PHYLOGENETIC STUDY OF TEN SPECIES FROM CENTAUREA (ASTERACEAE) IN DUHOK CITY, KURDISTAN REGION-IRAQ

Rizgar yaseen ismail1, Joytay jassim muhammed 2, Dalal Yousif Khudhur Sinjare2
1 Biology Department, Faculty of Science, University of Zakho, Kurdistan region-Iraq
Corresponding author: mimory@uod.ac
2 Department of Scientific Research Center, College of Science, University of Duhok, Duhok, Kurdistan region-Iraq

ABSTRACT

The current research aimed to estimate the evolutionary relationships of ten Centaurea L. species growing naturally in the Duhok City, Kurdistan region of Iraq. The combing Start Codon Targeted (SCoT) markers with Internal Transcribed Spacer (ITS) gene region barcode were performed. To detect the DNA sequence variations and phylogenetic tree reconstruction, the Dice similarity matrix, the unweighted pair group method with arithmetic mean (UPGMA) clustering and Maximum Likelihood (ML) methods were applied. 104 polymorphic bands were scored with an average of 10.4. The Polymorphic Information Content (PIC) and Resolving Power (Rp) values ranged between (0.24 to 0.36) and 3.4 in primer (SCoT1) to 12 in primer (SCoT53) with an average of 0.319 and 5.74 respectively. The lowest similarity value was 0.52 between C. behen L. and C. solstitialis L., while the highest was 0.82 between C. balsamita and C. rigida. The reconstructed polytomous dendrogram was as follows: clade one; C. solstitialis L.; clade 2, C. balsamita Lam. and C. virgata Lam.; clade three subdivide into two subclades: C. iberica Trev. ex Spreng., C. hayalolepis Boiss., C. brugueriana (DC) Hand. Mazz. and C. gigantea Sch. Bip. Ex Boiss., C. regia Boiss., C. rigida Banks & Sol., C. behen Lam. Furthermore, C. brugueriana (DC) Hand. Mazz., C. iberica Trev. ex Spreng, C. behen L., C. solstitialis L. and C. balsamita Lam. were nested with National Center for Biotechnology Information (NCBI). In contrast, the remaining taxa were mixed with other closely related species. Thus, SCoT markers and ITS DNA barcode were considerably effective for investigating the evolutionary relationships of Centaurea taxa.

KEY WORDS: Phylogenetic Study, Centaurea, Asteraceae, Duhok, Kurdistan, Iraq.

1. INTRODUCTION

Asteraceae is one of the biggest families of angiosperms, comprises of 43 tribes, 12 subgenera, 1600 genera, and up to 30000 species (Funk et al. 2009). The Cardueae tribe has been divided into 12 main subtribes including Centaerinae (Herrando-Moraira et al. 2019). Considerable studies conducted to delimit, determine, and classify the subtribe Centaerinae (Wagenitz and Hellwig, 2000; Garcia-Jacas et al., 2000; Greuter, 2003; Herrando-Moraira et al. 2019). Herrando-Moraira et al. 2019 recognized 12 monophyletic genera containing Centaurea. The genus Centaurea s.l., comprises approximately 400-700 species which considered as one of the largest genera of Daisy family. This genus is widely distributed in Mediterranean basin to Irano–Turanian areas (Hellwig, 2004).

Taxonomically, Centaurea regards as a problematic genus due to considerable variation on morphology, karyology, and pollen diversity (Wagenitz 1955, Uysal et al., 2017; Sirin et al., 2022). In the middle of fifteenth, Wagenitz 1955 separated the Centaurea species into three distinct groups based on pollen structure and diversity: Cyanus/Montana, Acrocentron and Jacea. Furthermore, Garcia-Jacas et al. (2001) approved the finding when evolutionary relationships were studied. Thirteen years later, Hilpold et al. (2014a) divided the Centaurea s. lat. morphologically into two distinct groups: C. sect. Phalolepis (Cass.) DC., and sect. Centaurea (Cass.) DC. On the same year, the genus Centaurea subdivided to 3 subgenera: Centaurea, Cyanus (Mill.) Cass. ex-Hayek and Lopholoma (Cass.) Dobrocz (Hilpold et al., 2014b).

This considerable variation was obvious in Iran with 70 species (Wagenitz, 1980), Turkey with 219 taxa (Özbek, 2021), and mediterranean region with 250 (Susanna & Garcia-Jacas, 2007). Consequently, the Centaurea as a genus represents as an outstanding model plant for evolutionary process analysis, introgression, hybridization, gene flow revealing (Hilpold et al., 2014b; Garcia-Jacas, et al., 2009).

