The complete chloroplast genome sequences and phylogenetics of Cornus sanguinea L. and Cornus sericea L. (Cornaceae)

Eugenia Nikonorova

gatiatulinaer@gmail.com

All-Russian Research Institute of Medicinal and Aromatic Plants (VILAR)

Alexandr Shevtsov

National Center for Biotechnology

Nailya Tursunbay

National Center for Biotechnology

Oxana Khapilina

National Center for Biotechnology

Dmitry Baleev

All-Russian Research Institute of Medicinal and Aromatic Plants (VILAR)

Research Article

Keywords: chloroplast genome, plastids, Cornaceae, Cornus, dogwood

Posted Date: February 10th, 2025

DOI: https://doi.org/10.21203/rs.3.rs-5965562/v1

Additional Declarations: No competing interests reported.

Abstract

This study provides an in-depth analysis of the chloroplast genomes of two *Cornus* species, *Cornus sanguinea* L. and *Cornus sericea* L., which are significant both in ornamental horticulture and traditional medicine. These species were collected from the Botanical Garden of the VILAR, providing a unique geographic context for genetic examination. Our results indicated that the plastomes of both species have typical quadripartite structure of chloroplast DNA, with slight variations in the size of the Large Single Copy (LSC) and Small Single Copy (SSC) regions compared to other *Cornus* species. The complete chloroplast genome size of *C. sericea* and *C. sanguinea* was 158 244 and 158 663 bp, respectively. A total of 131 genes, including 86 protein-coding genes, 37 tRNA genes, and 8 rRNA genes were found. The study highlighted the role of simple sequence repeats (SSRs) in genomic differentiation, with a notable absence of tetra-, penta-, and hexa-nucleotide repeats in the studied genomes. This aspect of the genome could be vital for understanding species differentiation and evolution within the genus. Phylogenetic analyses placed *C. sanguinea* and *C. sericea* within a broader clade of Cornaceae and reflected their close relationship to other species in the Cornaceae family. 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*, that is necessary for further studies.

Introduction

The family Cornaceae includes 3 genera and over a hundred species, which, along with such families as Curtisiaceae, Grubbiaceae, Hydrangeaceae, Hydrostachyaceae, Loasaceae, and Nyssaceae, are the part of *Cornales* order (Thomas et al. 2021). The majority of species within the family belong to the genus *Cornus*. The dogwoods (*Cornus* spp.) are not only prized for their ornamental beauty but play a significant role in nature, biology, and medicine (Kazimierski et al. 2019; Badoni et al. 2024). Tenuta et al. (2022) in the comprehensive review highlighted that the edible fruits of European and Asian *Cornus* species are a rich source of phytochemicals with nutritional and functional properties. For example, *C. mas* (cornelian cherry) fruits and leaves are used in traditional medicine for treatment of diabetes, obesity, atherosclerosis, skin diseases, gastrointestinal and rheumatic disorders (Dinda et al. 2016). It was shown that *C. officinalis* possess the high range of pharmacological properties, such as hepatic and renal protection, antitumor, neuroprotective, antidiabetic effect, anti-inflammatory, antioxidant and immunoregulatory activities, etc (Huang et al. 2018).

The conserved nature of the chloroplast (cp) genome, along with its well-defined structure and gene content has provided valuable insights into the genetic diversity and evolutionary relationships within the Cornaceae family (Daniell et al. 2016; Fu et al. 2017; Guan et al. 2024). The first study of chloroplast DNA (cpDNA) variability by PCR-RFLP analysis of the *C. sanguinea* was performed by H. Liesebach and B. Götz (2008), who reported the presence of eight different haplotypes across different European populations without association with geographic occurrence and low level of variation (Liesebach and Götz 2008). Lv et al. (2019) reported the complete cp genome sequences of *C. sunhangii*, giving a foundation for further studies on these species (Lv et al. 2019). Li et al. (Li et al. 2020) also contributed

to this field by characterizing the cp genome of *C. bretschneideri*, respectively, further expanding our understanding of the genetic features of these plants. However, despite the recognized importance of the Cornaceae, the phylogenetic relationships within this family have been controversial and complex (Xiang et al. 1996; Fu et al. 2017; Du et al. 2023; Guan et al. 2024). The classification of species within the *Cornus* genus has seen diverse interpretations, with the taxonomy undergoing multiple revisions as new evidence comes to light (Du et al. 2023). In other species it's been found that cp genomes can vary depending on the species and their geographic distribution, which is related to adaptation to local climatic conditions (Xiong et al. 2023). Keir et al. (2011) revealed the relatively low diversity in cpDNA haplotypes of *C. nuttallii*, collected from 20 native populations of southern California and northern Idaho, USA (Keir et al. 2011). In study of Guan et al. (2024) plastid data of species growing from China to the USA were studied (Guan et al. 2024) – however, they did not include any species native to Russia.

In the VILAR Botanical Garden, two species of dogwood (*C. sanguinea* and *C. sericea*) are cultivated and included in the collection. These species are thoroughly studied, including their genetic resources. However, there is no data on the structure and characteristics of the cp genome of *C. sericea* (no verified sequence was available in the NCBI database at the time of writing), and the studies on the chloroplast genome of *C. sanguinea* are limited (only one representative - FCN1796 from China - has been identified). Therefore, the aim of this study was to characterize and compare the structure and gene organization of the plastid genomes of *C. sanguinea* and *C. sericea* from the medicinal plant collection of the VILAR Botanical Garden.

Materials and Methods

Plant material.

Fresh mature leaves of *Cornus sanguinea* L. (voucher specimen VF 004.21) and *Cornus sericea* L. (voucher specimen VF 006.21) were collected in July at the Botanical Garden of the VILAR (55°33'51.8"N; 37°35'56.8"E), located in urban zone of Moscow with warm-summer humid continental climate (Fig. 1).

Chloroplasts isolation.

Chloroplasts were isolated using high ionic strength solutions based on a technique described by Bookjans et al. (Bookjans et al. 1984). Briefly, after incubation in the dark at 4°C, fresh leaves were homogenized in a Waring blender in 50 mM Tris, 25 mM EDTA, 10 mM mercaptoethanol, 0.1% BSA and 1.25 M NaCl (pH = 8.0). The homogenate was filtered and centrifuged for 5 minutes at 1500 *g* (Eppendorf, Germany), the pellet was resuspended in the homogenization buffer, and the centrifugation was repeated for 5 minutes at 1500 *g*. The resulting chloroplast pellet was resuspended in a 50 mM Tris, 10 mM EDTA (pH = 8.0) and used for DNA extraction.

