Phylogenetic Analysis of Four Selected Asteraceae Species Using ITS Molecular Markers

Zahraa earif yahyaa*[1], Ashwaq T. Hameed[2] and Rasha Abd Al Hadi Al Tarawneh3

[1,2]Department of Biology, College of Education of Women, University of Anbar, Anbar, Iraq
[3]Faculty of Agricultural Technology. Department of Agricultural Biotechnology and Genetic Engineering, Al-Ahliyya Amman University. Jordan
*Corresponding author's email: Zah24w4003@uoanbar.edu.iq
[2]E-mail ashwaq.talib@uoanbar.edu.iq , [3]E-mail r.tarawneh@ammanu.edu.jo

Abstract

The Asteraceae family is one of the largest and most diverse families of flowering plants worldwide, comprising over 23,000 species distributed across a wide range of geographic environments. Molecular markers have become essential tools in taxonomic and evolutionary studies of plants, particularly in groups where species are difficult to distinguish based solely on morphological characteristics. The Internal Transcribed Spacer (ITS) region is one of the most widely used molecular markers in plant systematics and phylogenetic studies due to its high evolutionary rate and strong discriminatory power among closely related species. This study aimed to assess genetic diversity and infer phylogenetic relationships among four plant species belonging to the Asteraceae family using ITS region sequencing (n = 4). Plant samples were collected from several areas in Anbar Governorate, Iraq. Total genomic DNA was extracted using the CTAB method, and the ITS region was amplified using polymerase chain reaction (PCR) with the primers ITS1F and ITS4R. The amplified products were confirmed on agarose gel, showing clear bands of approximately 600–700 bp, consistent with the expected size of the ITS region in angiosperms. The obtained sequences were compared with reference sequences in the GenBank using BLAST analysis, revealing high sequence similarity (97–100%), which confirmed accurate species identification. Phylogenetic analysis was performed using the Maximum Likelihood method with 1000 bootstrap replicates. The results showed that the studied species were divided into two main clades. The first clade included *Centaurea caballeroi*, *Artemisia transiliensis*, and *Sonchus brachyotus*, where the closest genetic relationship was observed between *C. caballeroi* and *S. brachyotus*, with genetic distance values ranging from 0.183 to 0.193, while *A. transiliensis* showed higher divergence (0.376). In contrast, *Silybum marianum* formed a distinct separate clade with the highest genetic distance, indicating clear genetic differentiation. Overall, the results demonstrate the effectiveness of ITS-based DNA barcoding in resolving phylogenetic relationships within the Asteraceae family and highlight its importance in plant molecular taxonomy and biodiversity studies.

ISSN 2072-3857

Keywords: Phylogenetic, ITS region, Molecular phylogeny, BLAST analysis

Introduction

The Asteraceae family is one of the largest and most diverse flowering plant families in the world, comprising more than 1,600 genera and over 23,000 species distributed across a wide range of climatic environments [1]. This family is characterized by its great morphological and chemical diversity, as well as its high adaptability to various environmental conditions, which has contributed to its widespread distribution in temperate and semi-arid regions [2]. The Asteraceae family includes a large number of plants of economic and medicinal importance, such as *Helianthus annuus* and *Lactuca sativa*, in addition to well-known medicinal plants like *Artemisia annua* and *Silybum marianum*, which contain active compounds of pharmaceutical interest [3]. Traditional plant classification has long relied on the morphological characteristics of plants; however, the morphological similarity between many species can lead to difficulties in accurately determining taxonomic relationships. Therefore, molecular techniques have become an essential tool in studying the evolutionary relationships of plants and re-evaluating their classification [4].

Among the molecular markers used in plant taxonomy are chloroplast genes such as matK and rbcL, as well as nuclear markers such as the Internal Transcribed Spacer (ITS) region within nuclear ribosomal DNA [5]. The ITS region is characterized by conserved regions that facilitate amplification using PCR, along with variable regions that allow for the detection of genetic differences between taxonomically related species [6]. This region has been successfully used in numerous plant evolutionary studies, demonstrating high efficiency in analyzing genetic relationships between species within the Asteraceae family, particularly in genera such as *Centaurea*, *Artemisia*, and *Lactuca* [7].

