Evaluation of genetic variability and relatedness among eight Centaurea species through CAAT-box derived polymorphism (CBDP) and start codon targeted polymorphism (SCoT) markers

Mohamed Atia Mohamed Atia a, Diaa Abd El-Moneim b, Taghreed Khaled Abdelmoneim a, Eman Hussein Reda c, Zienab Talaat Abdel Shakour c, Ali Mohamed El-Halawany d, El-Sayeda Ahmed El-Kashoury d, Khaled Ahmed Shams e, Nahla Sayed Abdel-Azim e and Mohamed-Elamir Fathy Hegazy a,f

Genome Mapping Department, Agricultural Genetic Engineering Research Institute (AGERI), Agricultural Research Center (ARC), Giza, Egypt; bDepartment of Plant Production (Genetic Branch), Faculty of Environmental and Agricultural Sciences, Arish University, Arish, Egypt; cPhytochemistry Laboratory, National Organization for Drug Control and Research, Giza, Egypt; dDepartment of Pharmacognosy, Faculty of Pharmacy, Cairo University, Cairo, Egypt; eChemistry of Medicinal Plants Department, National Research Centre, Giza, Egypt; fDepartment of Pharmaceutical Biology, Institute of Pharmaceutical and Biomedical Sciences, Johannes Gutenberg University, Mainz, Germany

ABSTRACT

Centaurea is a value-ultimate genus of medicinal plants showing high diversification levels, especially within the Mediterranean basin, and is still traditionally recognized as a complicated taxon. So far, few studies utilizing molecular markers have been done on Centaurea spp. towards a better dissection of its phylogeny and accurate assessment of genetic diversity. Here, two functional marker systems, start codon targeted (SCoT) polymorphism and CAAT box-derived polymorphism (CBDP), were implemented to assess the genetic diversity between eight wild Centaurea species in Egypt. Seventeen SCoT and 19 CBDP primers generated 197 and 179 bands, respectively. These primers generated 158 (80.2%) and 131 (73.18%) polymorphic amplicons with an average of 9.29 and 6.89 amplicons per primer, respectively. SCoT primers exhibited higher levels in % of polymorphism, % of heterozygosity, effective multiplex ratio, discriminating power, resolving power parameters compared to the CBDP. In contrast, no significant differences were observed between SCoT and CBDP in polymorphism information content, marker index and mean heterozygosity parameters. The UPGMA (unweighted pair group method with arithmetic mean) dendrogram of the combined data classified the Centaurea species under two major clades; the first comprised two species, whereas the second grouped five species. C. pallescens was kept separated as the most diverged among the eight Centaurea species. Principal component analysis (PCA) topology revealed highly similar results to those obtained from the cluster analysis. Ultimately, our results represent the first report utilising two gene-targeting marker systems as powerful techniques for assessing the genetic variability and relatedness among these eight valuable Centaurea species.

Introduction

Centaurea L. is the fourth biggest genus in the Asteraceae family, including more than 500 species worldwide, particularly in the Mediterranean regions [1]. Approximately 17 species of the genus Centaurea grow in Egypt [2, 3]. The Centaurea species are of strong interest in phytochemical and biological research due to their excellent medicinal value. Centaurea species are one of the most popular plants in folk medicine for the remedial treatment of cancer and microbial infections. In addition, Centaurea spp. has also been shown to have anti-diabetic, diuretic, antimalarial, anti-inflammatory, anti-pyretic, analgesic, anti-platelet, wound healing, anti-ulcerogenic, hepatoprotective, anti-plasmodial, cytotoxic, anti-pyretic, antioxidant, anti-bacterial, anti-fungal, and antirheumatic properties [4, 5]. Several reports have shown the presence of several compounds belonging primarily to the classes of sesquiterpene lactones, flavonoids and lignans. Besides, Centaurea was a subject of
interest for several phytochemical characterizations, due to the active constituents particularly sesquiterpene lactones [6, 7] and flavonoids [8]. The curative properties of Centaurea species may be due to the content of a wealth of bitter crystalline unsaturated lactones [9].

The species Centaurea alexandrina, Centaurea calcitrapa, Centaurea eryngioides, Centaurea glomerata, Centaurea pallescens, Centaurea pumilio, Centaurea scoparia and Centaurea sinaica are of great biological importance. C. alexandrina extract was suggested to have a potential anti-inflammatory analgesic, hypoglycemic activities, anticancer activity and was found to be active against Pseudomonas aeruginosa pathogenic activity for chest infection [10, 11]. C. calcitrapa was reported to have hypoglycemic effect and antimicrobial activity, and the seed extracts, also a potential ability for milk clotting [12]. C. eryngioides strongly inhibits the growth of the malaria parasite [13]. C. glomerata and C. pallescens have been documented to have antioxidant activity [14, 15]. C. pumilio extract has antimicrobial activity against skin infection-causing strains, suggesting its potential to maintain healthy skin [16]. Also, extracts from C. scoparia and C. sinaica have potential antitumor activity against human carcinoma cell lines [17, 18].