DNA extraction.

CpDNA was extracted following the method described by Shi et al. (Shi et al. 2012). The analysis of the extracted DNA fragments was performed by DNA separation in a 0.8% agarose gel containing ethidium bromide. Results were documented using a Gel Doc system (Bio-Rad, USA) with Quantity One software (Bio-Rad, USA). DNA concentrations were measured fluorometrically with a Qubit 2.0 Fluorometer (Thermo Fisher Scientific, MA, USA). DNA library preparation was performed using the Illumina DNA Prep, (M) Tagmentation kit (Cat. No. 20060060, Illumina, USA), following the manufacturer's guidelines. Libraries were normalized and then pooled to receive a 14 pM concentration, including a 1% PhiX control. Whole-genome sequencing was performed on the MiSeq platform (Illumina, USA) with the MiSeq Reagent Kit v3, 600 Cycles (#MS-102-3003).

Plastome annotation.

Read quality was assessed via fastqc, and adapter trimming and filtering of low-quality reads were performed using BBduk within Geneious (Kearse et al. 2012). The GetOrganelle toolkit (Jin et al. 2020) was used for *de novo* plastome assembly. Specifically for the plastome of *C. sanguinea*, graph correction using Bandage (Wick et al. 2015) and reassembly were executed using the scripts join_spades_fastg_by_blast.py and get_organelle_from_assembly.py to receive an organelle-sufficient graph as described at GetOrganelle flowchart (Jin et al. 2020). Plastome annotations were carried out with GeSeq (Tillich et al. 2017), PGA (Qu et al. 2019), and CPGAVAS2 (Shi et al. 2019), using *C. capitata* (NC_084212.1) as the reference genome, with subsequent manual review and annotation corrections in Geneious. Simple sequence repeats (SSRs) were determined using the MISA software v2.1, 2020-08-25 (Beier et al. 2017), with set parameters for unit size and the minimum number of repeats: (1/10) (2/6) (3/5) (4/5) (5/5) (6/5) (Beier et al. 2017).

Phylogenetic analysis.

For the phylogenetic study, 60 plastomes from related species within the Cornaceae, Hydrangeaceae, Nyssaceae, Garryaceae, Curtisiaceae, Grubbiaceae families, and the *Arabidopsis thaliana* plastome used as an outgroup were downloaded from NSBI. The list of accession numbers and species used for phylogenetic analysis is available at Supplementary Table 1. Multiple sequence alignment was performed using MAFFT (Katoh 2002), implemented in Geneious platform, followed by trimming using TrimAl (Capella-Gutiérrez et al. 2009). The phylogenetic analysis was carried out using the neighbor-joining method with 500 bootstrap replicates and the Tamura-Nei model in Geneious (Kearse et al. 2012).

Data visualization.

Visualization of plastomes was done using the R programming language ((R Core Team 2018), version 4.2.1) and package "chloroplot" (Zheng et al. 2020). The visual representation of Inverted Repeats (IRs), Small Single Copy (SSC), Large Single Copy (LSC), and the genes located on them was performed using the online version of the IRscope tool (Amiryousefi et al. 2018). The IRScope webserver allows to processes ten plastomes at a time only, so we analyzed several representative *Cornus* species to

compare their IR junctions. Phylogenetic tree was visualized using the R programming language ((R Core Team 2018), version 4.3.2) and package "ggtree".

Results

Characteristics of the C. sericea and C. sanguinea plastomes

Initially, we assembled and characterized the cp genome sequence of *C. sericea* and *C. sanguinea*. As shown in Table 1, the plastomes of the studied species have typical quadripartite structures and fell within the typical size parameters (130,000–160,000 bp), although they slightly exceeded the reference. The number of unique genes, rRNA, and tRNA matched the reference. The GC content showed minimal variation, whereas the lengths of LSC and SSC regions exceeded the reference by 130 to 619 bp; meanwhile, the size of IRs was slightly reduced. Notably, an additional gene and CDS were detected in the plastomes of *C. sericea* and *C. sanguinea* (see below).

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

Species		***C. sericea***	***C. sanguinea***	***C. capitata* (reference)**
Genome size (bp)		158 244	158 663	157 199
GC content (%)		37.9	37.8	38.2
SSC	Genome size (bp)	18 711	18 644	18 412
	GC content (%)	32.0	31.9	32.4
LSC	Genome size (bp)	87 459	87 925	86 306
	GC content (%)	36.0	36.0	36.4
IRs	Genome size (bp)	26 037	26 047	26 112
	GC content (%)	43.1	43.0	43.1
CDS		86 (79)	86 (79)	85 (79)
gene		131 (113)	131 (113)	130 (113)
rRNA		8 (4)	8 (4)	8 (4)
IRs		2	2	2
tRNA		37 (30)	37 (30)	37 (30)
Note: the number of unique genes is shown in brackets				

The studied cp genomes had a typical structure and only slightly differenced from the reference. Therefore, on the next step we have performed an annotation of the resulted assemblies.

Annotation of the C. sanguinea and C. sericea plastomes

The preliminary result of the annotations using three different packages, GeSeq, PGA, and CPGAVAS2, and comparison with the existing reference plastome showed the multiple genes with inaccurate coordinates of gene start and end, differences in coordinates of introns and exons, errors in the annotation of some tRNAs and genes with short exons, etc., were identified. Typically, these inaccuracies involved genes with short exons, such as rps16, pafI (ycf3), rpoC1, and rps12 (a trans-splicing gene composed of 3 exons from two pre-mRNAs with 1 exon in LSC and pairs of the 2nd and 3rd exons in IRs). Errors were also found in tRNAs with introns or those that were unusually short (under 100 nucleotides), including trnT-CGU or trnG-UCC, trnM-CAU or trnfM-CAU, trnM-CAU or trnI-CAU, trnE-UUC or trnI-GAU tRNAs. Such types of misannotation have been described previously (Qu et al. 2023). The list of annotated genes is presented in Table 2.