In Iraq, there are four morphological groups of Centaurea divided into fourteen sections including 43 native species distributed all over Iraq mainly in restricted mountain areas (Ghazanfar et al., 2019).

Thus, enormous molecular studies have been conducted to solve these taxonomic problems in the genus Centaurea. Since the early twentieth, the genus Centaurea enormously subjected to molecular applications in terms of DNA barcoding, predominantly Internal Transcribed Spacer (ITS) gene regions (Garcia-Jacas et al. 2000, 2001; Font et al., 2009; Hilpold et al. 2014b; Amelas et al. 2018; Lopez-Alvarado et al., 2020; Sirin et al., 2022), DNA markers; simple sequence repeats and Inter Simple Sequence Repeats (a López-Vinyallonga, et al., 2015; Atasagun, 2022). However, Start Codon Targeted (SCoT) is widely used on Asteraceae family and some of these researches focusing on Centaurea species (Feng et al., 2016; Kaminska et al., 2019; Jędrezejczyk 2020; Atia et al., 2021; Jamshidi et al., 2024). In the current study, the genetic variations and species relationships are investigated using DNA markers (SCoT) and nuclear gene (rDNA) barcodes. The achieved results might help for reconstruct concert phylogenetic tree for all remaining taxa in Iraqi flora.

* Corresponding author
This is an open access under a CC BY-NC-SA 4.0 license (https://creativecommons.org/licenses/by-nc-sa/4.0/).
2. MATERIAL AND METHODS

2.1 Plant samples

The plant materials were samples from the different locations of the Duhok city – Kurdistan Region of Iraq. The collected samples were dried and the herbarium specimens were prepared then deposited in the Duhok Province University Herbarium (DPUH) (Table 1 and Figure 1).

Table 1: showing Centaurea taxa codes, names, locations, coordinates, and herbarium specimens identification numbers.

<table>
<thead>
<tr>
<th>Codes</th>
<th>Taxa names</th>
<th>Location</th>
<th>Coordinates</th>
<th>Herbarium specimens ID No</th>
</tr>
</thead>
<tbody>
<tr>
<td>1C</td>
<td><em>C. solstitialis</em></td>
<td>Zakho City / sheransh</td>
<td>37.23168945N 42.84625196E</td>
<td>3937</td>
</tr>
<tr>
<td>2C</td>
<td><em>C. bruguierana</em></td>
<td>Duhok Baedre</td>
<td>36.74798584N 43.24991194E</td>
<td>3949</td>
</tr>
<tr>
<td>3C</td>
<td><em>C. balsamita</em></td>
<td>Zakho Sere Reza</td>
<td>37.11914062N 42.67837174E</td>
<td>4257</td>
</tr>
<tr>
<td>4C</td>
<td><em>C. behen</em></td>
<td>Zakho Sere Reza</td>
<td>37.11914062N 42.67837174E</td>
<td>4256</td>
</tr>
<tr>
<td>5C</td>
<td><em>C. rigida</em></td>
<td>Duhok Zawete</td>
<td>36.92932129N 43.14986068E</td>
<td>3945</td>
</tr>
<tr>
<td>6C</td>
<td><em>C. hyalolepis</em></td>
<td>Zakho – Duhok Road</td>
<td>37.1274189N 42.67811940E</td>
<td>3938</td>
</tr>
<tr>
<td>7C</td>
<td><em>C. regia</em></td>
<td>Zakho City Gali</td>
<td>37.17717896N 42.65238483E</td>
<td>3940</td>
</tr>
<tr>
<td>8C</td>
<td><em>C. iberica</em></td>
<td>Zakho City Gali</td>
<td>37.07717896N 42.65238483E</td>
<td>4255</td>
</tr>
<tr>
<td>9C</td>
<td><em>C. gigantea</em></td>
<td>Duhok Bablo</td>
<td>36.87799072N 43.13043568N</td>
<td>3947</td>
</tr>
<tr>
<td>10C</td>
<td><em>C. virga</em></td>
<td>Zakho Sheranesh</td>
<td>37.17327881N 42.74206720E</td>
<td>4258</td>
</tr>
</tbody>
</table>

Figure 1: A map illustrates the distribution of the collected samples of the Centaurea plants species in the Duhok governorate. The map is designed in ArcMap software version 10.7.1.