The genus of Centaurea still has many taxonomical problems. The systematics and the phylogenetic analysis of Centaurea have changed dramatically during the past two decades [1, 19]. To date, molecular analyses of the genus Centaurea have not shown a well-resolved phylogeny. Accordingly, it is essential to use modern molecular marker systems for enhanced taxonomy of the genus Centaurea. In the last few years, novel molecular marker techniques have been developed and used in various plant genetic studies [20]. Many new alternative and promising markers have evolved, called gene-targeting markers or functional markers [21].

The Start Codon Targeted (SCoT) technique is considered a modern DNA-marker system that targets regions of the conserved sequences surrounding the ATG start codon. For this reason, it employs a single primer as both a forward and reverse primer [22]. As the region flanking the ATG start codon is very conserved across a wide range of plant species, the SCoT marker system will be valuable for the sake of generating DNA markers in a vast range of plant species [23–25]. Recently, a genuine gene-targeted marker system called CAAT box-derived polymorphism (CBDP) has evolved to be one of the latest and easy-to-implement functional marker techniques. The CDBP marker system is designed to target specifically the CAAT box region of the promoters upstream of the genes in the plant genomes [26]. Given that the CAAT box has a distinct paradigm of nucleotides with a consensus sequence, GGCCAATCT, located -80 bp upstream of the start codon of the genes, it is a key player during the transcription of eukaryotic genes. Furthermore, it is more beneficial for several downstream applications in plant molecular genetics, taking into considerations that the recombination between marker and gene is lower than Random Amplification of Polymorphic DNA (RAPDs), Simple Sequence Repeats (SSRs) or Inter Simple Sequence Repeats (ISSRs) [27].

Accordingly, our study focused on investigating the genetic diversity and genetic relationships between eight medicinal herbs belonging to the genus Centaurea; C. alexandrina, C. calcitrapa, C. eryngioides, C. glomerata, C. pallescens, C. pumilio, C. scoparia and C. lipii (Volutaria lippii), using SCoT-PCR accompanying with CDBP-PCR techniques as gene-targeting marker systems. The fusion of the two marker systems is pivotal to identify, classify and authenticate these valuable species.

Materials and methods

Plant materials

Young fresh leaves from eight Centaurea species (Three independent individuals for each species) were collected during the flowering phase in May 2018 from six locations in Egypt (Table 1). A specimen of all collected Centaurea samples in the present study was deposited in the National Research Center’s herbarium. The Centaurea species were taxonomically identified by Prof. Dr. Ibrahim El-Garf, (Professor at Botany Department, Faculty of Science, Cairo University, Cairo, Egypt).

Plant DNA extraction

Genomic DNA was extracted from fresh parts (100 mg) of the collected eight Centaurea species using a DNAeasy Plant Mini Kit (QIAGEN, Santa Clarita, CA) and following the manufacturer’s protocol. The DNA was quantified with the Qubit 3.0 Fluorometer (Life Technologies) according to the manufacturer’s instruction. The quality of the extracted DNA was determined in terms of A260/280 ratio as an indicator for the purity of the DNA, also the DNA samples were electrophoresed using agarose gel (0.7%) to check the integrity of DNA.
SCoT-PCR amplification

The SCoT-PCR analysis was conducted according to the procedure described by [24]. A set of 17 SCoT primers was initially screened against the eight Centaurea species (Table 2). The amplified fragments were finally electrophoresed in 1.5% agarose gel and photographed using Gel Doc XR + Gel Documentation framework (Bio-Rad, USA).

CBDP-PCR amplification

Nineteen CBDP primers (Table 3), used in the present study, were designed according to [26]. The amplification fragments were electrophoresed in 1.5% agarose gel and finally photographed using Gel Doc XR + Gel Documentation framework (Bio-Rad, USA).

Data analysis

For SCoT and CBDP data analysis, the amplified bands were scored manually. The bands were scored as absent (0) or present (1) to create the binary data matrix. A similarity matrix was constructed according to the Jaccard similarity coefficient (Jaccard, 1901). For SCoT, CBDP marker systems, and combined data (SCoT + CBDP), dendrograms were developed using cluster analysis and the unweighted pair group method with arithmetic mean (UPGMA). To evaluate the efficiency of the SCoT and CBDP primers, seven informative indices were calculated as described by [24].

To determine the efficiency of applied primers, the following parameters were calculated: Expected Heterozygosity \( (H = 1 - \sum pi^2) \) [28], Polymorphism Information Content (PIC = \( 1 - \sum pi^2 - \sum \sum pi^2 pj^2 \)) [29], Effective Multiplex Ratio \( (E = \frac{n \beta}{\gamma}) \) [30], Marker Index \( (MI = E \ H_{av}) \) [30], Mean Heterozygosity \( (H_{av} = \sum H_n/np) \) [30], Discriminating Power \( (D = 1 - C) \) [31], Resolving Power \( (R = \sum I_b) \) [32], using the iMEC online tool [33].