Furthermore, the use of molecular comparison tools such as BLAST allows for the determination of the degree of similarity between different genetic sequences and the identification of the closest evolutionary relatives [8]. In addition, constructing phylogenetic trees using methods such as Maximum Likelihood or Neighbor-Joining provides a deeper understanding of the evolutionary relationships between different lineages [9]. With advances in Next-Generation Sequencing (NGS) techniques, whole-genome analysis has become a powerful tool for studying the genetic diversity of microorganisms [10]. Despite the ecological and economic importance of Asteraceae plants in many parts of the world, including the Middle

ISSN 2072-3857

East, molecular studies of the evolutionary relationships among species in this family remain relatively limited in some regions. Therefore, this study aimed to analyze genetic diversity and assess the evolutionary relationships among a group of Asteraceae species using ITS region sequencing.

Material and Methods

Plant Sample Collection

Plant samples were collected through field visits to several sites in Anbar Governorate, including the areas of Ramadi, Al-Qaim, Al-Karma, and Haditha, during the spring and winter seasons of 2023 and 2024. Samples were collected during the flowering and fruiting stages to ensure accurate taxonomic identification. After collection, the plant samples were cleaned to remove dirt and impurities using distilled water and then dried in the shade at room temperature. The dried plant samples were then ground into a fine, homogeneous powder and stored in airtight containers pending molecular analysis.

Phylogenetic Analysis

The obtained nucleotide sequences were aligned using the MAFFT multiple sequence alignment program. Phylogenetic analysis was performed using MVSP version 3.22 software. Genetic distances between sequences were calculated using the p-distance model, and the phylogenetic tree was constructed using the UPGMA (Unweighted Pair Group Method with Arithmetic Mean) algorithm.

The reliability of the phylogenetic tree branches was evaluated by bootstrap analysis with 1000 replicates.

Table 1. Geographic Locations and Coordinates of Plant Species Collected from Different Regions of Al-Anbar Governorate

Location (English)	Scientific Name	Latitude	Longitude
Falahat Ramadi	*Sonchus brachyotus*	N 33.34033°	E 43.66631°
Hayy al-Zira‘a, Ramadi	*Silybum marianum*	N 33.4311°	E 43.3414°
Karma	*Centaurea caballeroi*	N 33°23'34.1	E 43°56'07.7
Haditha	*Artemisia transiliensis*	N 34°07'51"	E 42°22'26"

Genomic DNA Extraction

Total DNA was extracted from plant samples using a modified CTAB method. Plant tissues were milled with liquid nitrogen, then a buffered CTAB solution was added, and the mixture was incubated in a water bath at 65°C for 25 minutes. Chloroform was then added to

ISSN 2072-3857

separate the cellular components, and the clear layer containing the DNA was centrifuged.

Electrophoresis

DNA quality and purity were determined using electrophoresis on a 1% agarose gel prepared with TBE buffer solution. The DNA samples were mixed with loading dye and electrophoresed at 100 mA for 45 minutes, after which the bands were visualized under UV light. In addition, the amplified ITS-PCR products were analyzed using 1.5% agarose gel electrophoresis to confirm successful amplification and determine the expected fragment size. The PCR products were then preserved for subsequent purification and sequencing procedures."

PCR

PCR amplification of the ITS region of nuclear ribosomal DNA was performed using the primers ITS1F and ITS4R (prepared by Macrogen) according to Wang et al., as shown in Table 2.

Table 2 shows the nitrogenous base sequences of the primers used in the molecular study.

Primer	Nucleotide Sequence (3′→5′)	Annealing Temperature (°C)
ITS1F	TCCGTAGGTGAACCTGCGG	65.6
ITS4R	TCCTCCGCTTATTGATATGC	55.8

The PCR reaction was carried out in a final volume of 25 µL containing 12.5 µL deionized distilled water (ddH_2O), 2.5 µL of 10× ThermoPol buffer, 2.5 µL $MgCl_2$, 5 µL dNTPs, 1 µL of each primer (ITS1F and ITS4R), 0.2 µL Taq DNA polymerase, and 1–2 µL of template DNA. The reaction mixture was mixed thoroughly and amplified in a thermal cycler according to the thermal program shown in Table 3. The final concentrations of $MgCl_2$ and dNTPs were adjusted according to the manufacturer's instructions.

Table 3. Thermal program for PCR reaction

Step	Number of Cycles	Temperature (°C)	Time
Initial Denaturation	1	95	5 min
Denaturation	30	95	30 sec
Annealing	30	56	45 sec
Extension	30	72	50 sec
Final Extension	1	72	7 min

Purification, Sequencing, and Family Tree Construction

The PCR product was purified to remove proteins, primer residues, and dNTPs, and the DNA was then sequenced. The sequencing data were used to study the evolutionary relationships between the plant samples, and a genetic family tree was constructed using MVSP v.3.22 software.

