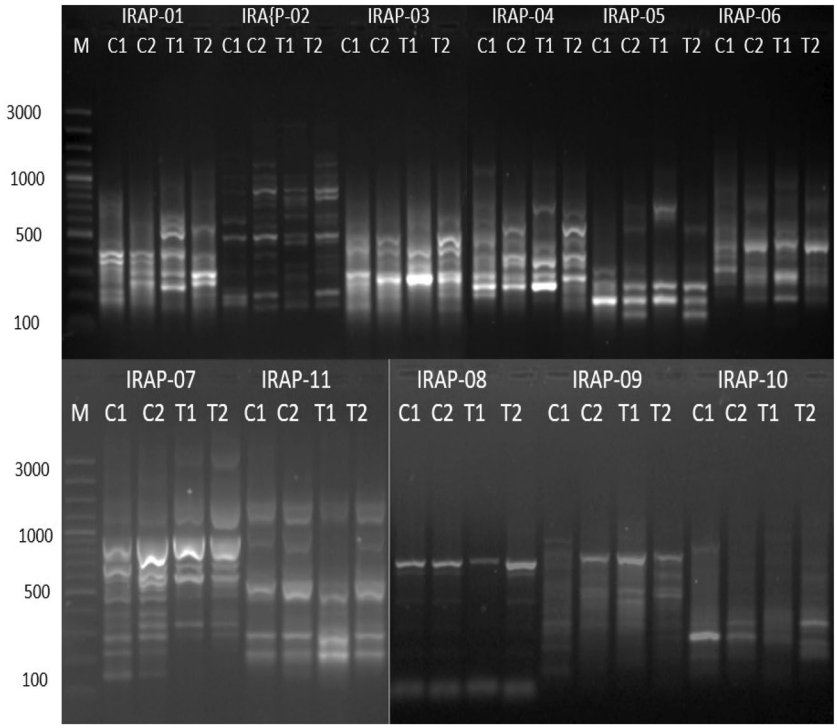


Fig. 3 Banding patterns of four samples *Calligounum* species (two *C. comosum*; C1, C2, and two *C. tetrapterum*; T1 and T2) amplified with the eleven IRAP primers, M: 1 kb DNA ladder

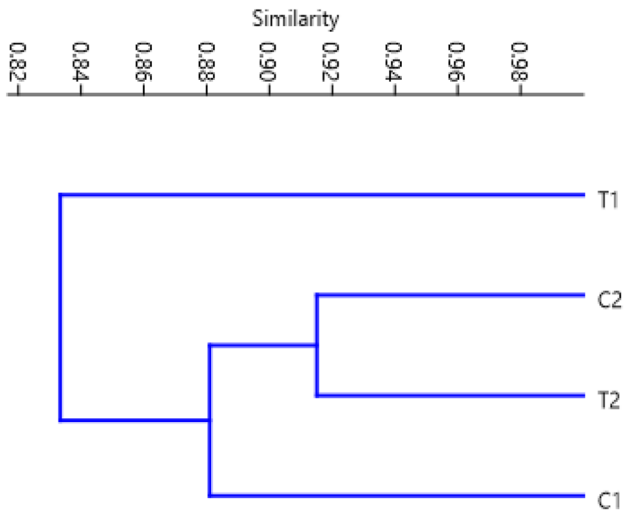


Fig. 4 Dendrograme for *Calligounum* species constructed from IRAP marker data using UPGMA and similarity matrix computed according to Dice coefficient

Table 4 Similarity matrix among the *Calligounum* species according to the coefficient of dice as shown by IRAP marker

	C1	C2	T1	T2
C1	100			
C2	88.4	100		
T1	82.0	82.8	100	
T2	87.9	91.5	85.3	100

551, and 515 nts were amplified for the PCR of C1, C2, T1, and T2 respectively, the sequences were documented in Gen-Bank under the accession numbers of OR933566.1, OR933567.1, OR933568.1 and OR933569.1 respectively (Table 6). We are the first to add sequence rpoC1 gene to *C. comosum* and *C. tetrapterum* in Gen-Bank.

Phylogenetic relationships

Phylogenetic relationships were constructed using the raxmlHPC algorithm with 10 bootstrap replica, and conserved motifs were identified using the MEME Suite algorithm. The phylogenetic tree of matK genes from *Calligonum* species shows distinct classes and sub-clads, with out-group sequences of *C. tetrapterum* and *C. comosum* showing early divergence. The tree also shows a nested pattern, highlighting hierarchical relationships.