Results

SCoT analysis

Seventeen SCoT primers were screened to assess the eight Centaurea species’ genetic relationships; all the tested SCoT primers produced reproducible and scorable bands in all species. The SCoT primers generated 197 scorable bands with an average of 11.58 bands per

Table 1. Geographical information of eight Centaurea species in this study.

| Scientific name       | Voucher ID | Date of collection | Location               | Latitude (N) | Longitude (E) | Elevation (m) |
|-----------------------|------------|--------------------|------------------------|--------------|---------------|---------------|
| Centaurea alexandrina | M/2282     | April 2016         | Marsa Marouh           | 31-23-378    | 27-01-764     | 171           |
| Centaurea pallescens  | M/2283     | April 2016         | Marsa Marouh           | 31-22-370    | 27-01-725     | 175           |
| Centaurea eryngioides | M/2284     | May 2016           | Saint-Catherine        | 28-33-208    | 33-56-913     | 81            |
| Centaurea scoparia    | M/2278     | May 2016           | Saint-Catherine        | 28-33-208    | 33-56-913     | 81            |
| Centaurea glomerata   | M/2280     | Sept. 2016         | Tanta                  | 30-56-525    | 30-58-331     | 16            |
| Centaurea lipii       | M/2281     | April 2017         | Suez                   | 29-38-165    | 32-18-237     | 25            |
| Centaurea calchropia  | M/2279     | May 2017           | Beheira                |              |               |               |
| Centaurea pumilio     | M/2285     | March 2016         | North-Coast            | 30-54-906    | 29-26-863     | 77            |

Table 2. Primer codes, primer sequences, number of total bands, polymorphic bands, percentage of polymorphism and marker efficiency parameters of SCoT primers.

| Primer Code | Primer sequence | Number of bands | % of polymorphism | H% | PIC | EMR | MI | H_av | D | R |
|-------------|-----------------|-----------------|-------------------|----|-----|-----|----|-------|---|---|
| SCoT-1      | CAACAATGGCTACCA | 6               | 100.00            | 0.469 | 0.379 | 2.250 | 0.010 | 0.022 | 0.864 | 3.000 |
| SCoT-3      | CAACAATGGCTACCCG | 12              | 91.67             | 0.489 | 0.369 | 6.875 | 0.005 | 0.035 | 0.674 | 6.250 |
| SCoT-4      | CAACAATGGCTACCT | 11              | 72.73             | 0.496 | 0.366 | 6.000 | 0.006 | 0.034 | 0.705 | 3.500 |
| SCoT-5      | CAACAATGGCTACCGA | 17              | 88.24             | 0.482 | 0.373 | 6.875 | 0.004 | 0.024 | 0.838 | 6.750 |
| SCoT-6      | CAACAATGGCTACCGC | 17              | 88.24             | 0.500 | 0.364 | 6.825 | 0.004 | 0.032 | 0.744 | 9.750 |
| SCoT-8      | CAACAATGGCTACCGT | 20              | 85.00             | 0.489 | 0.369 | 11.500 | 0.003 | 0.035 | 0.671 | 9.500 |
| SCoT-9      | CAACAATGGCTACCGA | 20              | 85.00             | 0.500 | 0.364 | 9.750 | 0.003 | 0.030 | 0.764 | 9.500 |
| SCoT-10     | CAACAATGGCTACCGC | 9               | 66.67             | 0.435 | 0.394 | 6.125 | 0.006 | 0.037 | 0.540 | 2.750 |
| SCoT-11     | AAGCAATGGTACCACCA | 15              | 80.00             | 0.427 | 0.398 | 10.375 | 0.004 | 0.037 | 0.523 | 6.250 |
| SCoT-12     | ACAGCAATGGGACACCG | 11              | 81.82             | 0.494 | 0.367 | 6.125 | 0.006 | 0.034 | 0.693 | 5.250 |
| SCoT-13     | ACAGCAATGGGACACCG | 12              | 58.33             | 0.430 | 0.396 | 8.250 | 0.004 | 0.037 | 0.530 | 4.000 |
| SCoT-14     | ACAGCAATGGGACACCG | 11              | 72.73             | 0.487 | 0.370 | 6.375 | 0.006 | 0.035 | 0.667 | 3.250 |
| SCoT-15     | ACAGCAATGGGACACCGG | 12             | 100.00            | 0.498 | 0.365 | 6.375 | 0.005 | 0.033 | 0.720 | 4.750 |
| SCoT-23     | CACATGATCCACACAG | 4               | 25.00             | 0.117 | 0.482 | 3.750 | 0.004 | 0.014 | 0.123 | 0.500 |
| SCoT-24     | CACATGATCCACACCAX | 8              | 62.50             | 0.342 | 0.430 | 6.250 | 0.005 | 0.033 | 0.392 | 3.500 |
| SCoT-25     | ACCATGATCCACACCGG | 7              | 57.14             | 0.337 | 0.432 | 5.500 | 0.006 | 0.033 | 0.386 | 3.000 |
| SCoT-27     | ACCATGATCCACACCGTG | 5            | 100.00            | 0.495 | 0.366 | 2.250 | 0.012 | 0.028 | 0.804 | 2.000 |

Total 197 158 80.20

Average 11.58 9.29 0.440 0.387 6.662 0.005 0.031 0.626 4.912
Table 3. Primer codes, primer sequences, number of total bands, polymorphic bands, percentage of polymorphism and marker efficiency parameters of CBDP primers.