Table 2
List of genes in annotated genomes of *C. sanguinea* and *C. sericea*

Group	*C. sanguinea* and *C. sericea*
tRNA	tRNA-Ala: trnA-UGC (2)
	tRNA-Arg: trnR-ACG (2), trnR-UCU
	tRNA-Asn: trnN-GUU (2)
	tRNA-Asp: trnD-GUC
	tRNA-Cys: trnC-GCA
	tRNA-Gln: trnQ-UUG
	tRNA-Glu: trnE-UUC
	tRNA-Gly: trnG-GCC, trnG-UCC*
	tRNA-His: trnH-GUG
	tRNA-Ile: trnI-CAU (2), trnI-GAU (2)*
	tRNA-Leu: trnL-CAA (2), trnL-UAA*, trnL-UAG
	tRNA-Lys: trnK-UUU*
	tRNA-Met: trnfM-CAU, trnM-CAU
	tRNA-Phe: trnF-GAA
	tRNA-Pro: trnP-UGG
	tRNA-Ser: trnS-GCU, trnS-GGA, trnS-UGA
	tRNA-Thr: trnT-GGU, trnT-UGU
	tRNA-Trp: trnW-CCA
	tRNA-Tyr: trnY-GUA
	tRNA-Val: trnV-GAC (2), trnV-UAC*
rRNA	rrn4.5 rRNA (2), rrn5 rRNA (2), rrn16 rRNA (2), rrn23 rRNA (2)
NADH-dehydrogenase	ndhA*, ndhB (2)*, ndhC, ndhD, ndhE, ndhF, ndhG, ndhH, ndhI, ndhJ, ndhK
Photosystem I	psaA, psaB, psaC, psaI, psaJ, pafI (ycf3)**, pafII (ycf4),
	pbf1
Photosystem II	psaA, psaB, psaC, psbD, psbH, psbI, psbJ, psbK, psbL, psbM, psbT, psbZ
Large subunit of the ribosome (LSU)	rpl14, rpl16*, rpl2 (2)*, rpl20, rpl22, rpl23 (2), rpl32, rpl33, rpl36

Group	*C. sanguinea* and *C. sericea*
Small subunit of the ribosome (SSU)	rps11, rps12 (2)**, rps14, rps15, rps16, rps18, rps19, rps2, rps3, rps4, rps7 (2), rps8
RNA-polymerase	rpoA, rpoB, rpoC1*, rpoC2
Cytochromes	ccsA, petA, petB*, petD*, petG, petL, petN, psbD, psbE
ATP-synthase	atpA, atpB, atpE, atpF*, atpH, atpI
Other	rbcL, cemA, clpP1**, infA, matK, accD, ycf1 (2), ycf2 (2)
* or ** – one or two introns (2) – two gene copies	

After annotation the visualization of the plastomes was performed. Complete genomic maps for *C. sanguinea, C. sericea*, and the reference genome of *C. capitata* are provided in Fig. 2. The duplicating genes located entirely or partially on the IRs encompass seven protein-encoding genes: ndhB, rpl2, rpl23, rps7, rps12, ycf2, and ycf1; seven tRNAs: trnI-GAU, trnA-UGC, trnL-CAA, trnI-CAU tRNA, trnR-ACG tRNA, trnV-GAC tRNA, trnN-GUU tRNA; along with all rRNAs – rrn4.5, rrn5, rrn16, rrn23. There were nine genes with a single intron – ndhA, ndhB, rpl16, rpl2, rps16, rpoC1, petB, petD, and atpF; and three genes featuring two introns – pafI (ycf3), rps12, and clpP1. Additionally, five tRNA genes include introns – trnG-UCC, trnI-GAU, trnL-UAA, trnK-UUU, trnV-UAC.

IR Contraction and Expansion

To evaluate the specific features of IRs, which, according to the literature data (Goulding et al. 1996; Zhu et al. 2016), may help in understanding the mechanisms of genome stability, and in identifying species-specific adaptations and phylogenetic relationships, we performed an analysis of the inverted repeats and their junctions with SSC and LSC in 8 representative *Cornus* species, revealing differences between the genomes studied (Fig. 3). The longest LSC was observed in *C. sanguinea*, whereas the shortest was in *C. multivernosa*. The length of SSC also differed between the studied species: the longest SSC was found in *C. bretschneideri* and the shortest in *C. kousa*. The rps19 gene was almost entirely located on the LSC, with its first 6 (*C. sericea*) and 46 (*C. sanguinea*) nucleotides on the IRB (LSC/IRb junction). The ycf1 gene was partially located on the SSC and IRa and was 5469 bp in both species. It starts 977 bp within IRa and extends 4,462 bp into the SSC in *C. sericea*, closely resembling the junction in *C. bretschneideri* and *C. wilsoniana*. When comparing IRs in some species shown in Fig. 3, it is evident that this gene was partially on the SSC (4331–4498 bp) and on the IRa (977–1114 bp) in the all studied species. Copies of the ycf1 gene 1044 bp in length in *C. sericea* and 1059 bp in *C. sanguinea* were also found, located at the SSC/IRb junction. In the genome of *C. capitata*, which was used as a reference, the ycf1 gene was annotated only once, as well as in *C. bretschneideri*.

Therefore, despite IRScope only generates visual data it has enabled a detailed evaluation of the structure of IRs and their junctions with SSC and LSC in the studied *Cornus* species.

The search for SSRs

SSRs or microsatellites are short tandemly repeated DNA motifs from 1 to 6 bp, and important sources of adaptive genetic variation (Kashi and King 2006). In the genomic analysis of *C. sanguinea* and *C. sericea*, a total of 53 SSRs were revealed for each species (Table 3). Notably, there was a complete lack of tetra-, penta-, and hexa-nucleotide repeats in the studied samples. When compare with reference *C. capitata*, the less frequency of SSRs was found. At the same time, in more distinct *A. thaliana* the other pattern of SSR was found, confirming the significance of SSRs search in species differentiation.