The phylogenetic tree from rpoC1 genes reveals significant evolutionary patterns, with a clear hierarchical structure and well-supported clads. Early divergence in *C. comosum*, distinct branches in *C. tetrapterum*, and similarity in *C. juochiangense*.

In summary, this gene-based phylogenetic analysis provides valuable insights into the evolutionary dynamics of matK and rpoC1 within *Calligonum* species. These findings contribute to our understanding of genetic relationships, with potential implications for taxonomic classification and the validation of molecular markers (Fig. 5).

Table 5 Accession numbers for matK genes in 22 identified species

Database accession	
Accession	Organism
NC_053267.1	*Calligonum yengisaricum*
NC_053266.1	*Calligonum taklimakanense*
NC_053265.1	*Calligonum squarrosum*
NC_053264.1	*Calligonum rubicundum*
NC_053263.1	*Calligonum roborowskii*
NC_053262.1	*Calligonum pumilum*
NC_053261.1	*Calligonum mongolicum*
NC_053260.1	*Calligonum leucocladum*
NC_049137.1	*Calligonum aphyllum*
NC_049148.1	*Calligonum korlaense*
NC_049149.1	*Calligonum klementzii*
NC_049147.1	*Calligonum junceum*
NC_049146.1	*Calligonum jeminaicum*
NC_049145.1	*Calligonum ebinuricum*
NC_049144.1	*Calligonum densum*
NC_049143.1	*Calligonum cordatum*
NC_049142.1	*Calligonum colubrinum*
NC_049141.1	*Calligonum caput-medusae*
NC_049140.1	*Calligonum arborescens*
NC_049139.1	*Calligonum gobicum*
NC_049138.1	*Calligonum juochiangense*
NC_062619.1	*Calligonum bakuense*
MK228108.1	*Calligonum comosum* 6H matK gene
KY293577.1	*Calligonum comosum* matK gene
Our Accession	
OR855414.1	*Calligonum comosum* C1 matK gene
OR855415.1	*Calligonum comosum* C2 matK gene
OR855416.1	*Calligonum tetrapterum* T1 matK gene
OR855417.1	*Calligonum tetrapterum* T2 matK gene

Table 6 Accession numbers for rpoC1 Genes in 22 identified species

Database accessions	
Accession	Organism
NC_053267.1	*Calligonum yengisaricum*
NC_053266.1	*Calligonum taklimakanense*
NC_053265.1	*Calligonum squarrosum*
NC_053264.1	*Calligonum rubicundum*
NC_049140.1	*Calligonum arborescens*
NC_049141.1	*Calligonum caput-medusae*
NC_049142.1	*Calligonum colubrinum*
NC_049143.1	*Calligonum cordatum*
NC_049144.1	*Calligonum densum*
NC_049145.1	*Calligonum ebinuricum*
NC_049146.1	*Calligonum jeminaicum*
NC_049147.1	*Calligonum junceum*
NC_049148.1	*Calligonum klementzii*
NC_049149.1	*Calligonum korlaense*
NC_062619.1	*Calligonum bakuense*
NC_049137.1	*Calligonum aphyllum*
NC_053260.1	*Calligonum leucocladum*
NC_053261.1	*Calligonum mongolicum*
NC_053262.1	*Calligonum pumilum*
NC_053263.1	*Calligonum roborowskii*
NC_049139.1	*Calligonum gobicum*
NC_049138.1	*Calligonum juochiangense*
Our Accessions	
OR933566.1	*Calligonum comosum* C1 RpoC1 gene
OR933567.1	*Calligonum comosum* C2 RpoC1 gene
OR933568.1	*Calligonum tetrapterum* T1 RpoC1 gene
OR933569.1	*Calligonum tetrapterum* T2 RpoC1 gene

Conserved motifs analysis

Our exploration of conserved motifs in matK and rpoC1 genes unveiled the detection of five distinct motifs, depicted in Fig. 6. Upon scrutinizing these motifs in our accessions and comparing them to those identified in the database, a noteworthy observation. The five motifs identified in the database genes shared a common region, approximately from 647 to 1225 bp for matK and from 1876 to 2363 bp for rpoC1 (refer to supplementary MEME files).

Intriguingly, our isolated genes were found to harbor the same conserved motifs, validating our investigation. Despite the shorter length of our sequences, the distances between the discovered motifs were consistently spaced. This observation suggests the isolated region aligns with the identified regions in *Calligonum* species genes retrieved from the NCBI database. The motifs' logos are illustrated in Fig. 7.

