Discovery of plastid gene mutations causing albinism in chimeric Cornus cultivars (Cornaceae)

Hayan Lee
Seoul National University
Jickerson Lado
Seoul National University
Hyeji Lee
Seoul National University
Yeon Jeong Kim
Seoul National University
Yun Sun Lee
Seoul National University
Jong-Soo Kang
Kangwon National University
Jee Young Park
Seoul National University
Jung Hwa Kang
Hantaek Botanical Garden
Hak Cheol Kwon
Korea Institute of Science and Teachnology
Tae-Jin Yang
tjyang@snu.ac.kr

Seoul National University

Article

Keywords: Cornus, chimera, plastome, pseudogenization, heteroplasmy, albino

Posted Date: June 26th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9675772/v1

Additional Declarations: No competing interests reported.

Abstract

Chimera plants are becoming popular due to their ornamental and aesthetic appeal. Here, we examined the complete plastid genomes (plastomes) and 45S nuclear ribosomal DNAs (nrDNAs) from green and albino leaf tissues (GLT and ALT) of five *Cornus* cultivars (*Cornus florida* Cherokee Daybreak and four *Cornus kousa* cultivars). We identified deleterious plastome mutations unique to the ALT of three cultivars: a missense mutation (Y29D) in *matK* (Cherokee Daybreak) impairing transcript splicing, a 1-bp deletion in *rpoB* (Melanie), and a 7-bp tandem repeat insertion in *ycf3* (Doudou). The latter two induced frameshifts and premature stop codons. Rutban and Tom Wood exhibited identical GLT and ALT plastome sequences, and no 45S nrDNA variations were detected across any cultivars. Furthermore, phylogenetic analyses clustered these cultivars within big-bracted dogwoods and resolved the maternal lineage of the hybrid Rutban as *C. florida*, correcting prior ambiguities. We also developed five specific molecular markers (two intergenic high-resolution melting and three gene-specific) that successfully discriminated these morphologically similar cultivars. These findings elucidate the diverse genetic causes of albinism and highlight plastid super-barcoding for precision breeding.

INTRODUCTION

The genus *Cornus* (Cornaceae) is a valuable plant lineage in ornamental horticulture due to its diverse array of colors, appealing flowers, and fruits[1–2]. This genus comprises approximately 60 distinct species, including the popular *C. florida* and *C. kousa*[3–4]. *Cornus florida* has gained recognition as a model tree in various scientific fields while *C. kousa* has been naturalized in the United States (U.S.) from Eastern Asia, primarily for its ornamental qualities and potential for hybridization with *C. florida*[5]. These two species have collectively given rise to 80 and 70 cultivars, respectively, owing to their practical utility[6–7]. *C. kousa* is among the popular *Cornus* species distributed in Korea, alongside *C. officinalis*, *C. alba*, *C. walteri*, *C. macrophylla*, and *C. controversa*5. *C. kousa* and *C. florida* are grouped under the big-bracted dogwoods, one of the four major subgroups within the genus *Cornus*, which also includes cornelian cherries, dwarf dogwoods, and blue- or white-fruited dogwoods. These groups are based on their distinct morphological characters and molecular phylogenetic studies[8–10].

Notably, *C. florida* Cherokee Daybreak and certain *C. kousa* cultivars Doudou, Melanie, Tom Wood, and Rutban, exhibit chimeric phenomena, displaying varied colors within the same leaves[11]. Many studies have explored various factors contributing to chimeric occurrences, including genetic alterations in the nuclear genomes and plastomes, transposable genetic elements, graft-induced chimeras, semigamy where male and female nuclei fail to fuse, and tissue culture synthesis of chimeric meristems[12–15].

Chimeras are closely related to the developmental stages and structures of shoot apical meristem (SAM) in dicotyledonous plants, which mainly comprises three layers: L1 (epidermal), L2 (sub-epidermal), and L3 (inner tissues)[7, 13, 16]. Chimeric patterns can be categorized as periclinal, mericlinal, or sectorial, depending on how heterogenomic cells are distributed across these layers[17–18]. Chimeras are associated with irregularities in the biogenesis, regulation, and development of the chloroplast, which is central to photosynthesis[19–21]. The chloroplast genome contains genes essential for photosynthesis located in its large single-copy (LSC), small single-copy (SSC), and two copies of inverted repeat (IR) regions. Plastomes are typically 115 to 165 kb in length, with each cell containing between 1,000 to 10,000 copies[22]. Plastome inheritance is typically maternal through three mechanisms: Lycopersicon, Solanum, and Triticum types, which excludes plastids from generative cells during pollen mitosis, degrades plastids and/or their DNA during pollen maturation, and eliminates paternal plastids during fertilization, respectively[23]. However, instances of plastid biparental inheritance have been reported, especially in angiosperms[24–25]. Moreover, plastomes are predominantly paternally inherited in gymnosperms[26]. Different inheritance patterns may be linked to a variegated phenotype and could be controlled by environmental and genetic factors[27–28].

Recent studies on ornamental chimeric plants showed that small deletion-mediated frameshift or single nucleotide substitution of plastid genes influences photosystem function and chloroplast development[29] while in the variegated ornamental plant with albino sectors, *Dianella tasmanica*[30], structural variation caused by intermolecular recombination generating a giant plastome was observed. Despite these general insights, the precise genetic underpinnings of variegation in commercially significant *Cornus* cultivars and the development of robust molecular tools for their authentication and lineage verification have been less comprehensively explored. Hence, we investigated the possible genetic causes of the chimeric phenotypes in *C. florida* and four *C. kousa* cultivars through whole-genome sequencing of their GLT and ALT. By assembling the plastome and 45S nrDNA and performing sequence variation and phylogenetic analyses, we demonstrated how DNA super-barcoding can explain chimeric phenomena and verify the parentage of hybrids.

RESULTS

Assembly and annotation of plastome and 45S nrDNA in Cornus cultivars

All *Cornus* cultivars used in this study showed a chimeric leaf pattern, characterized by a green center and distinctly colored margins (Fig. 1a). The leaf margins were white in Melanie, pale green in Doudou and Rutban, and pale yellow in Cherokee Daybreak and Tom Wood. Sequencing data for the plastome and 45S nrDNA were generated from the ALT and GLT of Cherokee Daybreak, Doudou, Melanie, Rutban, and Tom Wood (Supplementary Table S1). Illumina MiSeq produced NGS data ranging from 1.0 to 1.6 Gb per sample. The plastome coverage ranged from 230.17x to 626.94x, and 45S nrDNA coverage ranged from 249.85x to 970.52x (Supplementary Figure S1). The difference in plastome average read depth was 1.5 to 2.2 times higher in ALT than in GLT in all *C. kousa* cultivars (Supplementary Table S1). The lengths of the plastome regions varied among cultivars (Table 1; Supplementary Figure S2). LSC ranged from 86,598 bp to 86,780 bp, SSC from 18,270 bp to 18,455 bp, and the IR regions from 26,127 bp to 26,153 bp (Table 1).