| Primer code | Primer sequence | Number of bands | % of polymorphism | %H | PIC | EMR | MI | H.av | D | R |
|-------------|----------------|----------------|-------------------|-----|-----|-----|----|------|---|---|
| CAAT-1      | TGAGCACGATCCACAATGC | 11 11 11 100.00 | 0.4791 0.365 4.375 0.005 0.024 0.845 | 5.750 |
| CAAT-2      | TGAGCACGATCCACAATAT | 11 7 63.64 0.3750 | 0.409 8.250 0.004 0.035 | 0.440 | 5.000 |
| CAAT-3      | TGAGCACGATCCACAATACC | 8 6 75.00 0.4409 | 0.382 5.375 0.007 0.037 | 0.552 | 2.750 |
| CAAT-4      | TGAGCACGATCCACAATAG | 11 8 72.73 0.4994 | 0.355 7.250 0.005 0.032 | 0.734 | 6.000 |
| CAAT-5      | TGAGCACGATCCACAATCTA | 15 11 73.33 0.4728 | 0.368 9.250 0.004 0.036 | 0.622 | 6.500 |
| CAAT-6      | TGAGCACGATCCACAATCAG | 12 9 75.00 0.4991 | 0.365 9.125 0.004 0.033 | 0.714 | 8.750 |
| CAAT-7      | TGAGCACGATCCACAATCGA | 14 12 85.71 0.4994 | 0.355 7.250 0.005 0.032 | 0.734 | 6.000 |
| CAAT-8      | TGAGCACGATCCACAATTG | 17 | 13 76.47 | 0.4973 0.356 | 9.125 0.004 0.033 | 0.714 | 8.750 |
| CAAT-9      | TGAGCACGATCCACAATGAT | 9 4 44.44 0.4244 | 0.389 6.250 0.006 0.037 | 0.521 | 2.000 |
| CAAT-10     | TGAGCACGATCCACAATGTTT | 10 9 90.00 0.4688 | 0.369 6.250 0.006 0.037 | 0.612 | 5.000 |
| CAAT-11     | TGAGCACGATCCACAATGTCGC | 12 9 75.00 0.4922 | 0.358 5.250 0.005 0.027 | 0.811 | 4.000 |
| CAAT-12     | TGAGCACGATCCACAATATA | 9 5 55.56 0.4614 | 0.373 5.750 0.006 0.037 | 0.595 | 2.000 |
| CAAT-13     | TGAGCACGATCCACAATACGA | 4 2 50.00 0.1172 | 0.472 3.750 0.004 0.014 | 0.123 | 0.500 |
| CAAT-14     | TGAGCACGATCCACAATTGCG | 6 1 16.67 0.0408 | 0.478 5.875 0.001 0.005 | 0.042 | 0.250 |
| CAAT-15     | TGAGCACGATCCACAATA | 4 1 25.00 0.0408 | 0.478 5.875 0.001 0.005 | 0.042 | 0.250 |
| CAAT-16     | TGAGCACGATCCACAATCG | 7 6 85.71 0.4362 | 0.384 4.750 0.008 0.037 | 0.544 | 4.000 |
| CAAT-17     | TGAGCACGATCCACAATTGG | 6 4 66.67 0.4965 | 0.356 3.250 0.010 0.034 | 0.712 | 2.000 |
| CAAT-18     | TGAGCACGATCCACATTAG | 6 6 | 100.00 | 0.4783 | 0.365 | 2.375 0.010 0.024 | 0.848 | 3.250 |
| CAAT-19     | TGAGCACGATCCACATTAC | 7 7 100.00 | 0.4228 | 0.390 | 2.125 0.008 | 0.016 | 0.912 | 4.250 |
| **Total**   |                | **179** 131 | 73.18 | **9.42** 6.89 | **0.395** | **0.391** | **5.658** | **0.005** | **0.028** | **0.571** | **3.684** |

primer (Table 2). Out of this total number of bands, 158 (80.2%) were identified as polymorphic, with an average of 9.29 bands per primer. The number of amplicons per primer ranged from 4 bands (primer SCoT-23) to 20 bands (primers SCoT-8 and SCoT-9). Meanwhile, the polymorphic bands number per primer ranged from one band (primer SCoT-23) to 17 bands (primers SCoT-8 and SCoT-9). The expected heterozygosity values revealed a narrower range between 0.117 to 0.500, with an average of 0.440. Notably, 3 out of 17 SCoT primers showed the lowest expected heterozygosity values (SCoT-23, SCoT-24 and SCoT-25). Furthermore, the values range of the polymorphism information content was between 0.364 and 0.482, with an average of 0.387. The highest PIC values were observed in the SCoT primers SCoT-8 and SCoT-9. Also, the effective multiplex ratio (EMR) revealed relatively high values, ranging from 2.125 (primer SCoT-1) to 9.250 (primer SCoT-5), with an average of 5.658. The marker index values ranged from 0.001 (primer SCoT-14) to 0.010 (primers SCoT-17 and SCoT-18), with an average of 0.391. The expected heterozygosity values ranged between 0.355 (primers SCoT-6 and SCoT-7) and 0.478 (primer SCoT-14), with an average of 0.391. For the discrimination power, the values ranged from 0.042 (primer SCoT-14) to 0.912 (primer SCoT-19). Additionally, the resolving power values were found to be between 0.250 (primers SCoT-14 and SCoT-15) and 8.750 (primer SCoT-8) (Table 3).