These findings robustly validate the potential utility of our isolated genes as biomarkers. The shared conserved motifs emphasize the reliability and significance of our genetic findings, enhancing their applicability in molecular studies and contributing to the broader understanding of genetic relationships within *Calligonum* species.

Discussion

This study evaluated the genetic diversity of *Calligonum* species collected from Saudi habitats using SCoT, IRAP markers and the chloroplast DNA matK and rpoC1 barcodes. Genetic diversity was investigated using 121 markers created by 11 SCoT primers, the polymorphism ratio ranging from 78.6% for SCoT-2 to 14.3% for SCoT-9, with the average polymorphism ratio was 45.2%. Some wheat cultivars reported low levels of polymorphism ranging from 11 to 81% by the SCoT primers (Abouseada et al. 2023). Several wheat cultivars from Northern Africa have

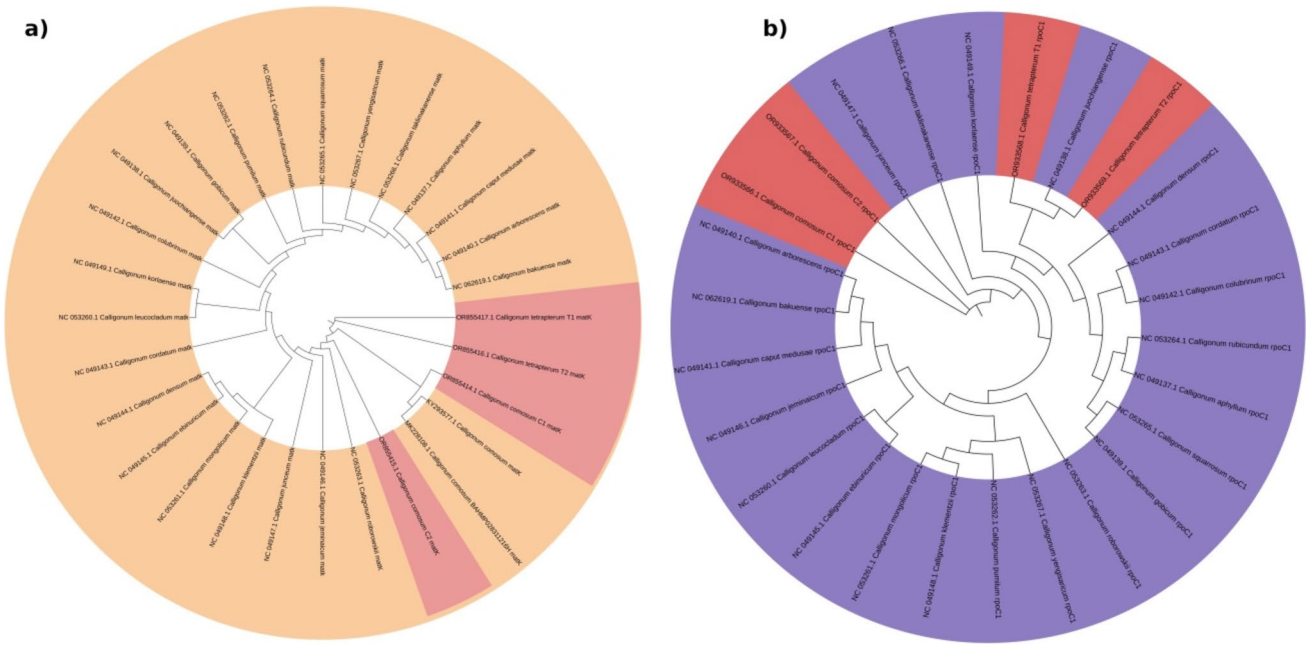


Fig. 5 Phylogenetic tree generated using *raxmlHPC* algorithm from **a** matK sequences and **b** rpoC1 sequences

