

Characterization of two chloroplast genomes in Cornaceae: *Cornus sanguinea* and *Cornus sericea*

Nikonorova E.^{1*}, Khapilina O.², Sereda A.¹, Bondarev S.¹, Krol' T.¹

¹ All-Russian Research Institute of Medicinal and Aromatic Plants (VILAR), Moscow, Russia

² National Center for Biotechnology, Astana, Kazakhstan

*gatiatulinaer@gmail.com

Key words: chloroplast genomes; *Cornus*; phylogenetics; genomic diversity

Motivation and Aim: The Cornaceae family encompassing over a hundred species, represents a significant group in the order *Cornales* [1]. The genus *Cornus*, particularly noted for its ornamental and medicinal value, comprises species whose phytochemical-rich fruits contribute to traditional medicine, treating a range of pathological conditions from diabetes to rheumatic disorders [2]. Despite extensive studies, the phylogenetic relationships within the family remain complex and controversial [3]. This study aims to clarify these relationships by analyzing the complete chloroplast (cp) genomes of *Cornus sanguinea* and *Cornus sericea*, focusing on their structure and gene content.

Methods and Algorithms: Mature leaves of *C. sanguinea* and *C. sericea* were collected in July at the Botanical Garden of VILAR, followed by chloroplast isolation using high ionic strength solutions and DNA extraction. Sequencing libraries were generated using Illumina DNA Prep, (M) Tagmentation kit and sequenced using the MiSeq Illumina platform. Read quality was assessed via fastqc, and adapter trimming and filtering of low-quality reads were performed using BBduk within Geneious. The plastome assembly was performed using the GetOrganelle toolkit and annotations were carried out with GeSeq, PGA, and CPGAVAS2 tools using *C. capitata* (NC_084212.1) as the reference genome, with subsequent manual review and annotation correction [4]. Simple sequence repeats (SSRs) were determined using the MISA software. For the phylogenetic study, plastomes within the Cornaceae, Hydrangeaceae, Nyssaceae, Garryaceae, Curtisiaceae, Grubbiaceae families, and the *Arabidopsis thaliana* plastome were downloaded from NSBI. Multiple sequence alignment was performed using MAFFT followed by trimming using TrimAl. The phylogenetic analysis was carried out using the neighbor-joining method with 500 bootstrap replicates and the Tamura–Nei model.

Results: The analysis revealed that both *C. sanguinea* and *C. sericea* possess a typical quadripartite structure of the chloroplast genome, with a slight variation in the size of the Large Single Copy (LSC) and Small Single Copy (SSC) regions compared to other *Cornus* species. The cp genomes had typical size being 158,244 and 158,663 bp for *C. sericea* and *C. sanguinea*, respectively. The cp genomes of *C. sericea* and *C. sanguinea* were found to contain 131 genes each, including 86 protein-coding genes, 37 tRNA genes, and 8 rRNA genes (Table 1, the number of unique genes is shown in brackets).

Table 1. The basic chloroplast genome characteristics of *C. sericea* and *C. sanguinea*

Species		<i>C. sericea</i>	<i>C. sanguinea</i>
Genome size (bp)		158,244	158,663
GC content (%)		37.9	37.8
SSC	Genome size (bp)	18,711	18,644
	GC content (%)	32.0	31.9
LSC	Genome size (bp)	87,459	87,925
	GC content (%)	36.0	36.0
IRs	Genome size (bp)	26,037	26,047
	GC content (%)	43.1	43.0
CDS		86 (79)	86 (79)
gene		131 (113)	131 (113)
rRNA		8 (4)	8 (4)
IRs		2	2
tRNA		37 (30)	37 (30)