**CBDP analysis**

Nineteen CBDP primers were used to assess the genetic relationships among the eight Centaurea species. All the utilized CBDP primers generated reproducible patterns with clearly scorable bands in all species. The CBDP primers yielded 179 scorable amplicons, out of which 131 amplicons (73.18%) were polymorphic. The number of amplicons per primer ranged from 4 amplicons (primers CAAT-13 and CAAT-15) to 17 amplicons (primer CAAT-8), with an average of 9.42 amplicons per primer. As a part of this, the number of polymorphic bands per primer ranged from one band (primers CAAT-14 and CAAT-15) to 13 bands (primer CAAT-8), with an average of 6.89 polymorphic bands per primer. The percentage values of the expected heterozygosity ranged from 0.040 (primer CAAT-14) to 0.499 (primers CAAT-6 and CAAT-7), with an average of 0.395. Moreover, the polymorphism information content values ranged between 0.355 (primers CAAT-6 and CAAT-7) and 0.478 (primer CAAT-14), with an average of 0.391. The effective multiplex ratio values ranged from 2.125 (primer CAAT-19) to 9.250 (primer CAAT-5), with an average of 5.658. The marker index values ranged from 0.001 (primer CAAT-14) to 0.010 (primers CAAT-17 and CAAT-18). Similarity, the expected heterozygosity values ranged from 0.007 (primer CAAT-15) to 0.037 (primers CAAT-3, −9, −10, −12 and CAAT-16). For the discrimination power, the values ranged from 0.042 (primer CAAT-14) to 0.912 (primer CAAT-19). Additionally, the resolving power values were found to be between 0.250 (primers CAAT-14 and CAAT-15) and 8.750 (primer CAAT-8) (Table 3).

**Analysis of molecular phylogeny and genetic similarities**

The phylogenetic trees based on the unweighted pair group method of the arithmetic averages analysis of SCoT, CBDP and combined data were constructed for...
the eight *Centaurea* species (Figure 1). For the SCoT marker system, the dendrogram comprised one major cluster (grouping seven out of the eight *Centaurea* species) in addition to another cluster that includes only *C. pallescens* (the most diverged among the eight *Centaurea* species). The major cluster was further divided into two subclusters; the first subcluster included only *C. eryngioides*, whereas the second subcluster was subsequently divided into two groups. The first group comprised four species (*C. pumilio*, *C. alexandrina*, *C. scoparia* and *C. lipii*), while the second group involved two species (*C. glomerata* and *C. calcitrapa*) (Figure 1A). On the other hand, principal component analysis (PCA) analysis of the SCoT data exhibited highly consistent results compared to the grouping obtained from the SCoT dendrogram (Figure 2A). The SCoT PCA plot revealed that the clustering topology is comparable to that obtained by the clustering analysis except for *C. pumilio*, which becomes closer to *C. lipii* than *C. alexandrina*.

For the CBDP marker system, the dendrogram was divided into two main clusters; the first cluster involved two species (*C. alexandrina* and *C. calcitrapa*). The second cluster grouped the rest of the eight *Centaurea* species. The second cluster was subdivided into two subclusters: the first subcluster comprised two species (*C. eryngioides* and *C. lipii*), while the second subcluster grouped four species (*C. pumilio* with *C. scoparia*, as the most genetically similar, *C. pallescens* and *C. glomerata*) (Figure 1B). Additionally, the PCA analysis of the CBDP data revealed a high degree of consistency with the CBDP dendrogram’s topology (Figure 2B). The CBDP PCA plot showed that the *Centaurea* species’ clustering remained the same as that obtained by the dendrogram analysis except for *C. pumilio*, which became closer to *C. pallescens* than *C. scoparia*.

Furthermore, the SCoT and CBDP scored data were further combined to reach more comprehensive genome coverage and generate deeper relationships among the eight *Centaurea* species. The topology of

![Figure 1](image1.png)

**Figure 1.** UPGMA cluster analysis based on Jaccard’s similarity coefficient of SCoT analysis (A), CBDP analysis (B) and combined (SCoT + CBDP) data (C) of the eight *Centaurea* species.

![Figure 2](image2.png)

**Figure 2.** Principal Component Analysis (PCA) of the PCR data showing the two-dimensional (PC1 and PC2) plot of SCoT analysis (A), CBDP analysis (B) and combined (SCoT + CBDP) data (C) of the eight *Centaurea* species.
the combined dendrogram was similar to the SCoT dendrogram with some variations. The combined dendrogram comprised one major cluster gathering seven out of the eight *Centaurea* species. The second cluster includes only *C. pallescens*; the most diverged among the eight *Centaurea* species. Again, the major cluster was subdivided into two subclusters; the first comprised two species (*C. alexandrina* and *C. calcitrapa*), while the second subcluster grouped five species (*C. pumilio* with *C. scoparia*, as the most genetically similar; *C. eryngioides*, *C. lipii* and *C. glomerata*) (Figure 1C). The genetic similarities based on Jaccard’s coefficient showed that the highest value was identified between *C. scoparia* and *C. pumilio*, whereas the lowest similarity value was detected between *C. eryngioides* and *C. calcitrapa* (Table 4). Furthermore, the PCA analysis of the combined data exhibited a supreme degree of consistency with the dendrogram’s topology of the combined data (Figure 2C).