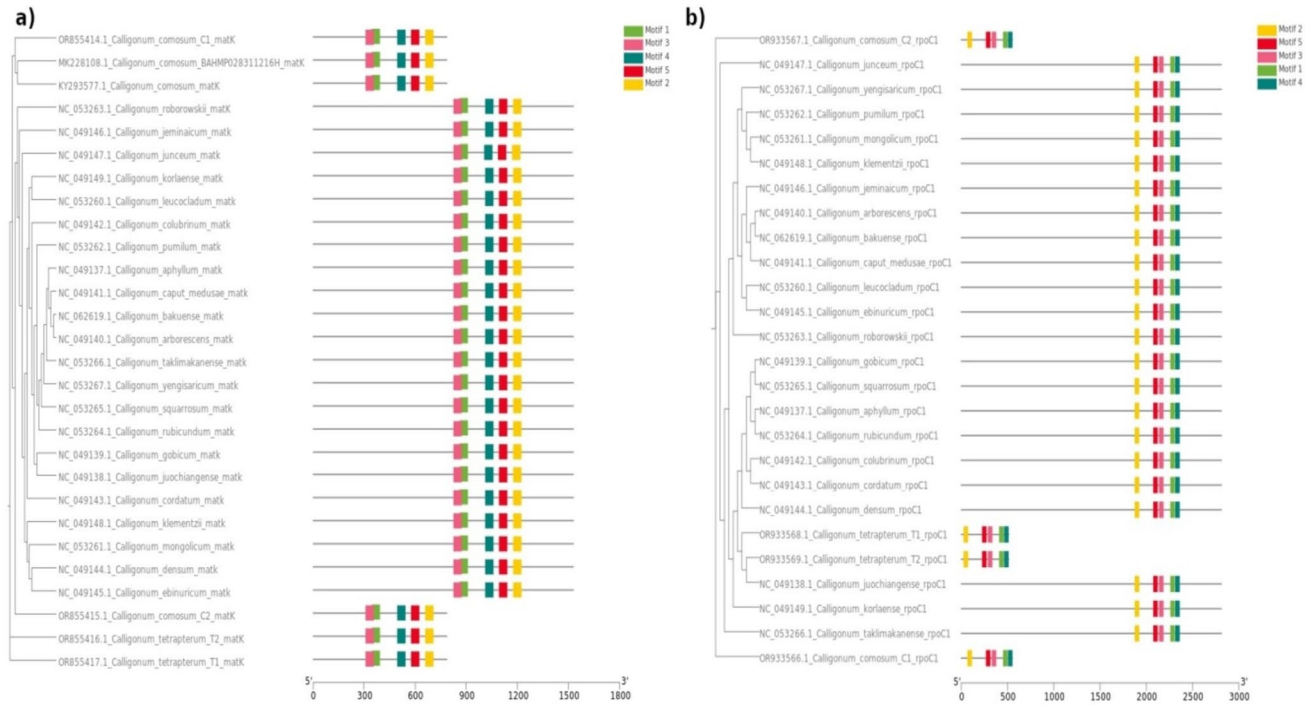


Fig. 6 Conserved motifs analyses using MEME suit algorithm from **a** matK sequences and **b** rpoC1 sequences

low levels of polymorphism, as measured by the SCoT primers, ranging from 8 to 57% (Mohamed et al. 2017). The low polymorphism percentage in wheat cultivars may be due to the limited diversity of genes and high preservation.

All populations had different levels of genetic polymorphism, according to the SCoT marker analysis. Of these, *C. comosum*1 and *C. tetrapterum*1 had the most polymorphic bands (37 each), suggesting higher genetic diversity and potentially more complex genomic structures. While *C.*

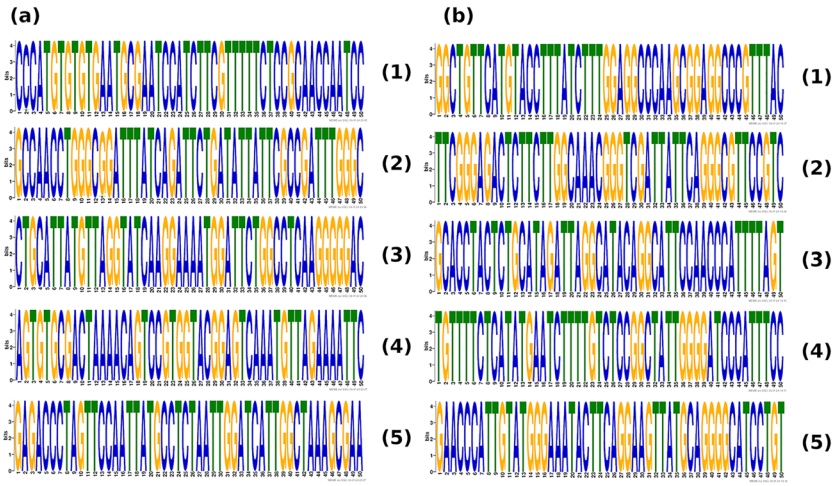


Fig. 7 Conserved motifs logos of the identified motifs from **a** matK sequences and **b** rpoC1 sequences