2.2 DNA extraction from Centaurea taxa:

The DNA samples were isolated from the fresh leaves of 10 Centaurea L. species following the standard cetyl-trimethylammonium bromide (CTAB) method (Weigand et al., 1993) with minor modifications by Hussein and Jubrael (2021).

The DNA quantity was estimated using the nanodrop spectrophotometer.

2.3 Polymerase Chain reaction (PCR) technique:

For the DNA barcoding of Nuclear Ribosomal DNA Internal Transcribed Spacer, ITS and (SCoT) markers amplifications, two different data set of primers were used.
(Table 2 and Table 4). 20μl PCR product was prepared as follows: 2 μl of DNA sample (50 ng), 2 μl (10 pmol) primers, 10 μl of DNA Master mix (Addbio, Korea) and 4 μl of sterile water. The following parameters were set to conduct the PCR amplifications; the thermocycler conditions were adapted for the nuclear DNA barcode as follows: 5 min at 95 °C followed by 35 cycles of 95 °C for 30 s as denaturation temperature, then annealing temperature was set to 30 s and 30 at 55 °C, 72 °C respectively and then 10 min at 72 °C as an extension, whereas different program was set for SCOT marker amplification: the initial denaturation temperature was 94°C for 3 min., then 35 cycles of 1 min at 94°C, 1 min of 50°C as annealing temperature, and the extension was 2 and 5 min at 72°C respectively.

Table 2: shows the oligonucleotides name, sequences and banding size used in current study.

<table>
<thead>
<tr>
<th>oligonucleotides name</th>
<th>forward primer</th>
<th>reverse primer</th>
<th>reference</th>
<th>banding size (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ITS region</td>
<td>5'TCCTCCGCTTATTGATATGC3'</td>
<td>5'TCCGTAGGTGAACCTGCGG3'</td>
<td>White et al., 1990</td>
<td>700</td>
</tr>
</tbody>
</table>

2.4 Capillary Standard Sequencing
The PCR products of the ITS gene region were packed and delivered to the Macrogen Inc company (Seoul, South Korea http://dna.macrogen.com) to sequence the single direction sanger sequencing. The amplified product was purified using the illustra GFX PCR DNA and Gel Band Purification Kit from (GE healthcare) following the manufacturer's instructions. The purified product was deposited in the Macrogen; a public biotechnology company in South Korea for DNA sequencing according to the business's instructions.

2.5 Data scoring and analysis
Each band was represented as a marker. Then, the presence (1) or absence (0) used as a code to score the SCoT marker fragments manually to generate a matrix for the raw data. The data analyzed applying the NTSYS software version 2.02 (Rohlf, 1998). The power of these markers was calculated based on the Polymorphic Information Content (PIC) and Resolving power (Rp) calculations. The PIC value was estimated using bellow formula PIC = 2f (1–f), where f= represent the bands frequency presence; and 1-f= the bands frequency absence (Roldan-Ruiz et al., 2000). On the other hand, the resolving power (Rp) applied to calculate the ability of oligonucleotides dissimilarity detection based on the Prevost and Wilkinson (1999): Rp = ΣIp where Ip (band informativeness) which is taking the following values: 1–[2(0.5–p)], where p is the bands present percentage. The dendrogram was reconstructed using the Dice similarity matrix and the unweighted pair group method with arithmetic mean (UPGMA) clustering procedures.

2.6 Phylogenetic analysis
DNA sequencing data were analyzed using the Geneious prime software (https://www.geneious.com) (Kearse et al. 2012) and Mega 11 software (Tamura et al. 2021). The DNA data was aligned using ClustalW (Thompson et al., 2003). The best model of evolution was estimated, then, Maximum Likelihood (ML) tree was conducted using Mega 11 (Tamura et al. 2021).

2.7 Results and discussion
The preliminary results of testing the total of 15 SCoT primers, only 10 of them were work properly and generate reproducible and clear polymorphic profiles (Figure 2). Consequently, the generated data of these primers, as shown in Table (3), then implemented to estimate the genetic relationships of the ten Centaurea species. The total of 104 amplified bands was scored as polymorphic bands, whereas the polymorphic bands ranged from 5 to 16 with an average of 10.4. The percentage of the polymorphism was 100% in all tested primers. Thus, the SCOT markers were indicated as an efficient and applicable marker on genetic diversity analysis for the Centaurea species. Similar finding has been approved on work of Mirzaei and Salari (2022).