These results in general are align with some of previous studies [5], but slightly differenced from genes number in *C. bretschnideri*, which was higher (132 vs 131) [6], and *C. alba*, where the number of genes and tRNA was 132 and 38, respectively [7]. At the same time, an only 113 unique genes, of which 79 were protein-coding genes, 30 tRNA genes, and 4 rRNA genes were identified in cp genome of *C. sunhangii* [8]. 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 [5, 9, 10]. In present study the differences in the inverted repeats and their junctions with SSC and LSC in 8 representative *Cornus* species were found. A total of 53 SSRs were revealed for each species, which was higher than that in reference genome *C. capitata*, and there was a complete absence of tetra-, penta-, and hexa-nucleotide repeats in the sequenced species. These results were differed from those obtained by Guan et al. [5], where the only 25–31 SSRs were identified in 10 taxa of *Cornus* subg. *Syncarpea*.

Conclusion: This study provided detailed insights into the chloroplast genomes of *C. sanguinea* and *C. sericea*, underscoring the genetic stability and minimal variation across the Cornaceae family. The findings support the hypothesis of conserved genome structures within the family, which could be pivotal for future phylogenetic and evolutionary studies. The detailed characterization of these genomes enhances our understanding of the adaptive strategies of these species, with implications for conservation biology and the use of genetic resources in breeding and cultivation of these species.

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

References

1. Simpson M.G. Diversity and Classification of Flowering Plants: Eudicots. In: Plant Systematics. Elsevier, 2010;275-448. doi 10.1016/B978-0-12-374380-0.50008-7

2. Dinda B., Kyriakopoulos A.M., Dinda S. et al. *Cornus mas* L. (cornelian cherry), an important European and Asian traditional food and medicine: Ethnomedicine, phytochemistry and pharmacology for its commercial utilization in drug industry. *J Ethnopharmacol.* 2016;193:670-690. doi 10.1016/j.jep.2016.09.042

3. Du Z.Y., Jenny Xiang Q.Y., Cheng J. et al. An updated phylogeny, biogeography, and PhyloCode-based classification of Cornaceae based on three sets of genomic data. *Am J Bot.* 2023;110:e16116. doi 10.1002/ajb2.16116
4. Amiryousefi A., Hyvönen J., Poczar P. IRscope: an online program to visualize the junction sites of chloroplast genomes. *Bioinformatics.* 2018;34:3030-3031. doi 10.1093/bioinformatics/bty220
5. Guan B., Wen J., Guo H., Liu Y. Comparative and phylogenetic analyses based on the complete chloroplast genome of *Cornus* subg. *Syncarpea* (Cornaceae) species. *Front Plant Sci.* 2024;15:1306196. doi 10.3389/fpls.2024.1306196
6. Li X., Ma Q., Zhou H., Yang Y., Li H., Wang J. Characterization of the complete chloroplast genome of *Cornus bretschneideri* (cornaceae). *Mitochondrial DNA B Resour.* 2020;5(1):543-544. doi 10.1080/23802359.2019.1710281
7. Yuan W., He S., Zhang S., Chang D., He Y. The complete chloroplast genome sequence of *Cornus Alba* L. (Cornaceae). *Mitochondrial DNA B Resour.* 2021;6(7):1997-1998. doi 10.1080/23802359.2021.1938727
8. Lv Z.-Y., Huang X.-H., Luo J., Zhang X., Deng T., Li Z.-M. The complete chloroplast genome sequence of *Cornus sunhangii* (Cornaceae). *Mitochondrial DNA B Resour.* 2019;4(2):3242-3243. doi 10.1080/23802359.2019.1669090
9. Keir K.R., Bemmels J.B., Aitken S.N. Low genetic diversity, moderate local adaptation, and phylogeographic insights in *Cornus nuttallii* (Cornaceae). *Am J Bot.* 2011;98:1327-1336. doi 10.3732/ajb.1000466
10. Fu C.N., Li H.T., Milne R. et al. Comparative analyses of plastid genomes from fourteen Cornales species: inferences for phylogenetic relationships and genome evolution. *BMC Genomics.* 2017;18(1):956. doi 10.1186/s12864-017-4319-9