Results and Discussion

The studied plant species underwent molecular analysis based on amplification of the transcribed internal spacer (*ITS*) region of rDNA. Electrophoresis on a 1.5% agarose gel showed distinct, well-defined bands in all samples, ranging in size from 600 to 700 base pairs, which is the expected size for the ITS region in plants. The presence of a single, distinct band per sample indicates the quality and integrity of the extracted DNA, as well as the efficiency of the PCR reaction and the absence of non-specific amplification. Thus, the electrophoresis analysis confirms the success of the ITS region amplification process, providing a scientific basis for sequencing analyses and the study of genetic relationships among the studied species. The results in Figure 1, based on sequence analysis and comparison of nitrogenous bases with the NCBI database, show genetic variation among the studied species, with high BLAST searches showed high sequence identity (97–100%) and low E-values, confirming the accuracy of the molecular identification of the species. The results indicate that the samples belong to several genera within the Asteraceae family, including *Sonchus, Centaurea*, *Artemisia,* and *Silybum.* To ensure sequence validation and availability for future research, the ITS sequences of the studied plant species were deposited in GenBank following confirmation of sequence identity. The assigned accession numbers were LC920497

for *A. transiliensis* and LC920496 for *S. marianum*. Phylogenetic divergence analysis also revealed genetic differences ranging from 0.183 to 1.248, reflecting significant genetic diversity among the species. These results were clearly reflected in the phylogenetic tree (dendrogram), where the species were grouped into two main clusters. The first cluster comprised species with low genetic distances, while the species with high divergence values formed a relatively separate group, reflecting a clear pattern of genetic relatedness and differentiation among the studied species. These results confirm the efficiency of the ITS region as an effective molecular marker for analyzing phylogenetic relationships within the Asteraceae family and revealing the level of genetic diversity among them.

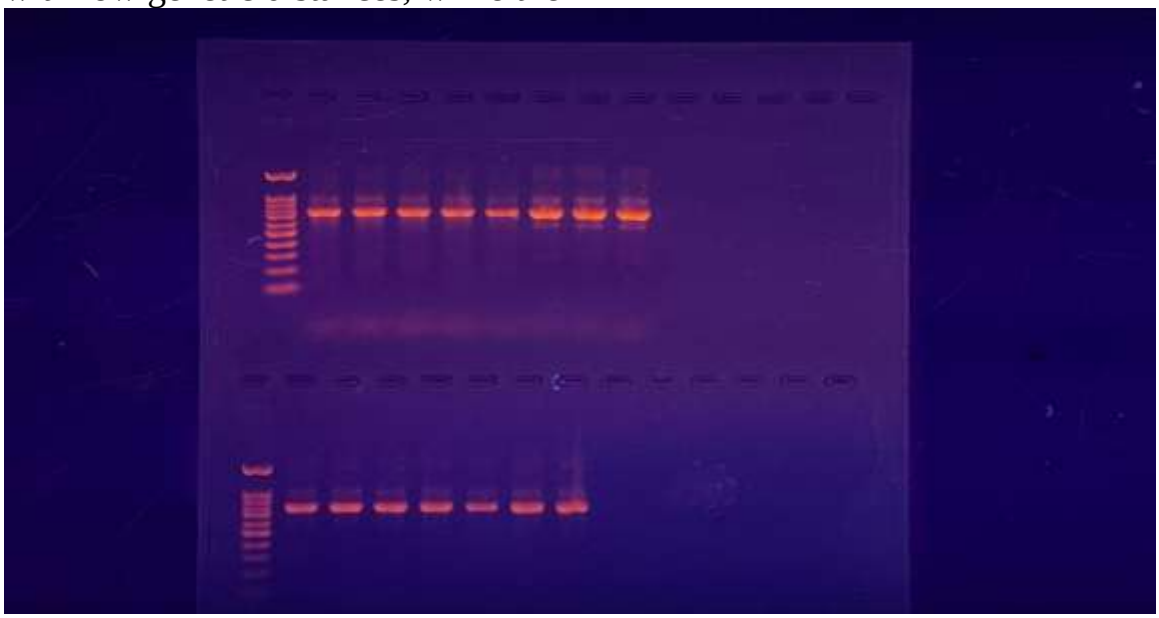

Figure 1. ITS region of nrDNA (Agarose gel 1.5%) extracted from 4 plants of the Asteraceae family