Table 3
Classification and frequency of SSRs within the chloroplast genomes of *C. sanguinea* and *C. sericea, C. capitata* and *A. thaliana*

Type	Repeat	*C. sanguinea*	*C. sericea*	*C. capitata*	*A. thaliana*
Mono-	A	22	22	8	30
	G	-	-	-	1
	T	29	29	10	38
Di-	AT	-	-	-	4
	TA	1	1	1	2
Tri-	TTA	1	1	-	1
	AAT	-	-	-	1

Phylogenetic study

To elucidate the evolutionary relationships between species belonging to the Cornaceae, Hydrangeaceae, Nyssaceae, Garryaceae, Curtisiaceae, Grubbiaceae families, and *Arabidopsis thaliana*, which served as the outgroup, a phylogenetic tree was constructed based on their complete plastid genomes (Fig. 4). The sequenced plastomes of *C. sericea* and *C. sanguinea* appear within a cluster of *Cornus* species. This cluster was a part of a larger clade including *Alangium* species, suggesting that these genera share a common ancestor which is in line with the existing literature data (Du et al. 2023; Guan et al. 2024). The studied *Cornus* species formed four distinct clades, as it is described earlier: 1) the blue, white, or black-fruited dogwoods (BW); 2) the cornelian cherries (CC); 3) the big-bracted dogwoods (BB); and 4) the dwarf dogwoods (DW) (Du et al. 2023). As it is seen from the results, the *C. sanguinea* sample was placed together with available in NCBI sample of *C. sanguinea* (MN380667.1) from China. The newly sequenced *C. sericea* plastome was placed together with *C. sanguinea* samples,

C. bretschneideri, C. macrophylla, C. alba, and *C. walteri*, forming the sister branches. Both *C. sericea* and *C. sanguinea* plastomes were within BW clade.

Pairwise alignment of *C. sanguinea* and *C. sericea* specimens showed an impressive 99.4% identity. However, this identity dropped to 87.8% when the reference sequence was added.

Discussion

In present study, the complete chloroplast genomes of *C. sanguinea* and *C. sericea* have been characterized, providing additional genetic information for these plants. The chloroplast genomes of *Cornus* species have been extensively studied due to their importance in understanding phylogenetic relationships and evolutionary adaptations within the genus. The chloroplast genome size for both *C. sericea* and *C. sanguinea* reported in our study shows slight deviations from the reference species *C. capitata*, with minor differences in the SSC and LSC regions. The cp genomes had typical size being 158 244 and 158 663 bp for *C. sericea* and *C. sanguinea*, respectively. For example, the cp genome size of *C. elliptica* reported by Lu et al. (2021) was 157,400 bp, while Li et al. (2020) described the 158,270 bp genome of *C. bretschneideri*. Lv et al. (2019) provided details on the 157,446 bp cp genome of *C. sunhangii*, and Yuan et al. (2021) reported the 158,451 bp genome of *C. alba*.

A total of 131 genes were found, including 86 protein-coding genes, 8 rRNA genes, and 37 tRNA genes. These results are align with some of previous studies (Lv et al. 2019; Guan et al. 2024). However, Li et al. (2020) characterized the complete chloroplast genome of *C. bretschneideri*, where the genome organization and gene content were comparable to the findings of our study with *C. sericea* and *C. sanguinea*, except for protein-coding genes, which number was a bit higher (132 vs 131) (Li et al. 2020). Additionally, at the study of Yuan et al. (2021) on *C. alba* the similar structural features, such as the size of LSC and SSC regions were found, whereas the number of protein-coding genes and tRNA was 132 and 38, respectively (Yuan et al. 2021).

The variation in the length of IR, LSC, and SSC regions observed across different *Cornus* species in our study provides insight into the plastome structural dynamics and their evolutionary implications (Goulding et al. 1996; Kashi and King 2006; Zhu et al. 2016). The contraction and expansion at the IR boundaries, particularly affecting the rps19 and ycf1 genes, are phenomena widely reported in the literature and are thought to contribute to genomic plasticity and adaptation. In early study of Goulding et al. (1996), the authors studied the movement of the IR of chloroplast DNA in flowering plants and noted the frequent small expansion and contraction near the rps19 gene even closely related flowering species (Goulding et al. 1996). In cucumbers, comparative chloroplast genome analyses revealed significant variations in the LSC and SSC regions, especially in Indian ecotype, which may correlate with the plant's adaptation to temperature variations (Xia et al. 2023). This study provides insights into how structural changes in chloroplast genomes can influence plant responses to environmental stresses. In study Zhu et al. (2016), the authors showed that genes moved from the SC into the IR exhibited lower

synonymous substitution rates comparable to other IR genes and vice versa – genes moved from the IR into the SC had higher rates of substitution like other SC genes (Zhu et al. 2016).

The number of SSRs was higher than that in reference genome *C. capitata*, and there was a complete lack of tetra-, penta-, and hexa-nucleotide repeats in the studied samples. The results obtained differ from the study of Guan et al. (2024), where the only 25–31 SSRs were identified in 10 taxa of *Cornus* subg. *Syncarpea*; at the same time, the mononucleotide repeats were more frequent in those cp genomes, which is consistent with the observed data (Guan et al. 2024).

The phylogenetic placement of *C. sericea* and *C. sanguinea* within the *Cornus* clade and their close relationship to other species in the Cornaceae family supports the data of a shared evolutionary ancestry (Zhang et al. 2020). The results of the study are consistent with the available data indicating the presence of four distinct clades within *Cornus* genus (Du et al. 2023). In present study, both *C. sericea* and *C. sanguinea* plastomes were within BW clade. *C. sanguinea* cp genome was placed together with available in NCBI database sample of *C. sanguinea* (MN380667.1, isolate FCN1796) from China, indicating a high similarity between these cpDNA. According to DEC analysis performed by Du et al. (2022), the “BW clade likely had an ancestor in eastern Asia, which diversified and spread into North America and Europe”. However, at the same study the authors showed that the plastid phylogeny without nuclear data cannot provide a comprehensive analysis of the biogeography and evolution, resulting in species nesting discordance (Du et al. 2023). Also, the high genomic identity between these two species compared to the reference sequence highlights relatively low divergence within the *Cornus* genus and consistent with existing studies (Keir et al. 2011; Guan et al. 2024).

Conclusion

In summary, the cpDNA of two *Cornus* species, *C. sericea* and *C. sanguinea*, while conserved in overall structure, does display slight variations. The study has provided new insights into the structure and features of the *C. sericea* cp genome, clarified the available data about *C. sanguinea* cpDNA and their classification within the *Cornus* clade. The comparative plastome analysis of the species growing in the Botanical Garden of VILAR provides data for further studies on the possible impact of the biogeography on cpDNA variation and resolving the phylogenetic positions of these species within the Cornaceae family.

Declarations

Conflicts of Interest:

The authors declare no conflicts of interest

Funding:

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

Author Contribution

EN designed the research, collected samples, analyzed the data, wrote the draft; AS and NT performed DNA isolation and sequencing; OK reviewed and commented on the manuscript; DB participated in the data analysis and edited the final version of the manuscript.