Table 1
Plastomes and 45S nrDNAs characteristics of *Cornus* chimeric cultivars.

No.	Scientific name	Leaf tissue color	Plastid genome									45S nrDNA Transcription unit (bp)
			Genome (bp)	LSC (bp)	SSC (bp)	IR (bp)	No. CDS	No. pseudogene	No. tRNA	No. rRNA	CG%	
1	*Cornus florida*	GLT	157,540	86,779	18,455	26,153	85	0	38	8	38.1	5,801
	Cherokee Daybreak	ALT	157,540	86,779	18,455	26,153	85	1	38	8	38.1	5,801
2	*Cornus kousa*	GLT	157,184	86,660	18,270	26,127	85	0	38	8	38.1	5,804
	Melanie	ALT	157,183	86,659	18,270	26,127	84	1	38	8	38.1	5,804
3	*Cornus kousa*	GLT	157,339	86,662	18,423	26,127	85	0	38	8	38.1	5,803
	Doudou	ALT	157,346	86,669	18,423	26,127	84	1	38	8	38.1	5,803
4	*Cornus kousa*	GLT	157,541	86,780	18,455	26,153	85	0	38	8	38.1	5,799
	Rutban (Aurora®)	ALT	157,541	86,780	18,455	26,153	85	0	38	8	38.1	5,799
5	*Cornus kousa*	GLT	157,283	86,598	18,431	26,127	85	0	38	8	38.1	5,803
	Tom Wood	ALT	157,283	86,598	18,431	26,127	85	0	38	8	38.1	5,803

Note: GLT, Green leaf tissue; ALT, Albino leaf tissue

Annotation of the assembled complete plastomes identified a total of 85 protein-coding sequences (Table 1; Supplementary Figure S2). Among them, one putative pseudogene was identified in the ALT of the Cherokee Daybreak, Melanie, and Doudou, which were the *matK*, *rpoB* and *ycf3*, respectively (Table 2; Fig. 1C). Additionally, 38 tRNAs, and eight rRNAs were identified (Table 1; Supplementary Figure S2). To obtain the complete 45S nrDNA unit sequences, we selected the major assembled contigs and performed manual curation. The length of one unit of 45S nrDNA ranged from 5,799 to 5,804 bp (Table 1).

Table 2. Plastid mutations and percentages of wild and mutant-type reads identified in three *Cornus* chimeric cultivars.

No.	Scientific name	Gene	Mutation			Genotype		Depth (Ratio)			
			Type	Length	Position/bp	GLT	ALT	GLT		ALT	
								Wild	Mutant	Wild	Mutant
1	*Cornus florida* Cherokee Daybreak	*matK*	SNP	1bp	3,425	A	C	346 (81.6%)	78 (18.3%)	76 (16.48%)	385 (82.8%)
2	*Cornus kousa* Melanie	*rpoB*	Deletion	1bp	26,887	T	-	111 (47.2%)	124 (52.8%)	8 (2.0%)	385 (98.0%)
3	*Cornus kousa* Doudou	*ycf3*	Insertion	7bp	45,697	-	TCTGTAA	80 (37.9%)	131 (62.1%)	37 (8.8%)	384 (91.2%)

Note: GLT, Green leaf tissue; ALT, Albino leaf tissue

Analysis of plastid heteroplasmy and associated gene mutations

Sequence alignment of the GLT and ALT plastomes revealed two distinct groups among the *Cornus* cultivars. The plastomes of Rutban and Tom Wood were homoplasmic, showing no variation between the two tissue types while Cherokee Daybreak, Melanie and Doudou were heteroplasmic, with each showing a single, unique variation site (Fig. 1b and 1c, Table 2).

Specifically, the polymorphic site in Cherokee Daybreak was an A3425C nucleotide substitution in the *matK* gene that led to a Y29D missense mutation (Fig. 1c-e; Supplementary Figure S3A). The PROVEAN analysis (Supplementary Table S2) predicted this amino acid change to be deleterious (score = -4.690)[31]. The Melanie ALT plastomes exhibited a 1bp-deletion (A) at position 26,887 in the *rpoB* plastid gene. This deletion caused a frameshift that resulted in a premature stop codon, producing a truncated protein of 223 amino acids (Fig. 1c; Supplementary Figure S3B). The Doudou ALT plastome showed a 7bp-insertion (ATTACAG) in the *ycf3* gene, starting at position 45,697. This insertion, which appears to be a tandem repeat, also induced a frameshift that truncated protein to 61 amino acids (Fig. 1c; Supplementary Figure S3C). These mutations were absent from publicly available *Cornus* plastome sequences in the NCBI database, indicating that they are novel variations specific to the ALT in these cultivars (Supplementary Figure S4).

Furthermore, we quantified the proportions of wild-type and mutant reads within each GLT and ALT of heteroplasmic cultivars by mapping raw reads to their respective plastome sequences (Table 2). In the GLT, the percentage of mutant reads was 18.3% for Cherokee Daybreak, 52.8% for Melanie, and 62.1% for Doudou. In contrast, the mutant allele was prevalent in the ALT of all three cultivars, constituting over 80% of the total reads (Table 2).

Analysis of intra- and interspecific variation and phylogeny of plastome and 45S nrDNA

To assess the plastome diversity among the five *Cornus* cultivars, we identified SNPs and InDels by comparing their plastome sequences (Table 3). The analysis revealed two distinct groups based on sequence variation. The first group, comprising Cherokee Daybreak and Rutban, showed high similarity, with only 3–5 SNPs and 1 InDel differentiating their GLT and ALT. In contrast, this group was highly divergent from the second group (Doudou, Melanie, and Tom Wood), from which they differed by 741–761 SNPs and 135–139 InDels. Moderate variation was also observed among the members of the second group, which differed from each other by 24–74 SNPs and 12–34 InDels.

Table 3
Variations among *Cornus florida* and *C. kousa* cultivars plastome.