Table 2 results indicated a high identity percentage (85-100%) associated with a low expected value (E value), and this confirms the close genetic relationship to the reference sequence. It also appeared that *Silybum marianum* dominated regions such as Ramadi-Tash, Hit, and Qa'im, while *Sonchus brachyotus* spread in regions like Anah and Haditha. However, Ramadi-Al-Hamdhiyah and Rutba were dominated by *Artemisia transiliensis*. This suggests the adaptability to various ecological zones

Table 5. Results of comparing ITS sequences of the studied species with the NCBI database using BLAST

No.	Species	Closest Match in NCBI	Accession Number	Identity (%)	Query Coverage (%)	E-value
1	*Sonchus brachyotus*	*Sonchus brachyotus*	DQ507909.1	92.0	85	4e-10
7	*Centaurea caballeroi*	*Centaurea solstitialis*	KF721040.1	98.0	98	2e-11
9	*Artemisia transiliensis*	*Artemisia annua*	PV900215.1	99.0	98	0.0
10	*Silybum marianum*	*Silybum marianum*	AY914831.1	97.0	100	0.0

The phylogenetic analysis revealed that the studied plant species were divided into two main clusters based on their genetic distances. The first cluster included *Centaurea caballeroi* (Sample 2), *Artemisia transiliensis* (Sample 3), and *Sonchus brachyotus* (Sample 4), indicating a relatively closer genetic relationship among these species. Within this cluster, *Centaurea caballeroi* and *Sonchus brachyotus* showed the closest relationship, as reflected by the lower genetic distance values (0.183 and 0.193), suggesting a higher level of sequence similarity between them. On the other hand, *Artemisia transiliensis* appeared slightly more distant within the same cluster (0.376), indicating moderate genetic divergence compared to the other two species. The second cluster consisted of *Silybum marianum* (Sample 1), which was clearly separated from the rest of the species, showing the highest genetic distance. This separation reflects a distinct genetic background and suggests that *Silybum marianum* is more evolutionarily divergent compared to the other studied species. Overall, the clustering pattern indicates varying degrees of genetic relatedness among the studied plants, with some species sharing closer evolutionary relationships while others exhibit significant divergence.

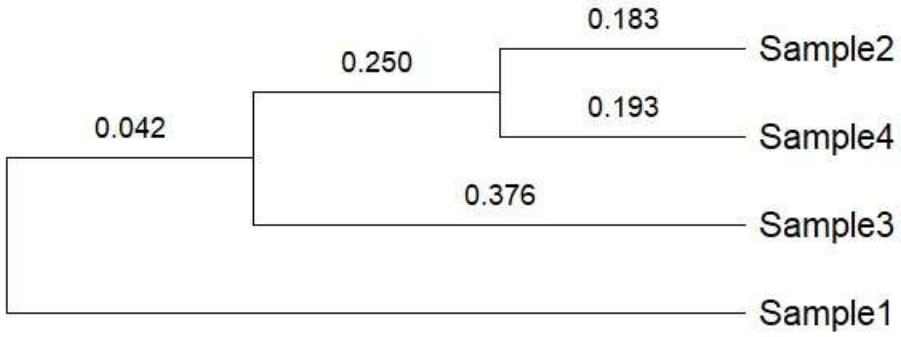

Figure 2. The dendrogram showing clustering of studied plant species based on ITS sequence analysis

The results of the present study demonstrated that the Internal Transcribed Spacer (ITS) region of rDNA is an efficient molecular marker for the identification and phylogenetic analysis of plant species belonging to the family [11]. The amplification of the ITS region produced clear bands ranging from 600–700 bp, which is consistent with the expected fragment size for ITS sequences in most angiosperms. Similar results were reported by Devi *et al*., and Okan *et al*.,[12, 13], who indicated that the ITS region is widely used in plant molecular systematics due to its high variability and suitability for distinguishing closely related species within Asteraceae . The sequence analysis and comparison with the NCBI database confirmed the taxonomic identity of the studied species with similarity values ranging from 97–100%, indicating accurate molecular identification. This high level of similarity supports the reliability of ITS markers in differentiating plant taxa. These findings agree with the study of Mandel *et al*.,[1], who reported that ITS sequencing provides a powerful approach for resolving phylogenetic relationships within the Asteraceae family and for identifying species belonging to different genera such as *Centaurea, Artemisia*, and *Sonchus*. This balance between conserved and variable regions has made this gene a cornerstone of molecular taxonomy for decades. The BLAST analysis results in this study showed that the studied gene sequence has a high similarity to a set of reference sequences in the GenBank database, indicating that the studied sample belongs to a specific evolutionary group within the bacteria studied. BLAST analysis is a crucial step in molecular studies because it allows for the identification of the closest related sequences to the studied sequence and the estimation of their similarity [14]. Furthermore, the use of global genetic databases contributes to improving the accuracy of genetic identification of microorganisms. The ITS-based phylogenetic analysis revealed a clear genetic differentiation among the studied species, grouping them into two main clusters. The first cluster included *Centaurea caballeroi*, *Artemisia transiliensis*, and *Sonchus*