Acknowledgements:

The authors express their gratitude to the staff at the Botanical Garden of the VILAR who took care of the plants and made it possible to obtain samples for the research.

Data Archiving Statement Sequence data that support the findings of this study have been deposited in GenBank (accession numbers PP811661 and PP818646) and will be released to the public database when they appear in print.

Data Availability

Statement Sequence data that support the findings of this study have been deposited in GenBank (accession numbers PP811661 and PP818646) and will be released to the public database when they appear in print.

References

1. Amiryousefi A, Hyvönen J, Poczai P (2018) IRscope: an online program to visualize the junction sites of chloroplast genomes. Bioinformatics 34:3030–3031. https://doi.org/10.1093/bioinformatics/bty220
2. Badoni S, Rawat D, Mahato AK, Jangwan NS, Ashraf GM, Alexiou A, Tayeb HO, Alghamdi BS, Papadakis M, Singh MF (2024) Therapeutic Potential of Cornus Genus: Navigating Phytochemistry, Pharmacology, Clinical Studies, and Advanced Delivery Approaches. Chemistry & Biodiversity e202301888. https://doi.org/10.1002/cbdv.202301888
3. Beier S, Thiel T, Münch T, Scholz U, Mascher M (2017) MISA-web: a web server for microsatellite prediction. Bioinformatics 33:2583–2585. https://doi.org/10.1093/bioinformatics/btx198
4. Bookjans G, Stummann BM, Henningsen KW (1984) Preparation of chloroplast DNA from pea plastids isolated in a medium of high ionic strength. Analytical Biochemistry 141:244–247. https://doi.org/10.1016/0003-2697(84)90452-4
5. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T (2009) trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25:1972–1973.

https://doi.org/10.1093/bioinformatics/btp348

6. Daniell H, Lin C-S, Yu M, Chang W-J (2016) Chloroplast genomes: diversity, evolution, and applications in genetic engineering. Genome Biol 17:134. https://doi.org/10.1186/s13059-016-1004-2
7. Dinda B, Kyriakopoulos AM, Dinda S, Zoumpourlis V, Thomaidis NS, Velegraki A, Markopoulos C, Dinda M (2016) 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. Journal of Ethnopharmacology 193:670–690. https://doi.org/10.1016/j.jep.2016.09.042
8. Du Z, (Jenny) Xiang Q, Cheng J, Zhou W, Wang Q, Soltis DE, Soltis PS (2023) An updated phylogeny, biogeography, and PhyloCode-based classification of Cornaceae based on three sets of genomic data. American J of Botany 110:e16116. https://doi.org/10.1002/ajb2.16116
9. Fu C-N, Li H-T, Milne R, Zhang T, Ma P-F, Yang J, Li D-Z, Gao L-M (2017) Comparative analyses of plastid genomes from fourteen Cornales species: inferences for phylogenetic relationships and genome evolution. BMC Genomics 18:956. https://doi.org/10.1186/s12864-017-4319-9
10. Goulding SE, Wolfe KH, Olmstead RG, Morden CW (1996) Ebb and flow of the chloroplast inverted repeat. Molec Gen Genet 252:195–206. https://doi.org/10.1007/BF02173220
11. Guan B, Wen J, Guo H, Liu Y (2024) Comparative and phylogenetic analyses based on the complete chloroplast genome of Cornus subg. Syncarpea (Cornaceae) species. Front Plant Sci 15:1306196. https://doi.org/10.3389/fpls.2024.1306196
12. Huang J, Zhang Y, Dong L, Gao Q, Yin L, Quan H, Chen R, Fu X, Lin D (2018) Ethnopharmacology, phytochemistry, and pharmacology of Cornus officinalis Sieb. et Zucc. Journal of Ethnopharmacology 213:280–301. https://doi.org/10.1016/j.jep.2017.11.010
13. Jin J-J, Yu W-B, Yang J-B, Song Y, dePamphilis CW, Yi T-S, Li D-Z (2020) GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. Genome Biol 21:241. https://doi.org/10.1186/s13059-020-02154-5
14. Kashi Y, King D (2006) Simple sequence repeats as advantageous mutators in evolution. Trends in Genetics 22:253–259. https://doi.org/10.1016/j.tig.2006.03.005
15. Katoh K (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Research 30:3059–3066. https://doi.org/10.1093/nar/gkf436
16. Kazimierski M, Regula J, Molska M (2019) Cornelian cherry (Cornus mas L.) – characteristics, nutritional and pro-health properties [pdf]. Acta Sci Pol Technol Aliment 18:5–12. https://doi.org/10.17306/J.AFS.2019.0628
17. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Meintjes P, Drummond A (2012) Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28:1647–1649. https://doi.org/10.1093/bioinformatics/bts199