InDel / SNP	CFC-GLT	CFC-ALT	CKR-GLT	CKR-ALT	CKM-GLT	CKM-ALT	CKT-GLT	CKT-ALT	CKD-GLT	CKD-ALT
CFC-GLT	-	0	1	1	138	139	135	135	139	139
CFC-ALT	2	-	1	1	138	139	135	135	139	139
CKR-GLT	3	5	-	0	138	139	135	135	138	139
CKR-ALT	3	5	0	-	138	139	135	135	138	139
CKM-GLT	761	763	760	760	-	1	32	32	32	33
CKM-ALT	761	763	760	760	0	-	33	33	33	34
CKT-GLT	742	744	741	741	70	70	-	0	12	13
CKT-ALT	742	744	741	741	70	70	0	-	12	13
CKD-GLT	755	757	754	754	74	74	24	24	-	1
CKD-ALT	755	757	754	754	74	74	24	24	0	-

Note: CFC, *C. florida* Cherokee Daybreak; CKD, *C. kousa* Doudou; CKM, *C. kousa* Melanie; CKR, *C. kousa* Rutban; CKT, *C. kousa* Tom Wood; GLT, Green leaf tissue; ALT, Albino leaf tissue; SNP, Single Nucleotide Polymorphism; InDel, Insertion/Deletion

A phylogenetic tree generated from the complete plastome sequences encompassing additional 27 *Cornus* species provided further insight into these relationships (Fig. 2a). In this phylogenetic tree, the *C. kousa* cultivars were separated into different clades: Melanie clustered with the reference *C. kousa* from NCBI, while Doudou and Tom Wood formed a distinct group that was sister to *C. elliptica*. Notably, Cherokee Daybreak and Rutban formed a strongly supported monophyletic clade (89% bootstrap value), grouping 100% with the other *C. florida* (NC_044820.1) under subgroup Cynoxylon, instead of the other *C. kousa* cultivars under the subgroup Syncarpea. These findings were strongly corroborated by a separate phylogenetic analysis using the 45S nrDNA sequences (Fig. 2b). In this tree, Tom Wood clustered with the reference *C. kousa*, while Melanie grouped with *C. elliptica* (83% bootstrap support), and Doudou was sister to them. The analysis supported the position of Rutban, which once again clustered with *C. florida* with maximum bootstrap support (100%).

Molecular marker design and validation

Based on sequence variations among the *C. florida* and four *C. kousa* cultivars, two high-resolution melting (HRM) markers were designed to discriminate between the group of Cherokee Daybreak and Rutban and the group of Doudou, Melanie, and Tom Wood (Supplementary Table S3). The marker C_HRM1 was derived from an intergenic sequence (IGS) between *matK* and *rps16* with product sizes of about 92–99 bp. C_HRM2 was based on an IGS between *psbZ* and *rps14*, producing about 149–152 bp amplicons. As expected from the amplicon size differences, the HRM analysis using C_HRM1 successfully separated the cultivars into two distinct clusters that corresponded to these predefined groups (Fig. 3a). The analysis using C_HRM2 also generated five distinct melting profiles, grouping the cultivars individually (Fig. 3b). Results of the three designed specific genic markers (*matK*, *rpoB*, and *ycf3*) showed successful discrimination of the GLT and ALT. CF_1 marker, specific for the *matK* nucleotide substitution, differentiated GLT (CFC_G, 78.5 ℃) and ALT (CFC_W, 78.15 ℃) of Cherokee Daybreak (Fig. 3c), and the clustering of melting temperature for all *C. kousa* cultivars (CK, 78.30 ℃). CK_1, an HRM designed specifically for a single nucleotide deletion in the *rpoB* of the Melanie ALT had a peak at 78.10 ℃ (CKM_W) compared to the GLT (CKM_G, 78.00 ℃) and the cluster of Cherokee Daybreak (CF) and other *C. kousa* cultivars (CK) (78.2 ℃) (Fig. 3d). CK_2 was designed to detect the 7-bp insertion in *ycf3* at the Doudou ALT. Results of the endpoint PCR showed 247 bp band for the Doudou ALT while 240 bp band for Doudou GLT and the other cultivars (Fig. 3e). These markers can be used to correctly identify these *C. florida* and *C. kousa* cultivars, since reports of mislabeling and mishandling of private breeders and sellers had been reported[32–33].

DISCUSSION

Albino genes in the plastome of chimeric plants

Our comprehensive whole-genome sequencing enabled the determination of both the complete plastomes and 45S nrDNA sequences of the *C. florida* and four *C. kousa* cultivars. While no variations were detected between GLT and ALT in the 45S nrDNA in all cultivars, heteroplasmic plastomes were identified in Cherokee Daybreak, Melanie, and Doudou with three plastid genes, *matK*, *rpoB*, and *ycf3*,

respectively (Fig. 1; Supplementary Figure S3). Rutban and Tom Wood exhibited no variations between the GLT and ALT in their plastomes. A similar previous study reported some heteroplasmic plastomes and mutated genes including *rpoB* and *ycf3* among chimeric plants29. These mutations induced frameshift and premature stop codon, leading to the pseudogenization of the genes, resulting in the observed albino phenotype (Supplementary Figure S3A-B). In white barley mutant *albostrians*[34], *matK* plays a crucial role in chloroplast function and intron excision. In tobacco, pull-down assays linked *matK* to seven group IIA introns[35] like those identified in *albostrians*, its overexpression caused a variegated cotyledon development phenotype, and its complete knockout has failed as it was lethal[36], indicating the essential role of *matK* in plant survival.

A nucleotide substitution in the 3425th position of the Cherokee Daybreak GLT and ALT led to a missense mutation, affecting the MatK/TrnK N-terminal region (Fig. 1c-e; Supplementary Figure S3A). This mutation affects the interaction between the amino acids, potentially leading to a different protein conformation. MatK proteins are encoded by the trnK intron in land plants and act as a general splicing factor for multiple group IIA introns in chloroplasts[37–38]. A recent study reported the structural interaction of MatK protein with MATURASE K INTERACTING PROTEIN1 (MKIP1), a crucial step for plastidial intron splicing[39]. Defects in MatK function can lead to severe splicing impairments, resulting in defective chloroplast development and function, often manifesting as an albino phenotype due to impaired photosynthesis[36, 40]. The *rpoB* gene product, the β subunit, is part of a larger multi-subunit complex of PEP, which plays a crucial role in transcribing all photosynthesis-related genes in chloroplast genomes during plant development[41–46]. The *rpoB* mutation in the ALT of the Melanie cultivar resulted in the deletion of essential *rpoB* domains, including several RNA polymerase Rpb2 domains and binding sites leading to an improper assembly and dysfunctional PEP (Supplementary Figure S3B). The *ycf3* plastid gene, with an observed mutation in the ALT of the Doudou cultivar, encodes an essential chaperone-like protein for photosystem I (PSI) formation[47–49] through tetratricopeptide repeat (TPR) domains, a degenerate 34-amino acid consensus sequence frequently repeated in tandem arrays[50]. The 7-bp insertion at the *ycf3* gene of the ALT of the Doudou cultivar led to an internal stop codon producing only 60 amino acids compared to the GLT with 168 residues. Protein classification and analysis revealed that this incomplete transcription led to the non-expression of the conserved three TPR domains in the Doudou ALT, potentially limiting PSI biogenesis and affected chloroplast development in these *ycf3*-mutant tissue regions (Supplementary Figure S3C).

The biased distribution of plastids in Melanie and Doudou likely influences the expression of their chimeric phenotypes. Specifically, their GLT were heteroplasmic, containing a balanced mix of wild-type and mutant plastomes. In contrast, their ALT were nearly homoplasmic due to a heavily biased accumulation of mutant plastomes (Table 3). This differential accumulation within an individual plant can be attributed to the 'sorting-out' process, where random segregation of plastids during cell division leads to varied plastid compositions in different tissues[27].

