

COMPARATIVE ANALYSIS AND PHYLOGENETICS OF *CORNUS SANGUINEA* AND *CORNUS SERICEA*

E.R. Nikonorova^{1*}, O.N. Khapilina², T.A. Krol'¹

¹ All-Russian Research Institute of Medicinal and Aromatic Plants (VILAR), Russian Federation, 117216, Moscow, Grina St., 7

² National Center for Biotechnology, Republic of Kazakhstan, 000010, Astana, Qorgalzhyn highway 13/5
e-mail: gatiatulinaer@gmail.com

The Cornaceae family, comprising three genera and over a hundred species, represents a significant group within the order *Cornales*. The genus *Cornus*, known for its ornamental and medicinal properties, includes species whose phytochemical-rich fruits and leaves are used in traditional medicine to treat various ailments, including diabetes and rheumatic disorders. Despite numerous studies, the phylogenetic relationships within this family remain complex and debated. This study aims to clarify these relationships by analyzing the complete chloroplast (cp) genomes of *Cornus sanguinea* L. and *Cornus sericea* L., focusing on their structure and gene content.

Mature leaves of *C. sanguinea* and *C. sericea* were collected at the Botanical Garden of VILAR. Chloroplasts were isolated using high ionic strength solutions, followed by DNA extraction. Libraries preparation was performed using the Illumina DNA Prep (M) Tagmentation kit and sequenced on the MiSeq Illumina platform. Read quality was assessed with FastQC, and adapter trimming and low-quality read filtering were performed using BBduk in Geneious. The plastome assembly was carried out using the GetOrganelle toolkit, and annotations were done with GeSeq, PGA, and CPGA-VAS2 tools, using *C. capitata* as the reference genome, followed by manual review and correction. Simple sequence repeats (SSRs) were identified using the MISA software. For the phylogenetic analysis, plastomes from the Cornaceae, Hydrangeaceae, Nyssaceae, Garryaceae, Curtisiaceae, and Grubbiaceae families, along with the *Arabidopsis thaliana* plastome, were downloaded from NSBI. Multiple sequence alignment was performed using MAFFT,

followed by trimming with TrimAl. The phylogenetic analysis was conducted using the neighbor-joining method with 500 bootstrap replicates and the Tamura-Nei model.

The analysis revealed that both *C. sanguinea* and *C. sericea* had a typical quadripartite structure of the chloroplast genome, with slight variations in the size of the Large Single Copy and Small Single Copy regions compared to other *Cornus* species. The cp genomes were 158,244 bp and 158,663 bp for *C. sericea* and *C. sanguinea*, respectively. Each cp genome contained 131 genes, including 86 protein-coding genes, 37 tRNA genes, and 8 rRNA genes. In this study, differences in the inverted repeats and their junctions with SSC and LSC regions in eight representative *Cornus* species were identified. A total of 53 SSRs were detected for each species, which was higher than in the reference genome *C. capitata*, with a complete absence of tetra-, penta-, and hexa-nucleotide repeats in the sequenced species. Phylogenetic analysis placed *C. sanguinea* and *C. sericea* within a broader clade of the Cornaceae family, reflecting their close genetic relationship with other species in the family, which is consistent with existing studies.

Overall, our study provides new data about the structure and features of the *C. sericea* cp genome and adds the valuable information on cp genome *C. sanguinea*, and supports the hypothesis of conserved genome structures within the family, which could be crucial for future phylogenetic and evolutionary studies.

Funding: The study was carried out in accordance with the State task on the topic FGUU-2024-0001.