18. Keir KR, Bemmels JB, Aitken SN (2011) Low genetic diversity, moderate local adaptation, and phylogeographic insights in *Cornus nuttallii* (Cornaceae). American J of Botany 98:1327–1336. https://doi.org/10.3732/ajb.1000466
19. Li X, Ma Q, Zhou H, Yang Y, Li H, Wang J (2020) Characterization of the complete chloroplast genome of *Cornus bretschneideri* (cornaceae). Mitochondrial DNA Part B 5:543–544. https://doi.org/10.1080/23802359.2019.1710281
20. Liesebach H, Götz B (2008) Low Chloroplast DNA Diversity in Red Dogwood (Cornus sanguinea L.). Silvae Genetica 57:291–300. https://doi.org/10.1515/sg-2008-0044
21. Lv Z-Y, Huang X-H, Luo J, Zhang X, Deng T, Li Z-M (2019) The complete chloroplast genome sequence of *Cornus sunhangii* (Cornaceae). Mitochondrial DNA Part B 4:3242–3243. https://doi.org/10.1080/23802359.2019.1669090
22. Qu X-J, Moore MJ, Li D-Z, Yi T-S (2019) PGA: a software package for rapid, accurate, and flexible batch annotation of plastomes. Plant Methods 15:50. https://doi.org/10.1186/s13007-019-0435-7
23. Qu X-J, Zou D, Zhang R-Y, Stull GW, Yi T-S (2023) Progress, challenge and prospect of plant plastome annotation. Front Plant Sci 14:1166140. https://doi.org/10.3389/fpls.2023.1166140
24. R Core Team (2018) R: A language and environment for statistical computing
25. Shi C, Hu N, Huang H, Gao J, Zhao Y-J, Gao L-Z (2012) An Improved Chloroplast DNA Extraction Procedure for Whole Plastid Genome Sequencing. PLoS ONE 7:e31468. https://doi.org/10.1371/journal.pone.0031468
26. Shi L, Chen H, Jiang M, Wang L, Wu X, Huang L, Liu C (2019) CPGAVAS2, an integrated plastome sequence annotator and analyzer. Nucleic Acids Research 47:W65–W73. https://doi.org/10.1093/nar/gkz345
27. Thomas SK, Liu X, Du Z, Dong Y, Cummings A, Pokorny L, Xiang Q (Jenny), Leebens-Mack JH (2021) Comprehending Cornales: phylogenetic reconstruction of the order using the Angiosperms353 probe set. American J of Botany 108:1112–1121. https://doi.org/10.1002/ajb2.1696
28. Tillich M, Lehwark P, Pellizzer T, Ulbricht-Jones ES, Fischer A, Bock R, Greiner S (2017) GeSeq – versatile and accurate annotation of organelle genomes. Nucleic Acids Research 45:W6–W11. https://doi.org/10.1093/nar/gkx391
29. Wick RR, Schultz MB, Zobel J, Holt KE (2015) Bandage: interactive visualization of *de novo* genome assemblies. Bioinformatics 31:3350–3352. https://doi.org/10.1093/bioinformatics/btv383
30. Xia L, Wang H, Zhao X, Obel HO, Yu X, Lou Q, Chen J, Cheng C (2023) Chloroplast Pan-Genomes and Comparative Transcriptomics Reveal Genetic Variation and Temperature Adaptation in the Cucumber. IJMS 24:8943. https://doi.org/10.3390/ijms24108943
31. Xiang Q-Y, Brunsfeld SJ, Soltis DE, Soltis PS (1996) Phylogenetic Relationships in Cornus Based on Chloroplast DNA Restriction Sites: Implications for Biogeography and Character Evolution. Systematic Botany 21:515. https://doi.org/10.2307/2419612
32. Xiong Y, Xiong Y, Jia X, Zhao J, Yan L, Sha L, Liu L, Yu Q, Lei X, Bai S, Ma X (2023) Divergence in *Elymus sibiricus* is related to geography and climate oscillation: A new look from pan-chloroplast

genome data. J of Sytematics Evolution jse.13020. https://doi.org/10.1111/jse.13020

33. Yuan W, He S, Zhang S, Chang D, He Y (2021) The complete chloroplast genome sequence of *Cornus Alba* L. (Cornaceae). Mitochondrial DNA Part B 6:1997–1998. https://doi.org/10.1080/23802359.2021.1938727
34. Zhang C, Zhang T, Luebert F, Xiang Y, Huang C-H, Hu Y, Rees M, Frohlich MW, Qi J, Weigend M, Ma H (2020) Asterid Phylogenomics/Phylotranscriptomics Uncover Morphological Evolutionary Histories and Support Phylogenetic Placement for Numerous Whole-Genome Duplications. Molecular Biology and Evolution 37:3188–3210. https://doi.org/10.1093/molbev/msaa160
35. Zheng S, Poczai P, Hyvönen J, Tang J, Amiryousefi A (2020) Chloroplot: An Online Program for the Versatile Plotting of Organelle Genomes. Front Genet 11:576124. https://doi.org/10.3389/fgene.2020.576124
36. Zhu A, Guo W, Gupta S, Fan W, Mower JP (2016) Evolutionary dynamics of the plastid inverted repeat: the effects of expansion, contraction, and loss on substitution rates. New Phytologist 209:1747–1756. https://doi.org/10.1111/nph.13743

Figures

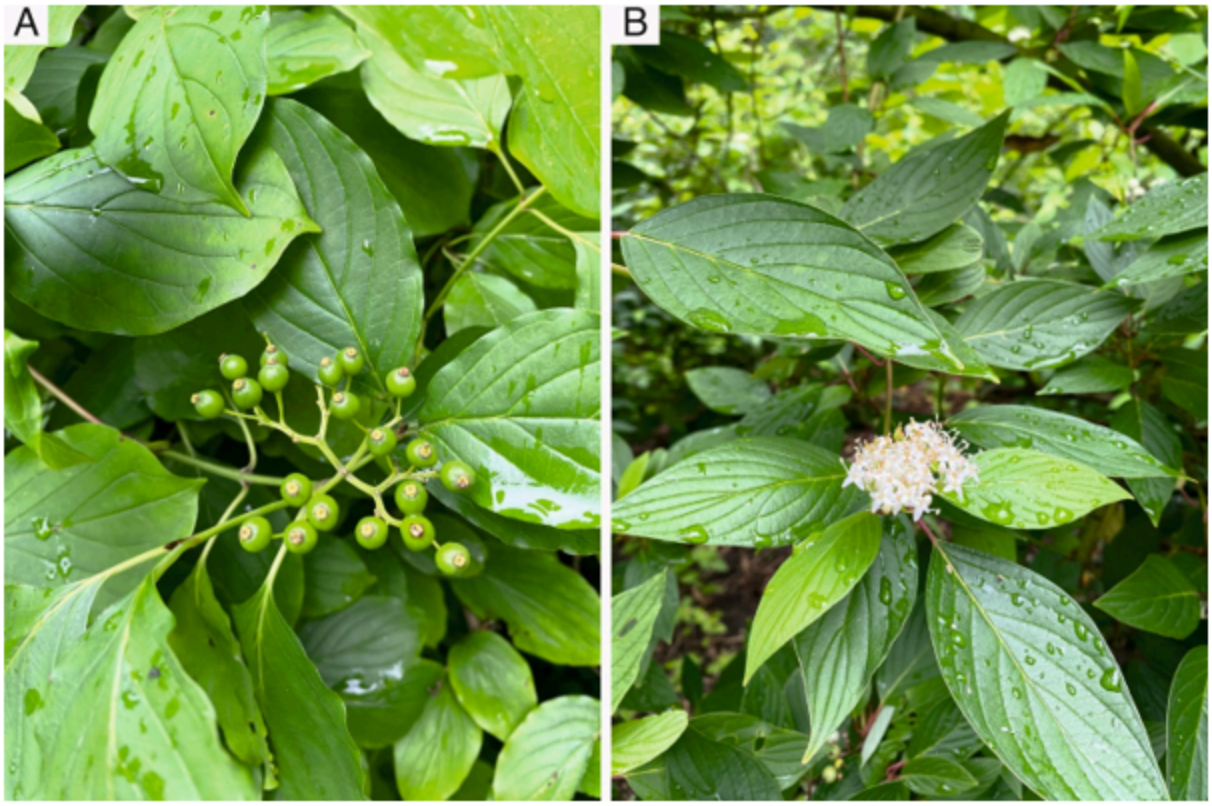

Figure 1

Photos of respective *Cornus sanguinea* L. (A) and *Cornus sericea* L. (B) used in this study