Phylogenetic relationships and marker development in C. kousa cultivars

Among the *C. kousa* species used in the phylogenetic tree reconstruction, most of the *C. kousa* cultivars clustered together, except for Rutban cultivar grouping closely with *C. florida* Cherokee Daybreak and *C. florida* (NC_044820.1) (Fig. 3). Interestingly, *Cornus kousa* Rutban, the only hybrid cultivar under investigation, originated from crossbreeding *C. kousa* 'K2' (maternal) and *C. florida* springtime (paternal), which is one of the *Cornus* × *rutgersensis* cultivars[7, 51]. From this background, however, the phylogenetic result suggests incongruence with the maternal inheritance pattern for the plastome. We further investigated this inference by conducting read mapping on the assembled Rutban cultivar plastid sequence using the public data of *C. kousa* 'K2' (SRR23225011) and *C. florida* springtime (SRR23224781). The coverage of *C. florida* springtime (172.76x) was higher than *C. kousa* 'K2' (120.96x). Given this result, Rutban has *C. florida* as its maternal source, and the potential occurrence of misidentification, mislabeling, or any errors in handling during hybridization may have occured. Several reports mentioned the possible mix-ups of these morphologically similar *Cornus* species[32–33]. Here, we have shown that DNA super-barcoding is a practical and essential tool for detection and authentication[52].

Additionally, we aimed to precisely identify cultivars within the genus *Cornus* by designing new HRM markers. With correct identification and authentication, these results can append the current population structure and genetic variations of *Cornus* cultivars. We presented another layer of plastome-based variations leading to the chimeric phenomena in plants, specifically using *C. florida* and *C. kousa* cultivars. This paper showcased the diverse and equally accurate use of low-coverage NGS data to build the plastome and 45S nrDNA in performing genetic diversity and evolutionary studies.

MATERIALS AND METHODS

Sample collection and sequencing

A total of five *Cornus* cultivars (*C. florida* Cherokee Daybreak and *C. kousa* Doudou, Melanie, Rutban, and Tom Wood) were collected from the Hantaek Botanical Garden, Yongin, Gyeonggi-do, the Republic of Korea (http://www.hantaek.co.kr/) under the permission of and assisted by the garden authority (Fig. 1a). Each GLT and ALT were isolated from the same leaf using scissors and then ground using a mortar and pestle in liquid nitrogen. Genomic DNA was extracted using a GeneAll PLANT SV DNA Extraction kit (GeneAll Biotechnology, Seoul, Republic of Korea) following the manufacturer's instructions. The quality and concentration of the extracted DNA were determined with a UV spectrophotometer (Nanodrop ND-1000; Thermo Fisher Scientific, Waltham, MA, USA) and agarose gel electrophoresis. The genomic DNA from GLT and ALT was used for paired-end library construction and next-generation sequencing (NGS) using the Illumina MiSeq platform (Illumina, USA) at Phyzen (www.phyzen.com, Seongnam, South Korea).

Plastome analysis and functional characterization of matK, rpoB and ycf3

High-quality paired-end reads were used for the *de novo* assembly of the complete plastome and 45S nrDNA sequences following the *de novo* assembly of low coverage WGS (dnaLCW) method[53–54] using a CLC genome assembler (ver. 4.21.104315, CLC Inc., Aarhus, Denmark). Raw paired-end reads were trimmed by CLC quality trim software with an offset of quality score value 33. The trimmed reads were then assembled de novo with an overlap distance ranging from 150 to 500 bp and a set window size of 32 for plastome and 64 for 45S nrDNA. The initial contigs were extracted by MUMmer and BLASTZ[55–56], subsequently arranged, and merged into a single draft sequence by comparison to the reference plastome from *C. florida* (NC_044820.1) and *C. kousa* (NC_044818.1) sequences from the National Center for Biotechnological Information (NCBI) database and reference 45S nrDNA sequence from *Deutzia scabra* (SRR14570892). For assembly validation, raw reads were remapped to the final assembled sequences utilizing an integrated mapping tool (clc_mapper) within the CLC assembly cell. BWA v.0.7.17-r118857[59] and samtools v.1.2258[58] generated the coverage data and visualized using a custom R script 4.3.059[59] with libraries ggplot2 4.0.360[60], dplyr 1.2.161[61], zoo 1.8–1562[62], and patchwork 1.3.263[63].

The resulting plastome sequences were annotated using GeSeq (https://chlorobox.mpimp-golm.mpg.de/geseq.html)[64] and manually corrected using the Artemis program[65]. Circular maps of the plastome were generated initially using the OGDRAW program (https://chlorobox.mpimpgolm.mpg.de/OGDraw.html)[66] and Chloroplot[67], both manually modified in Keynote [v.14.4 (7043.0.93), USA]. The 45S nrDNA region, including 18S, 5.8S, 26S, and two internal transcribed spacer (ITS) regions, was annotated using a RNAmmer[68], and BLASTN (https://blast.ncbi.nlm.nih.gov/)[69].

DNA and protein sequences of *matK*, *rpoB* and *ycf3* from GLT and ALT of Cherokee Daybreak, Melanie, and Doudou cultivars, respectively, were extracted using the Artemis program v.18.2.064. AlphaFold Protein Structure and Database (AFDB)[70] was used to predict protein structure models (https://alphafold.ebi.ac.uk) based on available *matK* (AF-A0A481WER4-F1-v4) for *C. florida*, and *rpoB* (AF-A0A5C0Q5B1-F1-v4) and *ycf3* (AF-A0A5C0Q6P5-F1-v4) for *C. kousa*. Similarly, SWISS-MODEL[71], accessed via the Expasy web server (https://swissmodel.expasy.org) was utilized for protein structure homology modeling. InterPro version 102.0[72] provided an integrative classification of protein sequences into families and identified functionally important domains and conserved sites[73]. PEP-PAP complex and PSI assembly protein models were downloaded from RCSB Protein Data Bank (RCSB PDB) (https://www.rcsb.org). MatK GLT and ALT protein pdb files were visualized using the web program Molstar-based Protein Visualizer (https://stevenyuyy.us/protein-viewer/), and the images exported were illustrated in Keynote [version 14.4 (7043.0.93), USA]. PROVEAN v1.1.3 (Protein Variation Effect Analyzer)[31], a software tool which can predict the effects of an amino acid substitution was utilized for *matK* protein of Cherokee Daybreak GLT and ALT.

Comparative and phylogenetic analyses of plastomes and 45S nrDNA

The assembled plastome sequences between GLT and ALT of the same individuals were aligned and compared using MAFFT 7.520[74] to identify sequence variations. For phylogenetic analysis, newly assembled plastome sequences of *C. florida* and *C. kousa* cultivars and other 27 *Cornus* species were downloaded from NCBI GenBank and the plastome sequence of *Alangium kurzii* (NC_044826.1) was included as an outgroup. Based on the aligned plastome sequences, phylogenetic analysis was performed using the Maximum likelihood (ML) method with 1000 bootstrap replications using the IQ-TREE[75] and visualized in FigTree v1.4.4 (http://tree.bio.ed.ac.uk/software/figtree/). Similarly, the 45S nrDNA of *C. florida* and *C. kousa* cultivars with five other *Cornus* species were retrieved from NCBI GenBank, and the 45S nrDNA of *Alangium barbatum* (SRR7122029) was utilized as an outgroup.