*comosum*2 had the fewest polymorphic bands (22), indicating comparatively less genetic variability, *C. tetrapterum*2 displayed an intermediate level of polymorphism with 29 bands. These results demonstrate the importance of SCoT markers in evaluating genetic diversity within the population as well as the notable genetic variation among the samples. The genetic varying qualities of 4 *Calligonum* genotypes were computed using the esteem of hereditary differing qualities parameters revealed by SCoT markers. Certain species demonstrated convergence, as seen in the case of the two *C. comosum* and *C. tetrapterum* in the clustering analysis. Other reports that estimated the genetic relationships using SCoT markers disagreed with these results, as in 29 wild plants in Al Jubail Saudi Arabian (Alotaibi and Abd-Elgawad 2022), and among indigenous EPNs belonging to *H. indica* species (Abd El Azim and Khashaba 2021).

This study evaluated hereditary differences using 128 markers generated by 11 IRAP primers. Polymorphism rates ranged from 0% for IRAP-11 to 65% for IRAP-02, with the average polymorphism ratio was 32.78%. The findings contradict Badr et al. (Badr et al. 2020), who used IRAP to study genetic diversity in *Medicago sativa* cultivars and found high polymorphism. The populations had different levels of genetic polymorphism, according to the IRAP marker analysis, with polymorphic band counts of 26, 24, 23, and 32. Although the genetic diversity levels of the first three samples were comparatively comparable, *C. tetrapterum*1 had a noticeably higher number of polymorphic bands, suggesting greater genetic variability. Accordingly, *C. tetrapterum*1 might have been subjected to different evolutionary pressures or have a more varied genetic background. Overall, the findings demonstrate how well IRAP markers identify genetic diversity and offer insightful information about the genetic makeup of the population. The low polymorphism percentage in *Calligonum* species may be due to the limited diversity of genes and high preservation. Certain species demonstrated convergence, as seen in the case of *C. comosum*2 and *C. tetrapterum*2 in the clustering analysis. These findings disagreed with previously recorded findings, IRAP markers was conducted on 12 sakura genotypes (Stepanov et al. 2020), and the Asian bamboo genus *Phyllostachys* (Li et al. 2020).

The results indicate that SCoT is highly reproducible and more effective than IRAP markers in assessing the relationships between the taxa under investigation within *Calligonum* species. Given that *C. comosum*1 and *C. tetrapterum*1 have more polymorphic bands, the SCoT analysis may be more successful in accurately differentiating between species. Their unique genetic diversity and evolutionary paths are reflected here. On the other hand, the IRAP analysis may be more useful for comparing traits between closely related species because it identifies differences in retrotransposons activity that could reveal traits and evolutionary relationships in common. Through the integration of findings from both IRAP and SCoT analyses, we can improve our comprehension of these species' evolutionary dynamics and genetic diversity. Although the start codon ATG is surrounded by a short, conserved sequence known as the SCoT sequence (Collard and Mackill 2009), IRAP marker analyses indicate that the instability of RTN insertion events and environmental conditions may contribute to the findings (Kalendar et al. 2011). Diverse biologic circumstances may be the reason for the distinct increases in the separated species under inspection. These conditions may also increase the genetic diversity among species, which in turn may promote the innate assortment among them. IRAP and SCoT markers showed high levels of polymorphism in genetic diversity analyses of 268 *Diospyros* accessions in China (Guan et al. 2020).

This study marks the inaugural exploration of the roles played by matK and rpoC1 genes within the *Calligonum* genus. The matK and rpoC1 sequences from the same plant species were obtained from NCBI. Subsequently, each sequence underwent a BLAST search in Gen-Bank using the mega-blast parameter, allowing for

an assessment of homology among barcoding sequences within the plant species and related taxa. The percentage of homology, as evident in the phylogenetic tree, was examined to identify the closest matches. We are the first to add a sequence of rpoC1 genes to *Calligonum comosum* and *Calligonum tetrapterum* in Gen-Bank. The matK gene phylogenetic tree displays a robust structure with distinct clades and subclades. The *C. tetrapterum* T1 and *C. tetrapterum* T2 outgroup sequences are noteworthy because they occupy a basal position, indicating an early divergence from the other *Calligonum* species. The close relationship between *C. comosum* 6H, *C. comosum*, and *C. comosum* C1 indicates high similarity in the matK sequences. The tree displays a nested pattern as it moves inward, emphasizing the hierarchical relationships between the matK genes of *Calligonum* species.