![Figure 2: Amplified DNA fragments of SCoT marker. The codes on the top of pictures refer to the Centaurea accession numbers whereas the bottom codes are primer names (see Table 1)](image-url)
The power of the oligonucleotides was estimated to differentiate the *Centaurea* species using Rp and PIC values. The Rp values were ranged from 3.4 to 12 in (SCoT1 and SCoT53) primers respectively with an average of 5.74. Similarly, the PIC value was ranged from 0.24 to 0.36 with an average of 0.319 (Table 3). The PIC values are considerably variable depending on the genotype variation (Manimekalai and Nagarajan, 2006).

Table 3: shows the genetic similarity matrix of *Centaurea* species.

<table>
<thead>
<tr>
<th>C. solstitialis</th>
<th>C. bruguierana</th>
<th>C. balsamita</th>
<th>C. behen</th>
<th>C. rigida</th>
<th>C. hyalolepis</th>
<th>C. regia</th>
<th>C. iberica</th>
<th>C. gigantea</th>
<th>C. virgata</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.000000</td>
<td>0.6767677</td>
<td>1.000000</td>
<td>0.747474</td>
<td>0.545454</td>
<td>0.737373</td>
<td>1.00000</td>
<td>0.595959</td>
<td>0.515151</td>
<td>0.575757</td>
</tr>
<tr>
<td>0.6767677</td>
<td>0.565656</td>
<td>0.565656</td>
<td>1.000000</td>
<td>0.585858</td>
<td>0.656565</td>
<td>0.636363</td>
<td>0.595959</td>
<td>0.515151</td>
<td>0.575757</td>
</tr>
<tr>
<td>0.747474</td>
<td>0.585858</td>
<td>0.656565</td>
<td>0.636363</td>
<td>1.000000</td>
<td>0.606060</td>
<td>0.636363</td>
<td>0.595959</td>
<td>0.515151</td>
<td>0.575757</td>
</tr>
<tr>
<td>0.545454</td>
<td>0.565656</td>
<td>0.656565</td>
<td>1.000000</td>
<td>0.606060</td>
<td>0.636363</td>
<td>1.00000</td>
<td>0.595959</td>
<td>0.515151</td>
<td>0.575757</td>
</tr>
<tr>
<td>0.737373</td>
<td>0.656565</td>
<td>0.636363</td>
<td>0.595959</td>
<td>0.515151</td>
<td>1.000000</td>
<td>0.636363</td>
<td>1.000000</td>
<td>0.515151</td>
<td>0.575757</td>
</tr>
<tr>
<td>1.000000</td>
<td>0.636363</td>
<td>0.595959</td>
<td>0.515151</td>
<td>0.575757</td>
<td>0.636363</td>
<td>1.00000</td>
<td>0.595959</td>
<td>0.515151</td>
<td>0.575757</td>
</tr>
</tbody>
</table>

The genetic relationships of 10 tested *Centaurea* species based on the Dice similarity coefficient matrix shown in the (Table, 3 and Figure 3). The lowest similarity was between *C. solstitialis* and *C. behen* with (0.52) and the highest value was 0.82 between *C. balsamita* and *C. rigida*. On the other hand, the Dice similarity matrix and the UPGMA methods exhibited the clusters that were identified at the 0.83 similarity level.

Figure 3: The illustrated dendrogram of cluster analysis of ten *Centaurea* species
Table 4: oligonucleotides names, sequences, number of bands, polymorphic bands proportions, polymorphism information contents (PIC), resolving powers (R)) in the Centaurea species.