2.4. Molecular marker development for cultivar identification of Cornus kousa

Insertions/deletions (InDels) and single nucleotide polymorphisms (SNPs) in the plastomes of Cherokee Daybreak, Doudou, Melanie, Rutban, and Tom Wood were examined by aligning all the GLT and ALT sequences in MAFFT 7.52074. SNPs and InDels were manually

curated following alignment. To develop molecular markers, two intergenic sequence sites and three specific genic sequences (*matK*, *rpoB*, and *ycf3*) were utilized to design and generate primers for high-resolution melting (HRM) analysis and end-point PCR. Primers were designed using Primer-BLAST[76]. The newly designed markers were applied to the same samples used for NGS to check their intraspecific and interspecific polymorphisms.

The plastome-based HRM analysis was performed using a LightCycler 96 Real-time PCR Machine (Roche Applied Science, Indianapolis, United States) under the following cycling program: pre-denaturation at 95℃ for 10 min; 45 cycles of 95℃ for 20 s, annealing at 58℃ for 20 s, extension at 72℃ for 20 s, and high-resolution melting at 95℃ for 1 min, 65℃ for 1 s. The reaction was conducted in 20-uL reactions containing 1 unit of Taq DNA polymerase (Inclone Biotech Co, Yongin, Republic of Korea) or 0.2 uL of Inclone™ Taq polymerase (5 U/uL), 1uL of DNA template (25 ng/uL), 2 uL of 10× Inclone Taq run buffer, 0.4 uL of dNTP mixture (each 10 mM), 0.2 uL of Inclone™ Taq polymerase (5 U/uL), 1 uL of syto9 (2.5 uM) (Invitrogen, Seoul, Republic of Korea).

Declarations

Competing Interests

The authors declare no competing interests.

Supplementary Information

Supplementary Material 1 (Figures)

Supplementary Material 2 (Tables)

Funding Statement

This research was supported by a research grant from Korea Institute of Science and Technology (KIST) Open Research Program (ORP) BlueBell Research Program (2E30650) and a grant of the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health and Welfare, Republic of Korea (grant number: HI23C1420).

Author Contribution

H.Y.L., J.P.L, J.S.K., J.Y.P., and T.-J.Y. designed the research. J.H.K provided the chimeric plant samples. H.Y.L., J.P.L, and H.J.L. assembled the chloroplast genomes of chimeric plants and conducted comparative genome analysis. H.Y.L., J.P.L, and H.J.L. designed the primers and performed the molecular marker tests. J.P.L., S.Y.L., and Y.J.K. performed the protein modeling and analysis. H.Y.L., J.P.L, and H.J.L. and T.-J.Y. wrote the manuscript, and H.Y.L., J.P.L, H.J.L., J.S.K., Y.J.K., S.L.Y., J.Y.P., J.H.K., H.C.K., and T.-J.Y. revised the manuscript.

Acknowledgement

We thank Hantaek Botanical Garden for kindly providing the Cornus chimera materials.

Data Availability

The information on the project, plant samples, and raw data used for the assembly can be accessed through NCBI BioProject PRJNA1458545. For the BioProject and associated SRA metadata, here is the NCBI-provided reviewer link in read-only format: https://dataview.ncbi.nlm.nih.gov/object/PRJNA1458545?reviewer=et49j8o939uv5g2htpbpakeupj. The complete chloroplast genome and 45S nrDNA sequences of Cornus cultivars used in this study has been submitted to NCBI GenBank. For the complete chloroplast genome, it is currently undergoing NCBI validation. The 45S nrDNA sequences are already uploaded and set for public availability once published in the journal. In order to access these datasets, kindly use this link: https://figshare.com/s/564ef3cfd04999544b7d.

