DNA BARCODING ANALYSIS OF THE ASTERACEAE SPECIES FROM KAZAKHSTAN

A.S. Akhmetollayeva, A.T. Zhumabek, M.Y. Sutula, D. Tussipkan, S.A.Manabayeva

National Center for Biotechnology, Kazakhstan, 010000, Astana, Qorgalzhyn hwy 13/5
e-mail: manabayeva@biocenter.kz

Asteraceae is the largest family of flowering plants in the world, containing more than 1.900 genera and more than 32.000 individual species. Many members of the *Asteraceae* family are important for medicinal, ornamental, and economic purposes. In Kazakhstan, the family is represented by 146 genera and 883 species.

Many species of *Asteraceae* are currently being used for medicinal purposes in Kazakhstan. The species of this family including *Artemisia dracunculus*, *Pentanema caspicum*, *Achillea nobilis*, *Leuzea altaica*, *Klasea cardunculus,* and their derivatives are effective in the treatment of diseases, possess anti-fatigue, and anti-inflammatory properties. They also have immunomodulatory properties that can to reduce inflammation, speed up wound healing, and boost the immune system in response to bacterial or viral infection. Commercially important plants of the *Asteraceae* family include food crops, aside from consumption, the seeds, roots, etc. can be used to produce cooking oil and spices. Other commercially important species of the *Asteraceae* family are members of the *Tanacetum* genera, which have insecticidal properties.

Given the many valuable members of the *Asteraceae* described above, a method to authenticate an *Asteraceae* species is essential to ensure drug and food safety.

DNA barcoding is a method that uses a short piece of DNA sequence from a standard locus as a tool for species identification.

In our study DNA was extracted from plant material using the CTAB method, followed by the selection of universal barcode primers (ITS*, matK*, *rbcL*) based on established literature. The result of PCR products were purified and subjected to Sanger sequencing. The assembled sequences were compared with existing DNA sequences using BLAST function of the National Center for Biotechnology (NCBI), with outgroup sequences (*Acicarpha tribuloides, Alseuosmia macrophylla*) and were analyzed in MEGA 11. Maximum Likelihood analysis was conducted using 24 different nucleotide substitution models including ITS (K2+G), *matK* (GTR+G), *rbcL* (K2+G). Basic indicators of genetic diversity were evaluated, including nucleotide divergence (Pi), the proportion of conservative (C), polymorphic, and segregating (S) regions. The ITS regions showed the highest divergence (Pi = 0.127), with conservative regions accounting for 68.7% and polymorphic regions varying within 29.6%. The *rbcL* regions showed higher conservatism (Pi = 0.02), with 94.6% conservative regions and only 5.4% variable regions. The *matK* regions exhibited intermediate values (Pi = 0.052), with 87.5% conservative and 12.5% variable regions. The G+C content of the aligned sequences ranged from 34.9% to 55.0%. The effectiveness of the BLAST tool in identifying identical sequences in the NCBI database was limited at the species level for the chloroplast DNA markers *rbcL* and *matK*. However, using the ITS DNA marker, the search within the NCBI database successfully identified only the species *Achillea millefolium* and *Achillea nobilis* with 100% accuracy. Intra-specific discrepancies were observed for other species such as *Psephellus sibiricus* (99.37%), *Scorzonera parviflora* (99.37%), *Klasea cardunculus* (99.79%). All investigated DNA barcodes were tested successfully identified species at the genus level with 100% accuracy.