### Discussion

Genus *Centaurea* is a typical example of complex development and problematic classification. Long ago, *Centaurea* was defined as a polyphyletic genus, and molecular studies showed that some clades, including the former section *Centaurea*, had to be excluded in order to make the genus monophyletic [34]. Recent studies identify three subgenera, namely *Lopholoma*, *Cyanus* and *Centaurea*, with the latter being the most species-rich and divided into several sections [34]. Species of the genus *Centaurea* are distributed throughout the Mediterranean and are featured by a basic chromosome number of \( x = 9 \) [34].

Although the genus *Centaurea* exhibited high diversity levels for its bioactive compounds, particularly essential oils, such as triterpenes, sesquiterpene, flavonoids and lignans [6, 35], this medicinally valuable genus still had morphologically numerous taxonomical issues. Therefore, to evaluate the genetic relationships and assess the level of polymorphism among the eight *Centaurea* species (*C. alexandrina*, *C. calcitrapa*, *C. eryngioides*, *C. glomerata*, *C. pallescens*, *C. pumilio*, *C. scoparia* and *C. lipii* (*Volutaria lippii*)), two efficient functional marker systems (SCoT and CBDP) were implemented.

Remarkably, during the last decade, few molecular studies were released to investigate the genetic diversity and relationships between *Centaurea* spp., focusing on ITS barcode or microsatellite markers for authentication. Recently, a study conducted by Dogan et al. [36] concluded that ISSR is a powerful tool in resolving the genetic relationships within problematic taxonomical entities including the *Centaurea* species (*C. ptosimopappoides*, and *C. straminicephala*). Moreover, Yildirim et al. [37], studied the genetic relatedness among 16 *Centaurea* species in the eastern Anatolia region of Turkey. They concluded that the use of RAPD marker system compared to the Fatty acid methyl ester (FAME) profiles gives a consistent result [37]. López-Pujol et al. [38] used the SSR marker system and demonstrated that there is a lack of correlation between genetic-based classification and morpho-based classification due to the allopatric diversification of the genus *Centaurea*.

These few studies, however, did not include any of the *Centaurea* species found in Egypt in their analysis and used less reproducible marker systems (RAPD, SSR, ISSR etc.) [36–39]. With the evolution of many improved novel marker techniques, currently, they have become a not preferred marker technique in plant genetic diversity studies. New functional marker systems proved their ability to provide good reproducibility and increased resolution power compared to the traditional random marker systems. For that, we implemented two functional marker systems (SCoT and CBDP) to investigate the genetic relationships and the diversity among eight Egyptian wild *Centaurea* spp.

Our study observed high levels in % of polymorphism, discriminating power, resolving power parameters in SCoT compared to the CBDP system, suggesting the advanced discriminatory capacity of both marker systems. Based on the resolving power values (\( R \)) of the primers SCoT-6, SCoT-7 and SCoT-8, these primers could identify more than 50 *Centaurea* genotypes according

|                  | *C. glomerata* | *C. calcitrapa* | *C. scoparia* | *C. pumilio* | *C. alexandrina* | *C. lipii* | *C. eryngioides* | *C. pallescens* |
|------------------|----------------|-----------------|--------------|--------------|------------------|------------|------------------|-----------------|
| *C. glomerata*   | 100%           | –               | –            | –            | –                | –          | –                | –               |
| *C. calcitrapa*  | 63%            | 100%            | –            | –            | –                | –          | –                | –               |
| *C. scoparia*    | 66%            | 59%             | 100%         | –            | –                | –          | –                | –               |
| *C. pumilio*     | 62%            | 57%             | 69%          | 100%         | –                | –          | –                | –               |
| *C. alexandrina*| 58%            | 65%             | 55%          | 60%          | 100%             | –          | –                | –               |
| *C. lipii*       | 61%            | 56%             | 65%          | 65%          | 57%              | 100%       | –                | –               |
| *C. eryngioides* | 53%            | 48%             | 58%          | 59%          | 56%              | 65%        | 100%             | –               |
| *C. pallescens*  | 52%            | 50%             | 58%          | 56%          | 54%              | 58%        | 55%              | 100%            |

Table 4. Genetic similarities between the eight *Centaurea* species based on Jaccard’s similarity coefficient of combined data (SCoT + CBDP).
to the equation of Prevost and Wilkinson \((0.15x+1.78-R,\) where \(x\) is the number of genotypes identified) [32]. Although both SCoT and CBPD markers can be effectively regarded as markers of choice for studying genetic diversity in *Centaurea*, it seems that SCoT can perform better than CBPD markers to interpret the polymorphism among the studied *Centaurea* spp. Our results are in line with previous reports that the SCoT markers showed a higher total number of bands, number of polymorphic bands, \(D\) and \(R\) values than CBPD markers when applied in *Triticum urartu* [40] and *Aegilops triuncialis* [41]. Other reports revealed that CBPD could perform a little better than the SCoT marker system in *Simmondsia Chinensis* [42] and *Andrographis paniculata* [43].

Ultimately, to the best of our knowledge, this is the first study focusing on the applications of functional marker systems such as SCoT and CBPD to assesses the genetic diversity and analyze the genetic relationships among those collections of wild Egyptian *Centaurea* species.