References

1. Cappiello, P. & Shadow, D. Dogwoods: Genus Cornus (Timber, (2005).
2. Morozowska, M., Woznicka, A. & Nowinska, R. The morphology and internal structure of dogwood (*Cornus* L.) endocarps in the taxonomy and phylogeny of the genus. *PeerJ* **9**, e12170. https://doi.org/10.7717/peerj.12170 (2021).
3. Bailey, L. H. *Manual of cultivated plants. Manual of Cultivated Plants (Rev. Edn)* (Macmillan, 1977).
4. Wadl, P. A. et al. Molecular identification keys for cultivars and lines of *Cornus florida* and *C. kousa* based on simple sequence repeat loci. *J. Am. Soc. Hortic. Sci.* **133**, 783–793 (2008).
5. Hamm, T. P. et al. Ornamental origins and genomic frontiers: a review of big-bracted dogwood research. *Front. Plant. Sci.* **16**, 1735902. 10.3389/fpls.2025.1735902 (2026).
6. Smith, K., Williams, D. J., Tilt, K. & Davidson, C. Selection and care of dogwoods. Resources N; Report No.: ANR-1077 (2013).
7. Moreau, E. L., Medberry, A. N., Honig, J. A. & Molnar, T. J. Genetic diversity analysis of big-bracted dogwood (*Cornus florida* and *C. kousa*) cultivars, interspecific hybrids, and wild-collected accessions using RADseq. *Plos one*, *19*(7), p.e0307326 (2024).
8. Eyde, R. H. The case for keeping *Cornus* in the broad Linnaean sense. *Systematic Botany* 505–518 (1987).
9. Fu, C. N. et al. Plastid phylogenomics and biogeographic analysis support a trans-Tethyan origin and rapid early radiation of Cornales in the Mid-Cretaceous. *Molecular Phylogenetics and Evolution*, 140, 106601 (2019). (2019).
10. Du, Z. Y. et al. An updated phylogeny, biogeography, and PhyloCode-based classification of Cornaceae based on three sets of genomic data. *American Journal of Botany*, *110*(2), p.e16116 (2023).
11. Lee, Y. M. et al. *Native Cornus and 200 Cultivar Cornus* (National Arboretum Exhibition Education Research Development, 2018).
12. Clowes, F. A. L. Chimeras and meristems. *Heredity* **11** (1), 141–148 (1957).
13. Tilney-Bassett, R. A. & Plant Chimeras, E. Arnold, London (1986).
14. Marcotrigiano, M. Chimeras and Variegation: Patterns of Deceit. *Hort Sci.* **32**, 773–784 (1997).
15. Chung, K. P. Cytoplasmic inheritance: The transmission of plastid and mitochondrial genomes across cells and generations. *Plant Physiol.* **198** (1), kiaf168 (2025).
16. Satina, S., Blakeslee, A. F. & Avery, A. G. Demonstration of the Three Germ Layers in the Shoot Apex of *Datura* by Means of Induced Polyploidy in Periclinal Chimeras. *Am. J. Bot.* **27**, 895–905 (1940).
17. Filippis, I., Lopez-Cobollo, R., Abbott, J., Butcher, S. & Bishop, G. J. Using a periclinal chimera to unravel layer-specific gene expression in plants. *Plant. J.* **75**, 1039–1049 (2013).
18. Frank, M. H. & Chitwood, D. H. Plant chimeras: The good, the bad, and the 'Bizzaria'. *Dev. Biol.* **419**, 41–53 (2016).
19. Legen, J., Wanner, G., Herrmann, R. G., Small, I. & Schmitz-Linneweber, C. Plastid tRNA genes trnC-GCA and trnN-GUU are essential for plant cell development. *Plant J.* **51** (5), 751–762 (2007).
20. Sadali, N. M., Sowden, R. G., Ling, Q. & Jarvis, R. P. Differentiation of chromoplasts and other plastids in plants. *Plant. Cell. Rep.* **38**, 803–818 (2019).
21. Shimada, H. & Sugiura, M. Fine structural features of the chloroplast genome: comparison of the sequence chloroplast genomes. *Nucleic Acids Res.* **19**, 983–995 (1991).
22. Dobrogojski, J., Adamiec, M. & Luciński, R. The chloroplast genome: a review. *Acta Physiol. Plant.* **42** (6), 98 (2020).
23. Bock, R. *Cell and Molecular Biology of Plastids* Vol. 19 (Springer Science & Business Media, 2007).
24. Hu, Y., Zhang, Q., Rao, G. & Sodmergen Occurrence of plastids in the sperm cells of Caprifoliaceae: biparental plastid inheritance in angiosperms is unilaterally derived from maternal inheritance. *Plant Cell Physiol.* **49** (6), 958–968 (2008).
25. Barnard-Kubow, K. B., McCoy, M. A. & Galloway, L. F. Biparental chloroplast inheritance leads to rescue from cytonuclear incompatibility. *New Phytol.* **213** (3), 1466–1476 (2017).
26. Adams, R. P. Inheritance of chloroplasts and mitochondria in conifers: a review of paternal, maternal, leakage and facultative inheritance. *Phytologia* **101** (2), 134–138 (2019).
27. Chung, K. P., Gonzalez-Duran, E., Ruf, S., Endries, P. & Bock, R. Control of plastid inheritance by environmental and genetic factors. *Nat. Plants*. **9**, 68–80 (2023).
28. Sakamoto, W. & Takami, T. Maternal plastid inheritance: two abating factors identified. *Trends Genet.* **39**, 342–343 (2023).
29. Park, H. S. et al. High-throughput discovery of plastid genes causing albino phenotypes in ornamental chimeric plants. *Hort Res*. 10(1); (2023). 10.1093/hr/uhac246 (2023).
30. Zhou, S., Ma, K., Mower, J. P., Liu, Y. & Zhou, R. Leaf variegation caused by plastome structural variation: an example from *Dianella tasmanica*. *Hortic. Res.* **11** (3). 10.1093/hr/uhae009 (2024).

31. Choi, Y. & Chan, A. P. PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. *Bioinformatics* **31** (16), 2745–2747 (2015).
32. Windham, M. T. & Trigiano, R. N. Are 'Barton'and 'Cloud 9' the same cultivar of *Cornus florida* L.? *Journal of Environmental Horticulture*, 16(3), 163–166 (1998). (1998).
33. Trigiano, R. N., Ament, M. H., Windham, M. T. & Moulton, J. K. Genetic profiling of red-bracted *Cornus kousa* cultivars indicates significant cultivar synonomy. *Hort Sci.* **39** (4), 4 (2004).
34. Vogel, J., Hübschmann, T., Börner, T. & Hess, W. R. Splicing and intron-internal RNA editing of trnK-matK transcripts in barley plastids: support for MatK as an essential splice factor. *J. Mol. Biol.* **270** (2), 179–187 (1997).
35. Zoschke, R. et al. An organellar maturase associates with multiple group II introns. *Proceedings of the National Academy of Sciences*, 107(7), 3245–3250 (2010). (2010).
36. Qu, Y. et al. Ectopic transplastomic expression of a synthetic MatK gene leads to cotyledon-specific leaf variegation. *Front. Plant Sci.* **9**, 1453 (2018).
37. Hausner, G. et al. Origin and evolution of the chloroplast trnK (matK) intron: a model for evolution of group II intron RNA structures. *Mol. Biol. Evol.* **23** (2), 380–391 (2006).
38. Zeng, C., Jiao, Q., Jia, T. & Hu, X. Updated progress on group II intron splicing factors in plant chloroplasts. *Curr. Issues. Mol. Biol.* **44** (9), 4229–4239 (2022).
39. Liang, Y. et al. Maturase K forms a plastidial splicing complex with a neofunctionalized branching enzyme. *Nat. Commun.* 10.1038/s41467-026-70734-3 (2026).
40. Muino, J. M. et al. MatK impacts differential chloroplast translation by limiting spliced tRNA-K (UUU) abundance. *The Plant Journal*, 119(6), 2737–2752 (2024). (2024).
41. Serino, G. & Maliga, P. R. N. A. Polymerase Subunits Encoded by the Plastid rpo Genes Are Not Shared with the Nucleus-Encoded Plastid Enzyme. *Plant. Physiol.* **117**, 1165–1170 (1998).
42. Zhang, Y., Tian, L. & Lu, C. Chloroplast gene expression: Recent advances and perspectives. *Plant Communications*, **4**(5) (2023).
43. do Prado, P. F. et al. Structure of the multi-subunit chloroplast RNA polymerase. *Molecular Cell*, 84(5), 910–925 (2024). (2024).
44. Vergara-Cruces, Á. et al. Structure of the plant plastid-encoded RNA polymerase. *Cell* **187** (5), 1145–1159 (2024).
45. Wang, T. et al. Architecture of the spinach plastid-encoded RNA polymerase. *Nature Communications*, **15**(1), 9838 (20240.
46. Wu, X. X. et al. Cryo-EM structures of the plant plastid-encoded RNA polymerase. *Cell* **187** (5), 1127–1144 (2024).
47. Naver, H., Boudreau, E. & Rochaix, J. D. Functional studies of Ycf3: its role in assembly of photosystem I and interactions with some of its subunits. *Plant. Cell.* **13**, 2731–2745 (2001).
48. Nellaepalli, S., Ozawa, S. I., Kuroda, H. & Takahashi, Y. The photosystem I assembly apparatus consisting of Ycf3–Y3IP1 and Ycf4 modules. *Nat. Commun.* **9** (1), 2439 (2018).
49. Rolo, D., Schöttler, M. A., Sandoval-Ibáñez, O. & Bock, R. Structure, function, and assembly of photosystem I in thylakoid membranes of vascular plants. *Plant. Cell.* **36** (10), 4080–4108 (2024).
50. Lamb, J. R., Tugendreich, S. & Hieter, P. Tetratricopeptide repeat interactions: to TPR or not to TPR? *Trends Biochem. Sci.* **20** (7), 257–259 (1995).
51. Mattera, R., Molnar, T. & Struwe, L. *Cornus* × *elwinortonii* and *Cornus* × *rutgersensis* (Cornaceae), new names for two artificially produced hybrids of big-bracted dogwoods. PhytoKeys, 55, 93–111 (2015).
52. Park, H. S., Mohanan, P. & Yang, T. J. Super-Barcoding of Medicinal Materials based on Complete Plastid Genomes and Nuclear Ribosomal DNA. In Authentication of Chinese Medicinal Materials by DNA Technology: Techniques and Applications. 363–389 (2023).
53. Kim, K. et al. Comprehensive Survey of Genetic Diversity in Chloroplast Genomes and 45S nrDNAs within *Panax ginseng* Species. *PLoS One*. **10**, e0117159 (2015).
54. Kim, K. et al. Complete chloroplast and ribosomal sequences for 30 accessions elucidate evolution of *Oryza* AA genome species. *Sci. Rep.* **5**, 15655 (2015).
55. Kurtz, S. et al. Versatile and open software for comparing large genomes. *Genome Biol.* 5, R12; (2004). 10.1186/gb-2004-5-2-r12
56. Schwartz, S. et al. Human–mouse alignments with BLASTZ. *Genome Res.* **13**, 103–107 (2003).
57. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Preprint at arXiv:1303.3997v2 (2013).
58. Danecek, P. et al. Twelve years of SAMtools and BCFtools. *GigaScience*. 10(2): (2021). 10.1093/gigascience/giab008