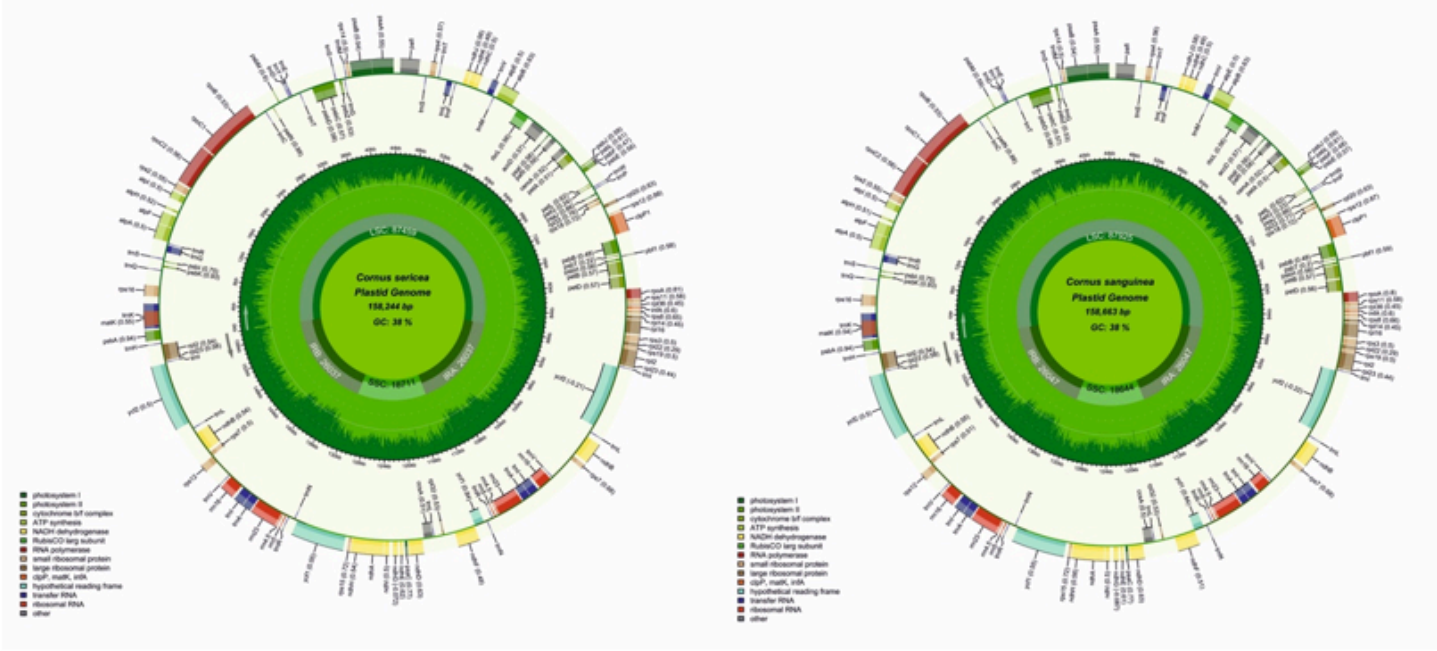

Figure 2

Maps of the complete chloroplast genome of *C. sanguinea* and *C. sericea.* The circles are commented from the inner to outer: the length and location of the corresponding SSC, IRa and IRb, and LSC regions are shown in the inner circle; the GC content is represented as the proportion of the shaded parts where the zero-level based on the outer part of circle; the gene names and their codon usage bias are shown on the outermost layer, with pseudogenes marked with asterisks; genes are color-coded by their functional classification; genes inside the circle are transcribed clockwise, genes outside the circle anticlockwise (also represented with arrows).