**Conclusions**

Our results revealed a high level of genetic diversity among the eight Egyptian *Centaurea* species. These findings were supported by various statistical analyses such as UPGMA cluster analysis, PCA and Jaccard’s similarity analysis, which showed a high divergence among the studied species. Hence, we concluded that the SCoT combined with CBPD marker systems could be used efficiently to evaluate the genetic diversity, especially in complex and problematic classification genus such as Centaurea.

**Disclosure statement**

The authors report no conflict of interest.

**Data availability statement**

All data that support the findings reported in this study are available from the corresponding author upon reasonable request.

**ORCID**

Mohamed Atia Mohamed Atia http://orcid.org/0000-0003-2462-1269
Diaa Abd El-Moneim http://orcid.org/0000-0003-3285-0563
Taghreed Khaled Abdelmoneim http://orcid.org/0000-0001-9171-863X

**References**

[1] Garcia-Jacas N, Susanna A, Mozaffarian V, et al. The natural delimitation of *Centaurea* (Asteraceae: Cardueae): ITS sequence analysis of the *Centaurea jacea* group. Plant Syst Evol. 2000;223(3–4):185–199.

[2] Mohamed TA, Elshamy AI, Abd‐ElGawad AM, et al. Cytotoxic and chemotaxonomic study of isolated metabolites from *centaurea aegyptiaca*. J Chin Chem Soc. 2021;68(1):159–168.

[3] Bakr RO, Mohamed SA, Ayoub N. Phenolic profile of centaurea aegyptiaca L. growing in Egypt and its cytotoxic and antiviral activities. Afr J Tradit Complement Altern Med. 2016;13(6):135–143.

[4] Formisano C, Cirignano C, Rigano D, et al. Antiproliferative activity against leukemia cells of sesquiterpene lactones from the turkish endemic plant *Centaurea drabifolia* subsp. *detonsa*. Fitoterapia. 2017;120:98–102.

[5] Köse YB, Işcan G, Demirci B, et al. Antimicrobial activity of the essential oil of *Centaurea aladagensis*. Fitoterapia. 2007;78(3):253–254.

[6] Reda EH, Shakour ZTA, El-Halawany AM, et al. Comparative study on the essential oils from five wild Egyptian centaurea species: effective extraction techniques, antimicrobial activity and in-silico analyses. Antibiotics. 2021;10(3):252.

[7] López-Rodríguez M, García VP, Zater H, et al. Cynaratriol, a sesquiterpene lactone from *Centaurea musimonum*. Acta Crystallogr Sect E Struct Rep . 2009;65(Pt 8):o1867-8.

[8] Flamini G, Stoppelli G, Morelli I, et al. Secondary metabolites from *Centaurea isaurica* from Turkey and their chemotaxonomical significance. Biochem Syst Ecol. 2004;32(6):553–557.

[9] Kurian A, Sankar MA. Medicinal plants. New Delhi: New India Publishing; 2007.

[10] Hegazy MEF, Abdelfatah S, Hamed AR, et al. Cytotoxicity of 40 egyptian plant extracts targeting mechanisms of drug-resistant cancer cells. Phytomedicine. 2019;59:152771.

[11] Kubacey TM, Haggag EG, El-Toumy SA, Ahmed AA, et al. Biological activity and flavonoids from *centaurea alexanderina* leaf extract. J. Pharm. Res. 2012;5(6):3352–3361.

[12] Dimkić I, Petrović M, Gavrilović M, et al. New perspectives of purple starthistle (*centaurea calcitrapa*) leaf extracts: phytochemical analysis, cytotoxicity and antimicrobial activity. AMB Expr. 2020;10(1):1–21.

[13] Senatore F, Arnold NA, Bruno M. Volatile components of *Centaurea eryngiodes* Lam. and *Centaurea iberica* Trev.
var. hermonis Bois. Lam., two Asteraceae growing wild in Lebanon. Nat Prod Res. 2005;19(8):749–754.

[14] Abdallah HM, Mohamed MA, Abdou AM, et al. Protective effect of centaurea pallescens del. against CCl 4-induced injury on a human hepatoma cell line (Huh7). Med Chem Res. 2013;22(12):5700–5706.

[15] El Toumy SA, Omara EA, Brouard I, et al. Flavonoids from centaurea glomerata and antioxidant activity of its extract. Planta Med. 2011;77(12):PG11.

[16] Mostafa E, Fayed MA, Radwan RA, et al. SSRome: an integrated database of genetic characterization, agro-morphological and physiological evaluation of grafted tomato under salinity stress conditions. Agronomy. 2020;10(12):1948.

[17] Dogan B, Duran A, Çetin Ö, et al. Study of phylogenetic relationship of turkish species of klasea (asteraceae) based on ISSR amplification. PhytoKeys. 2015;(56):29–40.

[18] Dogan B, Duran A, Çetin Ö, et al. Study of phylogenetic relationship of turkish species of klasea (asteraceae) based on ISSR amplification. PhytoKeys. 2015;(56):29–40.

[19] Ahmed SA, Kamel EM. Cytotoxic activities of flavonoids of certain desert plants growing in Saudi Arabia. Saudi Pharm J. 2018;26(2):198–204.