59. R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria (2025). https://www.R-project.org/
60. Wickham, H. ggplot2: Elegant Graphics for Data Analysis, (2016). https://ggplot2.tidyverse.org (Springer- New York.
61. Wickham, H., François, R., Henry, L., Müller, K. & Vaughan, D. dplyr: A Grammar of Data Manipulation. R package version 1.1.4. (2023). https://CRAN.R-project.org/package=dplyr
62. Zeileis, A., Grothendieck, G. & zoo S3 Infrastructure for Regular and Irregular Time Series. *J. Stat. Softw.* **14** (6), 1–27. 10.18637/jss.v014.i06 (2005).
63. Pedersen, T. L. & Patchwork The Composer of Plots. CRAN: Contributed Packages. (2019). https://patchwork.data-imaginist.com
64. Tillich, M. et al. GeSeq–versatile and accurate annotation of organelle genomes. *Nucleic Acids Res.* **45**, W6–W11 (2017).
65. Carver, T., Harris, S. R., Berriman, M., Parkhill, J. & McQuillan, J. A. Artemis: An integrated platform for visualization and analysis of high-throughput sequence-based experimental data. *Bioinformatics* **28**, 464–469 (2012).
66. Greiner, S., Lehwark, P. & Bock, R. OrganellarGenomeDRAW (OGDRAW) version 1.3.1: Expanded toolkit for the graphical visualization of organellar genomes. *Nucleic Acids Res.* **47**, W59–W64 (2019).
67. Zheng, S., Poczai, P., Hyvönen, J., Tang, J. & Amiryousefi, A. Chloroplot: An Online Program for the Versatile Plotting of Organelle Genomes. *Front. Genet.* **11**, 576124 (2020).
68. Lagesen, K. et al. RNAmmer: Consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res.* **35**, 3100–3108 (2007).
69. Camacho, C. et al. BLAST+: Architecture and applications. *BMC Bioinform*. **10**, 1–9 (2009).
70. Varadi, M. et al. AlphaFold Protein Structure Database in 2024: providing structure coverage for over 214 million protein sequences. *Nucleic Acids Res.* **52** (D1), D368–D375 (2024).
71. Waterhouse, A. et al. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res.* **46** (W1), W296–W303 (2018).
72. Paysan-Lafosse, T. et al. InterPro in 2022. *Nucleic Acids Res.* **51** (D1), D418–D427 (2023).
73. Blum, M. et al. The InterPro protein families and domains database: 20 years on. *Nucleic Acids Res.* **49** (D1), D344–D354 (2021).
74. Katoh, K., Rozewicki, J. & Yamada, K. D. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Brief. Bioinform*. **20**, 1160–1166 (2019).
75. Minh, B. Q. et al. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* **37** (5), 1530–1534 (2020).
76. Ye, J. et al. Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC Bioinform.* **13**, 1–11 (2012).