<table>
<thead>
<tr>
<th>oligonucleotides names</th>
<th>oligonucleotides sequences (5′-3′)</th>
<th>Number of bands</th>
<th>Polymorphic bands</th>
<th>polymorphic bands proportion</th>
<th>PIC</th>
<th>RP</th>
</tr>
</thead>
<tbody>
<tr>
<td>SCOT 1</td>
<td>CAACAATGGCTACCACCA</td>
<td>5</td>
<td>5</td>
<td>100</td>
<td>0.24</td>
<td>3.4</td>
</tr>
<tr>
<td>SCOT 6</td>
<td>CAACAATGGTCACCCACCCG</td>
<td>13</td>
<td>13</td>
<td>100</td>
<td>0.33</td>
<td>6.2</td>
</tr>
<tr>
<td>SCOT 11</td>
<td>AAGCAATGGCTACCACCCA</td>
<td>7</td>
<td>7</td>
<td>100</td>
<td>0.30</td>
<td>3.6</td>
</tr>
<tr>
<td>SCOT 14</td>
<td>ACGACATGGGCCACCACGC</td>
<td>13</td>
<td>13</td>
<td>100</td>
<td>0.34</td>
<td>7.2</td>
</tr>
<tr>
<td>SCOT 15</td>
<td>ACGACATGGCCGACCGGCA</td>
<td>9</td>
<td>9</td>
<td>100</td>
<td>0.36</td>
<td>5.2</td>
</tr>
<tr>
<td>SCOT 32</td>
<td>CCATGGCTACCACCGCAC</td>
<td>8</td>
<td>8</td>
<td>100</td>
<td>0.32</td>
<td>3.6</td>
</tr>
<tr>
<td>SCOT 33</td>
<td>CCATGGCTACCACCGCAC</td>
<td>13</td>
<td>13</td>
<td>100</td>
<td>0.31</td>
<td>5.4</td>
</tr>
<tr>
<td>SCOT 34</td>
<td>ACCATGGCTACCACCGCA</td>
<td>10</td>
<td>10</td>
<td>100</td>
<td>0.32</td>
<td>6.0</td>
</tr>
<tr>
<td>SCOT 47</td>
<td>ACAATGGCTACCACGCT</td>
<td>10</td>
<td>10</td>
<td>100</td>
<td>0.35</td>
<td>4.8</td>
</tr>
<tr>
<td>SCOT 53</td>
<td>ACAATGGCTACCACGCGA</td>
<td>16</td>
<td>16</td>
<td>100</td>
<td>0.32</td>
<td>12.0</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td><strong>104</strong></td>
<td></td>
<td></td>
<td><strong>3.19</strong></td>
<td><strong>57.4</strong></td>
</tr>
<tr>
<td><strong>Mean</strong></td>
<td></td>
<td><strong>10.4</strong></td>
<td></td>
<td></td>
<td><strong>0.319</strong></td>
<td><strong>5.74</strong></td>
</tr>
</tbody>
</table>

Phylogenetic analysis

The Nuclear ITS gene region, as a barcode, shows considerable amplifications with an approximate of 700 bp sequencing length (Figure 4). In plants, barcoding is considerably applied to resolve the evolutionary relationships of closely related species suggesting that this technique regarded as the significant tools for species differentiation (Alvarez & Wendel, 2003; Hollingsworth et al. 2011)

Figure 4: Amplified nuclear DNA fragments of ITS gene region: The codes on the top of pictures refer to the Centaurea accession numbers (see Table 1)

Despite that, the amplified gene length was 700 bp, only 630 bp was sequenced due to a single direction primer sequencing. The 10 Centaurea species were aligned using ClustalW. The best fitting substitution model of molecular evolution was chosen based on the Akaike information criteria (AIC) and Bayesian information criterion (BIC) using Modeltest v3.7 (Posada & Crandall 1998). The Tamura 3-parameter model by Tamura (1992) was selected as best. Consequently, polytomous phylogenetic trees was constructed as follows: clade one; C. solstitialis; clade 2, C. balsamita and C. virgata; clade three subdivide into two subclades: C. iberica, C. hayalolepis, C. brugueriana and C. gigantea, C. regia, C. rigida, C. behen (Figure 5). Although the current taxa are much restricted to Iraq and surrounding area, some of current results has been approved by (Sirin et al., 2022).
Further investigations were carried out to find out the taxonomic position as well as checking the correct names of the studied taxa. Accordingly, 34 taxa including the current 10 investigated data were aligned and phylogenetic trees were reconstructed. Subsequently, the current taxa were positioned all around tree and clustered with either same name taxa or different closely related taxa; *C. bruguieriana*, *C. iberica*, *C. behen*, *C. solstitialis* and *C. balsamita* are clustered with the same names accession from gene bank, whereas *C. regia*, *C. rigida*, and *C. gigantea* were nested and mixed with other closely related species (Figure 6). Suggesting that genetic introgression or hybridization might occur among these taxa, this result significantly agrees with previous findings (Garcea, 1992; Font et al., 2009; Hilopd, 2014).
CONCLUSIONS

The current study clarifies a new insight into the genetic diversity and species identification of some Centaurea species. The combination of the nr DNA region barcode with SCoT marker considers an effective method for reconstructing the evolutionary relationships and species identification of various Centaurea species. The SCoT marker approved that as a powerful tool for genetic variation determination. On the other hand, the internal transcriber spacer region barcoding confirms the marker results and further investigated the taxonomic position of closely related species on a phylogenetic tree. These findings will help to discover and build up the phylogenetic relationships of remaining Centaurea species in Iraq as well as recommend applying rDNA barcoding and SCoT marker for interspecific and intraspecific identification.

REFERENCES


283