[20] Mostafa E, Fayed MA, Radwan RA, et al. SSRome: an integrated database of genetic characterization, agro-morphological and physiological evaluation of grafted tomato under salinity stress conditions. Agronomy. 2020;10(12):1948.

[21] Mostafa E, Fayed MA, Radwan RA, et al. SSRome: an integrated database of genetic characterization, agro-morphological and physiological evaluation of grafted tomato under salinity stress conditions. Agronomy. 2020;10(12):1948.

[22] Collard BC, Mackill DJ. Start codon targeted (SCoT) polymorphism: a simple, novel DNA marker technique for generating gene-targeted markers in plants. Plant Mol Biol Rep. 2009;27(1):86–93.

[23] Abdeldym EA, El-Mogy MM, Abdellateaf HR, et al. Genetic characterization, agro-morphological and physiological evaluation of grafted tomato under salinity stress conditions. Agronomy. 2020;10(12):1948.

[24] Abouseadah HH, Atia MA, Younis IF, et al. Gene-targeted molecular phylogeny, phytochemical profiling, and antioxidant activity of nine species belonging to family cactaceae. Saudi J Biol Sci. 2020;27(6):1649–1658.

[25] Gowayed SMH, Abd El-Moneim D. Detection of genetic divergence among some wheat (Triticum aestivum L.) genotypes using molecular and biochemical indicators under salinity stress. PLoS One. 2021;16(3):e0248890.

[26] Singh AK, Rana MK, Singh S, et al. CAAT box-derived polymorphism (CBDP): a novel promoter-targeted molecular marker for plants. J Plant Biochem Biotechnol. 2014;23(2):175–183.

[27] Andersen JR, Lübbertstedt T. Functional markers in plants. Trends Plant Sci. 2003;8(11):554–560.

[28] Liu BH. Statistical genomics: linkage, mapping, and QTL analysis. London, New York: CRC Press; 2017.

[29] Botstein D, White RL, Skolnick M, et al. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am J Hum Genet. 1980;32(3):314–331.

[30] Powell W, Morgante M, Andre C, et al. The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. Mol Breeding. 1996;2(3):225–238.

[31] Arendt E, Heide-Marie M, Claudine D, et al. Optimization of the choice of molecular markers for varietal identification in Vitis vinifera L. Theor Appl Genet. 1999;98(1):171–177.

[32] Prevost A, Wilkinson MJ. A new system of comparing PCR primers applied to ISSR fingerprinting of potato cultivars. Theor Appl Genet. 1999;98(1):107–112.

[33] Amiryousefi A, Hyvonen J, Poczai P. iMEC: online marker efficiency calculator. Appl Plant Sci. 2018;6(6):e01159.

[34] Hilpold A, Garcia-Jacas N, Vilatersana R, et al. Taxonomical and nomenclatural notes on centaurea: a proposal of classification, a description of new sections and subsections, and a species list of the redefined section centaurea. Collect Bot. 2014;33:1–29.

[35] Shoeb M, Jaspers M, MacManus SM, et al. Anti-Colon cancer potential of phenolic compounds from the aerial parts of centaurea gigantea (asteraceae). J Nat Med. 2007;61(2):164–169.

[36] Dogan B, Duran A, Çetin Ö, et al. Study of phylogenetic relationship of turkish species of klasea (asteraceae) based on ISSR amplification. PhytoKeys. 2015;(56):29–40.

[37] Dogan B, Duran A, Çetin Ö, et al. Study of phylogenetic relationship of turkish species of klasea (asteraceae) based on ISSR amplification. PhytoKeys. 2015;(56):29–40.

[38] López-Pujol J, López-Vinyallonga S, Susanna A, et al. Speciation and genetic diversity in centaurea subsect. Sci Rep. 2016;6(1):37814–37818.

[39] López-Alvarado J, Mameli G, Farris E, et al. Islands as a crossroad of evolutionary lineages: a case study of centaurea sect. Centaurea (compositae) from Sardinia (mediterranean basin). PloS One. 2020;15(2):e0228776.

[40] Cholamian F, Etminan A, Changizi M, et al. Assessment of genetic diversity in triticum urartu thumanjan ex gidilayn accessions using start codon targeted polymorphism (SCoT) and CAAT-box derived polymorphism (CBDP) markers. Biotechnol Biotechnol Equip. 2019;33(1):1653–1662.

[41] Khodaei L, Azizinezhad R, Etminan AR, et al. Assessment of genetic diversity among iranian aegilops triuncialis accessions using ISSR, SCoT, and CBDP markers. J Genet Eng Biotechnol. 2021;19(1):1–9.

[42] Heikrujum M, Kumar J, Agrawal V. Genetic diversity analysis among male and female jojoba genotypes employing gene targeted molecular markers, start codon targeted (SCoT) polymorphism and CAAT box-derived polymorphism (CBDP) markers. Meta Gene. 2015;5:90–97.

[43] Tiwari G, Singh R, Singh N, et al. Study of arbitrarily amplified (RAPD and ISSR) and gene targeted (SCoT and CBDP) markers for genetic diversity and population structure in kalmehg [andrographis paniculata (burm. f.) nees]. Ind Crops Prod. 2016:86:1–11.