Figures

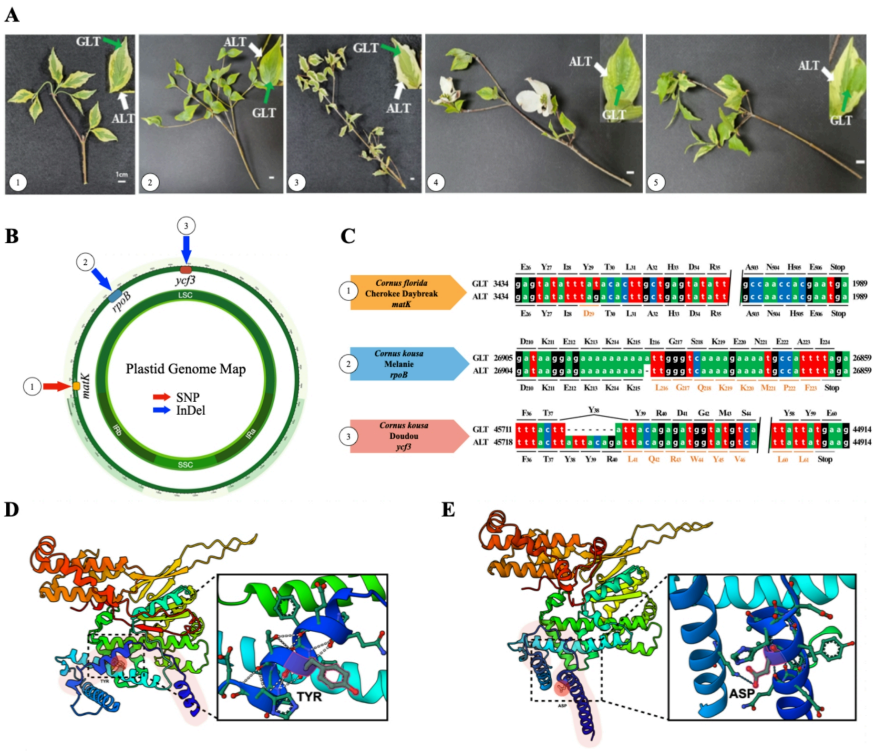

Figure 1

Chimeric plant materials, gene loci, sequence variations, and protein models between the plastomes of green (GLT) and albino leaf tissues (ALT) of *Cornus florida* Cherokee Daybreak and two *C. kousa* cultivars, Melanie and Doudou. (A) Samples and representative chimeric leaves of 1-*C. florida* Cherokee Daybreak, 2-*C. kousa* Melanie, 3-*C. kousa* Doudou, 4-*C. kousa* Rutban, 5-*C. kousa* Tom Wood; Inset (top right): zoomed image of a single chimeric leaf; (B) Representative circular plastome map showing the specific location of *matK* (1, orange), *rpoB* (2, light blue), and *ycf3* (3, light red). Arrows indicate the location, color the type of sequence variations between GLT and ALT in their plastome; SNP, single nucleotide polymorphism; InDel, Insertion/Deletion; (C) Partial DNA sequence alignment displaying variation and corresponding amino acid alteration between GLT and ALT in *matK*, *rpoB*, and *ycf3*, respectively. The numbers and colors are the same as in Fig. 1b; (D) *C. florida* Cherokee Daybreak *matK* amino acid sequence in GLT and (E) in ALT. A zoom of amino acid sequences with emphasis on position 29, TYR = tyrosine and ASP = aspartic acid, and their altered structural conformation.

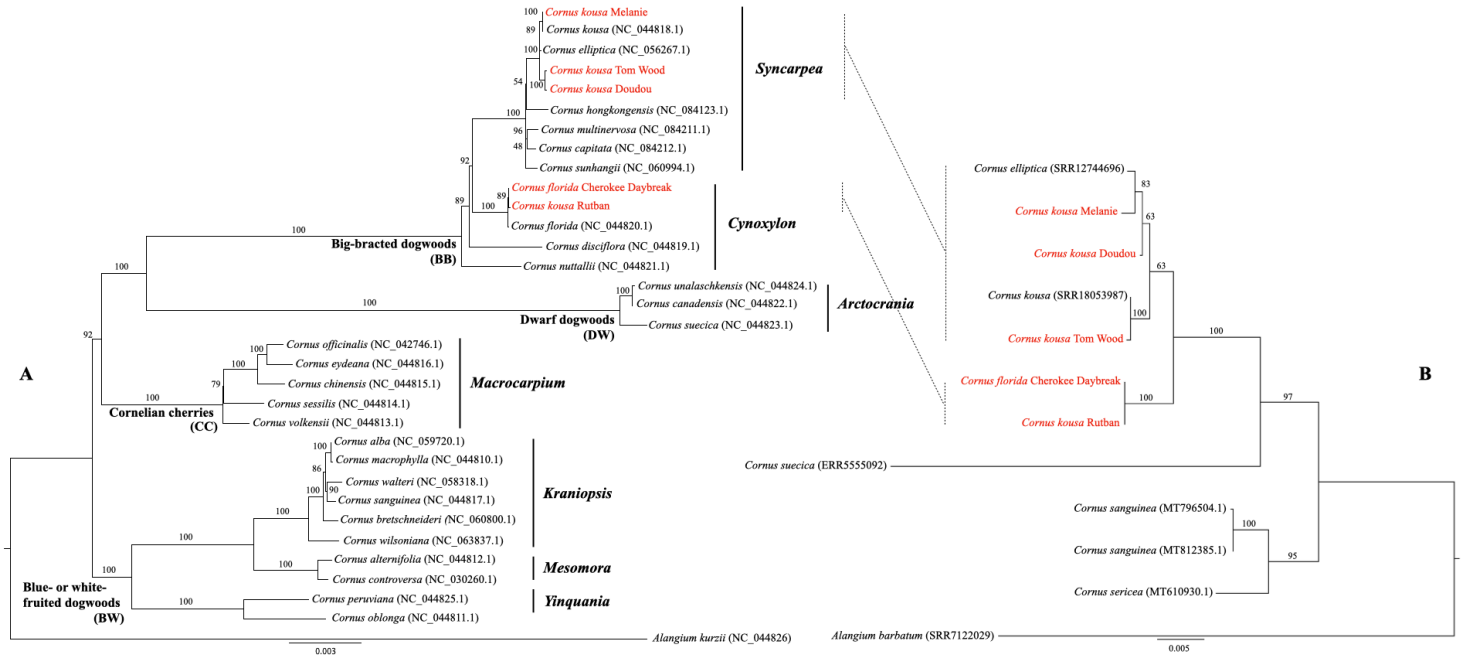

Figure 2

Plastome and 45S nrDNA-based trees of genus *Cornus*. (A) Phylogenetic trees were generated based on complete plastomes and (B) 45S nrDNA sequences. The numbers on the nodes indicate bootstrap support values of 1000 replications. Red, newly assembled.

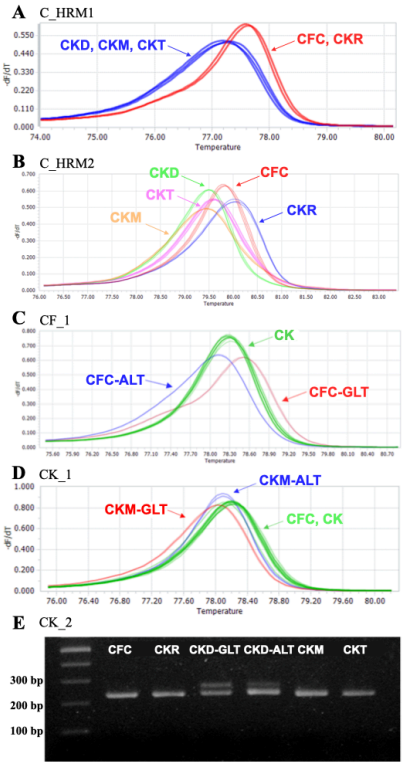

Figure 3

Marker authentication and validation of Cornus cultivars.

(A) C_HRM1 marker using *matK-rpo16* region; (B) C_HRM2 marker using *psbZ-rps14* region; (C) CF_1 marker using *matK* gene; (D) CK_1 marker using *rpoB* gene; (E) CK_2 marker using *ycf3* gene. A-D, normalized melting peaks of High-Resolution Melting (HRM) analysis; E, gel-based end-point PCR. Note: CFC, *C. florida* Cherokee Daybreak; CFC-GLT, *C. florida* Cherokee Daybreak green leaf tissue; CFC-ALT, *C. florida* Cherokee Daybreak albino leaf tissue; CK, all *Cornus* kousa cultivars; CKD, *C. kousa* Doudou; CKD-GLT, *C. kousa* Doudou green leaf tissue; CKD-ALT *C. kousa* Doudou albino leaf tissue; CKM, *C. kousa* Melanie; CKM-GLT, *C. kousa* Melanie green leaf tissue; CKM-ALT, *C. kousa* Melanie albino leaf tissue; CKT, *C. kousa* Tom Wood; CKR, *C. kousa* Rutban.

Supplementary Files

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

- SupplementaryInformationNatureScientificReportsCornus260511.docx