The rpoC1 gene phylogenetic tree derived from unveils significant evolutionary patterns. The tree displays a clear hierarchical structure with well-supported clades. The outgroup of *C. comosum* C2 represents an early divergence, followed by a distinct branch leading to *C. comosum* C1 at the base. *C. tetrapterum* T1 and *C. tetrapterum* T2 share a clade with *C. juochiangense*, indicating a degree of similarity between rpoC1 genes in these species. Our comparative analyses of isolated matK and rpoC1 genes demonstrated high similarities with the corresponding genes in other *Calligonum* species from the NCBI database. The phylogenetic relationships between our accessions and those in the database served as a validating method, confirming the potential of these genes as valid barcoding markers across *Calligonum* species. Both genes have previously been acknowledged for their efficacy in evaluating genetic diversity. In China, cpDNA, ITS, and ETS sequences verified the link between *C. jeminaicum* and *C. mongolicum* (Shi et al. 2019), while full plastome sequencing elucidated taxonomic relationships within the genus (Song et al. 2020). This approach aligns with using rpoC1 and matK of spinach cultivars to find associations among genotypes through phylogenetic analysis (Manzoor et al. 2022). Similarly, successfully utilized matK, among other loci, as a potential barcode for identifying *Paphiopedilum vietnamese* species, with 17 out of 22 *Paphiopedilum* species accurately identified (Vu et al. 2019). The phylogenetic analysis of *Poaceae* species is similar to our methodology in that it emphasizes conserved DNA motifs and selects barcode regions (matK and rpoC1) (Bhatt and Thaker 2021). For instance, genetic diversity was assessed in wheat cultivars using ISSR and SCoT markers, along with rbcL and matK, which revealed discernible variations in matK sequences, distinguishing cultivars from distinct regions (Abouseada et al. 2023).

Conclusions

The genetic variation study of *Calligonum* species in Saudi habitats provided species-specific insights. *C. tetrapterum* has demonstrated early divergence in both matK and rpoC1 phylogenetic trees, indicating a distinct evolutionary path from other species. *C. comosum*'s accessions showed high genetic similarity, particularly in matK sequences, indicating a more homogeneous genetic structure. SCoT markers effectively distinguished between *Calligonum* species, whereas IRAP markers identified traits shared by closely related species. MatK and rpoC1 genes exhibit greater genetic diversity across species, highlighting their potential for species breeding, medicinal uses, and biodiversity conservation efforts.

Acknowledgements Researchers Supporting Project number (RSPD2024R931), King Saud University, Riyadh, Saudi Arabia.

Author contributions Conceptualization; methodology; writing—review and editing; software, M.E.A.; investigation; validation; resources; funding acquisition, A. M. A. All authors have read and agreed to the published version of the manuscript.

Data availability statement All data generated or analyzed during this study are included in this published article. The sequence of five isolates were submitted to NCBI. The accession numbers for matK gene of *C. comosum* and *C. tetrapterum* are as follows: C1—OR855414.1, C2—OR855415.1, T1—OR855416.1, T2—OR855417.1, and for rpoC1 gene of C1—OR933566.1, C2—OR933567.1, T1—OR933568.1, and T2—OR933569.1.

Declarations

Conflict of interest The authors declare no conflicts of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

Abd El Azim AM, Khashaba EHK (2021) Genetic variability among three Egyptian isolates of Heterorhabditis indica using a new marker technique (SCoT). Egypt J Biol Pest Control 31:71. https://doi.org/10.1186/s41938-021-00419-0

Abouseada HH, Mohamed A-SH, Teleb SS et al (2023) Genetic diversity analysis in wheat cultivars using SCoT and ISSR