ISSN 2072-3857

brachyotus, indicating a relatively closer genetic relationship among these species despite belonging to different genera. This clustering may reflect shared evolutionary pressures or partial similarity in ecological adaptation. Within this group, the relatively low genetic distance between *Centaurea caballeroi* and *Sonchus brachyotus* suggests a higher degree of sequence similarity, which could be attributed to conserved regions within the ITS marker. In contrast, *Artemisia transiliensis* showed a higher genetic distance within the same cluster, indicating moderate divergence, which may be linked to its distinct ecological adaptation or evolutionary lineage. On the other hand, *Silybum marianum* formed a separate cluster, reflecting a clear genetic divergence from the other studied species. This separation indicates a distinct genetic background, which may be associated with differences in ecological requirements, life history traits, or taxonomic classification. This high divergence reflects significant differences in the ITS region sequences and suggests distant evolutionary relationships between these species despite belonging to the same botanical family. Comparable results were reported by Chen et al.,[6], who noted that the Asteraceae family exhibits high genetic diversity among genera due to its wide evolutionary radiation and adaptation to diverse ecological environments. ITS-based DNA barcoding has proven to be an effective method to highlight the genetic relationships among species belonging to the same genus in various geographical regions [15]. However, the low diversity (97.8%) in Rawa might reflect intraspecific variation, or crossing was common in species such as Chenopodium [16]. The biodiversity observed in the current investigations highlights the importance of maintaining those species as a gene bank for domesticated plants, particularly in regions affected by climate change.

Recent studies indicate that using Maximum Likelihood methods in phylogenetic tree analysis provides higher accuracy compared to some traditional methods, due to their ability to model molecular evolutionary processes more realistically. This method has become widely used in many modern genomic studies, especially with the development of analytical software such as IQ-TREE and MEGA [17]. These findings highlight the usefulness of ITS markers was also demonstrated through DNA diversity analyses which are useful in estimating genetic diversity and evolutionary relationships among herbaceous species of the Asteraceae family distributed in different ecological regions (e.g. Ramadi, Karma, Haditha and Al-Qaim). Further diversification of species genetics can also arise due to ecology, geography and adaptation to different conditions. The importance of environmental diversity on genetic variation among members of the Asteraceae family has been previously highlighted [18]. Overall, the current study confirms that

ISSN 2072-3857

ITS sequence analysis provides a reliable molecular approach for identifying plant species and understanding their evolutionary relationships, particularly within large and taxonomically complex plant families such as Asteraceae.

Conclusion

These results underscore the importance of using molecular bioengineering tools in studying and understanding the genetic relationships between different species, especially in cases where traditional methods are insufficient to distinguish between closely related species. Overall, this study demonstrates that combining gene sequence analysis, comparing sequences using global databases, analyzing gene distances, and constructing phylogenetic trees provides a comprehensive scientific framework for studying evolutionary relationships and accurately determining the taxonomic identity of microorganisms. These findings also open the door to further research that can utilize advanced gene sequencing techniques to analyze genomes more deeply and understand the genetic diversity of microorganisms in different environments.

Acknowledgment

The authors would like to express their sincere gratitude to the Department of Biology, College of Education for Women, University of Anbar, Anbar, Iraq, for their support and facilities provided during the course of this study.

Funding sources

None.

References

[1] Mandel JR, Dikow RB, Siniscalchi CM, Thapa R, Watson LE, Funk VA. 2019. A fully resolved backbone phylogeny reveals numerous dispersals and explosive diversifications throughout the history of Asteraceae. *Proceedings of the National Academy of Sciences*, 116(28):14083–14088.