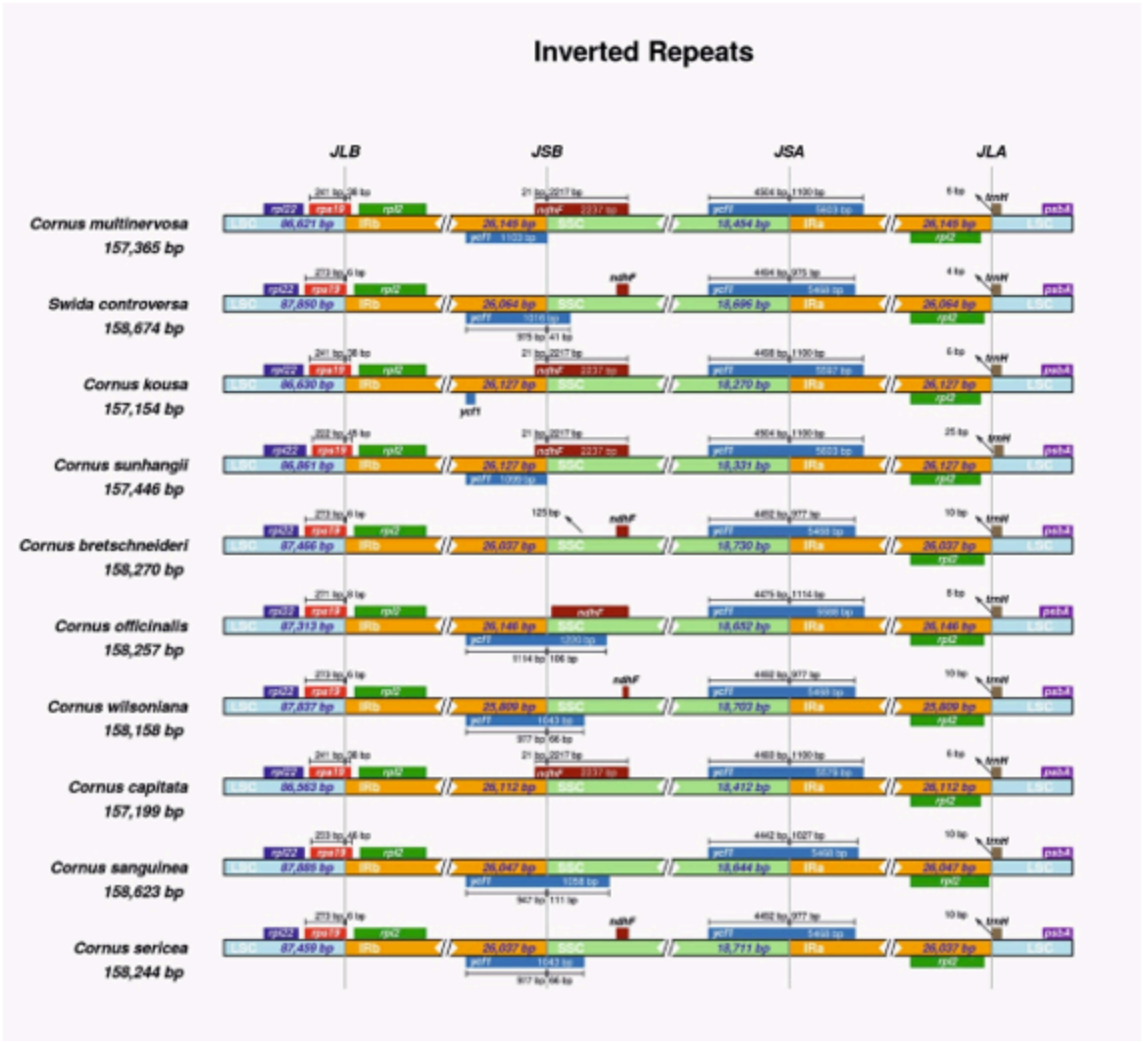

Figure 3

LSC/IR and IR/SSC of 8 representative *Cornus* species. Note: Two taxa of Cornus used in this study are depicted at the very bottom. JLB (IRb /LSC), JSB (IRb/SSC), JSA (SSC/IRa) and JLA (IRa/LSC) denote the junctions between each corresponding region. The genes located near the junctions are the realistically scaled with the bp distances. The arrows indicate the bp distance of the start or end coordinate of gene from the junction site.

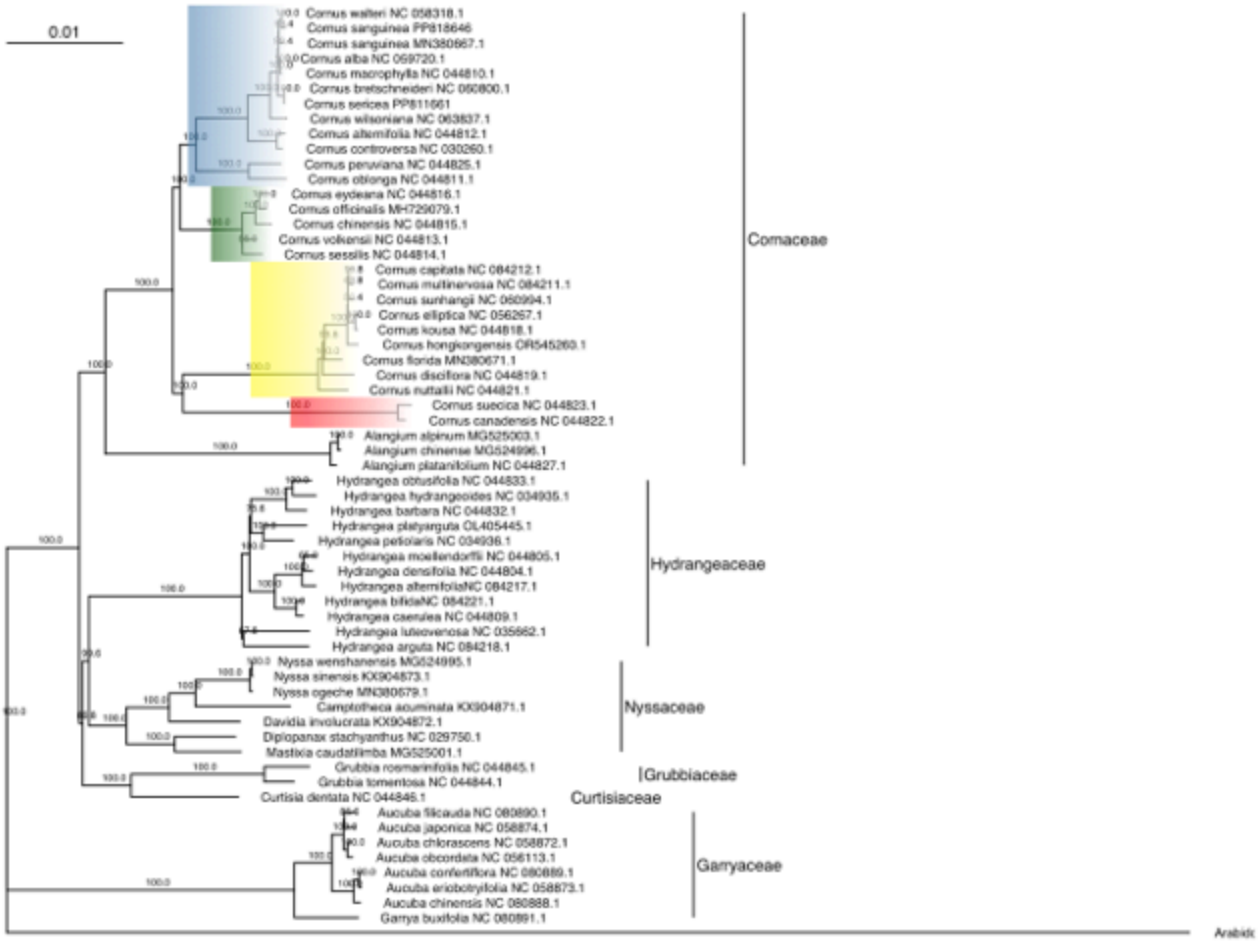

Figure 4

Phylogenetic tree inferred by the neighbor-joining method with 500 bootstrap replicates and the Tamura-Nei model based on 60 representative species. Clades withing the *Cornus* genus are highlighted by different colors: blue color for BW clade, green – for CC clade, yellow – for BB clade, and red – for DW clade.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supp.Table1.xlsx