markers, chloroplast DNA barcoding and grain SEM. BMC Plant Biol 23:193. https://doi.org/10.1186/s12870-023-04196-w
Adawy SS, Jiang J, Atia MA (2014) Identification of novel sex-specific PCR-based markers to distinguish the genders in Egyptian date palm trees. Int J Agric Sci Res 4:45–54
Alaklabi A (2022) Identification and molecular phylogeny of morphologically of the medicinal plant Calligonum L'Herit. Pakistan J Med Health Sci 16:383–383
Alotaibi MO, Abd-Elgawad ME (2022) ISSR and SCoT for evaluation of hereditary differences of 29 wild plants in Al Jubail Saudi Arabian. Saudi J Biol Sci. https://doi.org/10.1016/j.sjbs.2022.01.053
Amiteye S (2021) Basic concepts and methodologies of DNA marker systems in plant molecular breeding. Heliyon. https://doi.org/10.1016/j.heliyon.2021.e08093
Antil S, Abraham JS, Sripoorna S et al (2023) DNA barcoding, an effective tool for species identification: a review. Mol Biol Rep 50:761–775. https://doi.org/10.1007/s11033-022-08015-7
Arvas YE, Marakli S, Kaya Y, Kalendar R (2023) The power of retrotransposons in high-throughput genotyping and sequencing. Front Plant Sci 14:1174339
Avarave S, Thomas J (2021) Biodiversity of South Indian tea clones with detection of plant-based adulterants in tea dust using DNA barcoding. Nat Prod Res 36:4608–4613. https://doi.org/10.1080/14786419.2021.2002325
Badr A, El-Sherif N, Aly S et al (2020) Genetic diversity among selected Medicago sativa cultivars using inter-retrotransposon-amplified polymorphism, chloroplast DNA barcodes and morpho-agronomic trait analyses. Plants 9:995
Bailey TL, Johnson J, Grant CE, Noble WS (2015) The MEME suite. Nucleic Acids Res 43:W39–W49
Barati S, Movahedin M (2021) The antioxidant effects of Calligonum extract on oxidative stress in spermatogonial stem cells culture. Pharmaceut Sci 27:521–527
Bhatt P, Thaker V (2021) A comparative study on 193 plastomes of Poaceae for validity and implications of individual barcode genes and concatenated protein coding sequences with selected plastomes of grasses from the desert of India. Meta Gene 29:100921
Chen C, Chen H, Zhang Y et al (2020) TBtools: an integrative toolkit developed for interactive analyses of big biological data. Mol Plant 13:1194–1202
Chouikh A, Chemsa AE, Aounallah C, Aounallah I, Fatma AL (2020) Phytochemical study, nutritive value, antioxidant and anti-inflammatory activities of phenolic extracts from desert plant Calligonum comosum LHér. Algerian J Biosci 1(02):68–75. https://doi.org/10.5281/ZENODO.4395515
Collard BC, Mackill DJ (2009) Start codon targeted (SCoT) polymorphism: a simple, novel DNA marker technique for generating gene-targeted markers in plants. Plant Mol Biol Report 27:86–93
DeSalle R, Goldstein P (2019) Review and interpretation of trends in DNA barcoding. Front Ecol Evol 7:302
Fal'kovich MI (2023) Biological types of case-bearers and their food associations. Lepidopterous Fauna of the USSR and Adjacent Countries: a collection of papers dedicated to Professor Alexandr Sergeevich Danilevskii 297
Gasmi A, Triki T, Benabderrahim MA, Guasmi F (2022) Assessing phenolic and molecular diversity of arta (Calligonum comosum L.), a wild Tunisian desert plant. S Afr J Bot 151:166–174
Guan C, Chachar S, Zhang P et al (2020) Inter-and intra-specific genetic diversity in Diospyros using SCoT and IRAP markers. Horticult Plant J 6:71–80
Hammer Ø, Harper DA, Ryan PD (2001) PAST: Paleontological statistics software package for education and data analysis. Palaeontol Electron 4:9
Ibrahim SD, Abd El-Hakim AF, Ali HE, Abd El-Maksoud RM (2019) Genetic differentiation using ISSR, SCoT and DNA Barcoding for Quinoa genotypes. Arab J Biotech 22:103–118
Kalendar R, Flavell AJ, Ellis THN et al (2011) Analysis of plant diversity with retrotransposon-based molecular markers. Heredity 106:520–530
Kalendar R, Shustov AV, Schulman AH (2021) Palindromic sequence-targeted (PST) PCR, version 2: an advanced method for high-throughput targeted gene characterization and transposon display. Front Plant Sci 12:691940
Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol 30:772–780
Kiani K, Rudzitis-Auth J, Scheuer C et al (2019) Calligonum comosum (Escanbil) extract exerts anti-angiogenic, anti-proliferative and anti-inflammatory effects on endometriotic lesions. J Ethnopharmacol 239:111918
Kress WJ (2017) Plant DNA barcodes: applications today and in the future. J Sytem Evol 55:291–307. https://doi.org/10.1111/jse.12254
Lennartz C, Kurucar J, Coppola S et al (2021) Geographic source estimation using airborne plant environmental DNA in dust. Sci Rep 11:16238
Leray M, Knowlton N, Ho S-L et al (2019) GenBank is a reliable resource for 21st century biodiversity research. Proc Natl Acad Sci USA 116:22651–22656. https://doi.org/10.1073/pnas.1911714116
Li S, Ramakrishnan M, Vinod KK et al (2020) Development and deployment of high-throughput retrotransposon-based markers reveal genetic diversity and population structure of Asian Bamboo. Forests 11:31. https://doi.org/10.3390/f11010031
Liu J, Liu J, Shan Y-X et al (2019) The use of DNA barcodes to estimate phylogenetic diversity in forest communities of Southern China. Ecol Evol 9:5372–5379. https://doi.org/10.1002/ece3.5128
Luo C, He X-H, Chen H et al (2010) Analysis of diversity and relationships among mango cultivars using Start Codon Targeted (SCoT) markers. Biochem Syst Ecol 38:1176–1184
Mahdy E, Rizk RM (2023) Genetic variation of arta populations (Calligonum polygonoides subsp. comosum) in Egypt: Genepools for biodiversity and afforestation. J Water Land Dev https://doi.org/10.24425/jwld.2023.143748
Manzoor MT, Ali S, Fatima A et al (2022) Analyzing local spinach plant diversity using the rpoC1 and matK DNA barcode technique. Pakistan J Agric Sci 59:1029–1034
Mishra MK, Huded AKC, Jingade P (2020) Assessment of the suitability of molecular SCoT markers for genetic analysis of coffee species. Botanica 26:184–196
Mohamed ASH, Ibrahim M, Teleb SS, Tantawy ME (2017) SEM and SCoT markers unveil new taxonomic and Genetic Insights about Some Northern African Triticum aestivum L. Cultivars Vegetos 30:1. https://doi.org/10.4172/2229-4473.1000206
Negi A, Pare A, Meenatchi R (2021) Emerging techniques for adulterant authentication in spices and spice products. Food Control 127:108113
Nehal N, Choudhary B, Nagpure A, Gupta RK (2021) DNA barcoding: a modern age tool for detection of adulteration in food. Crit Rev Biotechnol 41:767–791. https://doi.org/10.1080/07388551.2021.1874279
Purohit CS, Kumar R (2020) A review on genus Calligonum L. (Polygonaceae) from India and report Calligonum crinitum an addition for Flora of India. J Asia-Pacific Biodive 13:319–324
Rokas A (2011) Phylogenetic Analysis of Protein Sequence Data Using the Randomized Axelerated Maximum Likelihood (RAXML) Program. CP Mol Biol https://doi.org/10.1002/0471142727.mb1911s96
Sheth BP, Thaker VS (2017) DNA barcoding and traditional taxonomy: an integrated approach for biodiversity conservation. Genome 60:618–628. https://doi.org/10.1139/gen-2015-0167