[2] Xia M, Cai M, Comes HP, Zheng L, Ohi-Toma T, Lee J, et al. 2022. An overlooked dispersal route of Cardueae (Asteraceae) from the Mediterranean to East Asia revealed by phylogenomic and biogeographical analyses of Atractylodes. *Annals of Botany*, 130(1):53–64.

[3] Rolnik A, Olas B. 2021. The plants of the Asteraceae family as agents in the protection of human health.

International Journal of Molecular Sciences, 22(6):3009.

[4] Stull GW, Soltis PS, Soltis DE, Gitzendanner MA, Smith SA. 2020. Nuclear phylogenomic analyses of asterids conflict with plastome trees and support novel relationships among major lineages. *American Journal of Botany*, 107(5):790–805.

[5] Hollingsworth TD, Adams ER, Anderson RM, Atkins K, Bartsch S, Basáñez M-G, et al. 2015. Quantitative analyses and modelling to support achievement of the 2020 goals for nine neglected tropical diseases. *Parasites & Vectors*, 8(1):630.

[6] Chen H, Li T, Chen X, Qu T, Zheng X, Luo J, et al. 2024. Insights into comparative genomics, structural features, and phylogenetic relationship of species from Eurasian Aster and its related genera (Asteraceae: Astereae) based on complete chloroplast genome. *Frontiers in Plant Science*, 15:1367132.

[7] Novaković J, Janaćković P, Susanna A, Lazarević M, Boršić I, Milanovici S, et al. 2022. Molecular insights into the Centaurea calocephala complex (Compositae) from the Balkans—Does phylogeny match systematics? *Diversity*, 14(5):394.

[8] Janda JM, Abbott SL. 2007. 16S rRNA gene sequencing for bacterial identification in the diagnostic laboratory: pluses, perils, and pitfalls. *Journal of Clinical Microbiology*, 45(9):2761–2764.

[9] Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, Von Haeseler A, et al. 2020. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Molecular Biology and Evolution*, 37(5):1530–1534.

[10] Quince C, Walker AW, Simpson JT, Loman NJ, Segata N. 2017. Shotgun metagenomics, from sampling to analysis. *Nature Biotechnology*, 35(9):833–844.

[11] Saif W, Elsherbeny EA, Mohamed MA, Ahmed MZ, ELSarag EI. 2025. The internal transcribed spacer (ITS) as a DNA barcode in identifying some plants from family Amaranthaceae. *Egyptian Journal of Botany*, 65(4):337–345.

[12] Devi ML, Thorat SS, Devi KK, Sharma KC, Singh YD, Mishra A, et al. 2022. Internal Transcribed Spacer (ITS) region of nuclear ribosomal DNA as a suitable DNA Barcode for identification of *Zanthoxylum armatum* DC. from Manipur. *Molecular Biotechnology*, 64(12):1454–1467.

[13] Okan K, Sevindik E, Paksoy MY. 2024. Phylogenetic analysis of the endemic *Bornmuellera* Hausskn. spp. (Brassicaceae) in Türkiye based on nuclear ITS and chloroplast trnL intron, trnLF, rbcL and trnQ-rps16 DNA sequences. *Genetic Resources and Crop Evolution*, 71(4):1529–1539.

ISSN 2072-3857

[14] Moreira GL, Panero JL, Inglis P, Zappi D, Cavalcanti T. 2023. A time-calibrated phylogeny of *Verbesina* (Heliantheae–Asteraceae) based on nuclear ribosomal ITS and ETS sequences. *Edinburgh Journal of Botany*, 80:1–22.

[15] Chen S, Yao H, Han J, Liu C, Song J, Shi L, et al. 2010. Validation of the ITS2 region as a novel DNA barcode for identifying medicinal plant species. *PLOS ONE*, 5(1):e8613.

[16] Mandák B, Krak K, Vít P, Lomonosova MN, Belyayev A, Habibi F, et al. 2018. Hybridization and polyploidization within the *Chenopodium album* aggregate analysed by means of cytological and molecular markers. *Molecular Phylogenetics and Evolution*, 129:189–201.

[17] Gasperskaja E, Kučinskas V. 2017. The most common technologies and tools for functional genome analysis. *Acta Medica Lituanica*, 24(1):1.

[18] Zhang G, Yang J, Zhang C, Jiao B, Panero JL, Cai J, et al. 2024. Nuclear phylogenomics of Asteraceae with increased sampling provides new insights into convergent morphological and molecular evolution. *Plant Communications*, 5(6).

ISSN 2072-3857