Shi W, Liu P-L, Wen J, et al (2019) New morphological and DNA evidence supports the existence of Calligonum jeminaicum Z. M. Mao (Calligoneae, Polygonaceae) in China. PhytoKeys 132:53. https://doi.org/10.3897/phytokeys.132.34981

Skuza L, Szućko I, Filip E, Adamczyk A (2019) DNA barcoding in selected species and subspecies of Rye (Secale) using three chloroplast loci (matK, rbcL, trnH-psbA). Notulae Botanicae Horti Agrobotanici Cluj-Napoca 47:54–62

Song F, Li T, Burgess KS et al (2020) Complete plastome sequencing resolves taxonomic relationships among species of Calligonum L. (Polygonaceae) in China. BMC Plant Biol 20:261. https://doi.org/10.1186/s12870-020-02466-5

Stepanov I, Balapanov I, Drygina A (2020) Search of effective IRAP markers for sakura genotyping. In: BIO Web of Conferences. EDP Sciences, p 03006

Tahmasebi F, Movahedin M, Mazaheri Z (2018) Antioxidant effects of calligonum extract on ovarian tissue of PCO model: an experimental study. Int J Reprod BioMed 16:641

Tizard J, Patel S, Waugh J et al (2019) DNA barcoding a unique avifauna: an important tool for evolution, systematics and conservation. BMC Evol Biol 19:52. https://doi.org/10.1186/s12862-019-1346-y

Vu H-T, Vu Q-L, Nguyen T-D et al (2019) Genetic diversity and identification of Vietnamese Paphiopedilum species using DNA Sequences. Biology 9:9

Wu Y, Li M, Yang Y et al (2018) Authentication of small berry fruit in fruit products by DNA barcoding method. J Food Sci 83:1494–1504. https://doi.org/10.1111/1750-3841.14177

Yu J, Wu XI, Liu C et al (2021) Progress in the use of DNA barcodes in the identification and classification of medicinal plants. Ecotoxicol Environ Saf 208:111691

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.