

Phylogenetic Structure and Differentiation Analysis of *Carpinus tientaiensis* Based on Chloroplast Genome SNPs: Postprint

Authors: Chen Moshun, Yang Zhongyi

Date: 2022-05-21T20:03:13+00:00

Abstract

Carpinus tientaiensis is an endangered plant endemic to China, discontinuously distributed only within Zhejiang Province with sparse populations, currently in a critically endangered status. This study investigates the phylogeographic structure and phylogenetic differentiation of *Carpinus tientaiensis* through cpDNA single nucleotide polymorphism (SNP) analysis of six natural populations (including maternal plants from all populations), to assess its endangered status and propose corresponding conservation strategies. Genomic DNA was extracted using the TIANGEN kit method, high-throughput sequencing was performed with Illumina NovaSeq 6000, complete chloroplast genome sequences were obtained, the online program OGDRAW was used to create chloroplast genome maps, DnaSP was used to analyze nucleotide diversity, PopART software was used to construct haplotype networks, RAxML software was used to construct ML trees, and MrBayes was used to construct Bayesian trees. The results showed: (1) Through analysis of the complete chloroplast genome sequences of *Carpinus tientaiensis*, most protein-coding genes and amino acid sequences exhibited obvious codon bias; 32 cp LTR forward repeats, 25 palindromic repeats, and 22 reverse repeats were detected; SSR repeats comprised 87 different types, most of which were A/T-rich, with mononucleotide repeats being the most numerous. (2) A total of 314 single nucleotide polymorphisms were identified in the chloroplast genome; single nucleotide substitutions revealed that *Carpinus tientaiensis* populations are monophyletic, divided into the Tiantai County population (THS) and Jingning County population (JST), with evolutionary relationships among population haplotypes displaying a star-like central radiation pattern. (3) Nucleotide diversity was low across all populations ($Pi < 0.005$), and haplotype diversity was low in both JST and THS populations ($Hd = 0.5-0.6$), indicating that *Carpinus tientaiensis* experienced local expansion after a historical bottleneck, with substantial genetic differentiation among populations and low genetic variation within populations coupled with

high differentiation levels among populations. Through research on chloroplast genome SNPs, this study reveals the genetic diversity and lineage differentiation of *Carpinus tientaiensis*, providing a theoretical basis for germplasm resource conservation and genetic rescue of this endangered plant.

Full Text

Preamble

Genealogical Structure and Differentiation Analysis of *Carpinus tientaiensis* Based on Chloroplast Genome SNPs

CHEN Moshun, YANG Zhongyi

School of Life Sciences, Taizhou University; Zhejiang Provincial Key Laboratory of Plant Evolutionary Ecology and Conservation, Taizhou 318000, Zhejiang, China

Abstract: *Carpinus tientaiensis* is an endangered plant species endemic to China, with sparse populations restricted to Zhejiang Province and currently in a critically endangered state. This study investigated the genealogical structure and systematic differentiation of *C. tientaiensis* through chloroplast DNA (cpDNA) single nucleotide polymorphism (SNP) analysis of six natural populations (including mother plants from all populations) to assess its endangerment status and propose conservation strategies. Genomic DNA was extracted using the TIANGEN kit method and subjected to high-throughput sequencing on the Illumina NovaSeq 6000 platform. Complete chloroplast genome sequences were obtained, annotated using the online program OGDRAW, analyzed for nucleotide diversity with DnaSP, and used to construct haplotype networks with PopART software. Phylogenetic trees were built using RAxML (maximum likelihood) and MrBayes (Bayesian inference). Results showed: (1) Analysis of the complete chloroplast genome revealed distinct codon preferences in most protein-coding genes and amino acid sequences, with 32 forward repeats, 25 palindromic repeats, and 22 reverse repeats detected as cpLTRs. A total of 87 SSR repeat sequences were identified, most of which were A/T-rich, with mononucleotide repeats being the most abundant. (2) Three hundred fourteen SNPs were identified in the chloroplast genome. Single-nucleotide substitutions indicated that *C. tientaiensis* populations are monophyletic, divided into the Tiantai County (THS) and Jingning County (JST) populations, with haplotype evolutionary relationships showing a star-like central radiation pattern. (3) All populations exhibited low nucleotide diversity variation ($P_i < 0.005$) and low haplotype diversity in both JST and THS populations ($H_d = 0.5-0.6$), suggesting that *C. tientaiensis* experienced local expansion after a historical bottleneck, with low genetic variation within populations but high differentiation between them. This study reveals the genetic diversity and lineage differentiation of *C. tientaiensis*, providing a theoretical basis for germplasm resource conservation and genetic rescue of this endangered species.

Keywords: *Carpinus tientaiensis*, chloroplast genome, single nucleotide polymorphism, genealogical structure, phylogeny

Funding: Supported by the Basic Public Welfare Research Project of Zhejiang Province (LY19C060001)

Corresponding Author: CHEN Moshun (1962-), Associate Professor, Bachelor's degree. Research focuses on plant ecology, population genetics, and conservation biology of endangered plants. Email: cmshoh@tzc.edu.cn

Introduction

Carpinus tientaiensis (Betulaceae) is a Tertiary relict plant species endemic to and endangered in China, designated as a second-class nationally protected wild plant (Zhang et al., 1993). China represents the distribution center for *Carpinus*, with approximately 33 species and 8 varieties (Fu, 2003). The southeastern coastal region is a major distribution area for *Carpinus*, with *C. tientaiensis* and *C. putoensis* being endemic to Zhejiang Province (Chen, 1994). The *Flora of China* documented *C. tientaiensis* as occurring in eastern Zhejiang (Editorial Committee of the Flora of China, 1979), while the *Flora of Zhejiang* reported only five remaining individuals (Zhang et al., 1993). Recent discoveries in Pan'an and Jingning counties have revealed disjunct distributions in Tiantai County, Pan'an County, Qingtian County, and Jingning She Autonomous County, Zhejiang Province, with fewer than 50 adult plants in the wild and a lack of juvenile individuals—falling below the threshold for stable wild population survival and placing the species in a critically endangered state. Habitat fragmentation reduces wild population sizes and increases isolation, leading to loss of genetic variation and inbreeding depression, ultimately increasing extinction risk (Aguilar et al., 2008; Wei et al., 2012). According to the IUCN Red List, *C. tientaiensis* is classified as Critically Endangered (CR).

Carpinus tientaiensis is a 14-ploid species ($2n = 14x = 112$), representing the polyploid with the highest ploidy level in Betulaceae (Chen et al., 2020). The species holds significant scientific value for research on Betulaceae taxonomy, paleofloristics, and endangerment mechanisms (Zhang et al., 1993; Wang & Ye, 2007).

Previous studies have examined tissue anatomy, photosynthetic characteristics in response to light intensity, and community features of *C. tientaiensis* (Chen et al., 2013, 2020). However, the evolutionary mechanisms and intraspecific phylogeny of *C. tientaiensis* populations remain unclear, necessitating population-level sampling and genomic analysis to deepen understanding of population dynamics at the genomic level. Chloroplast genomes, with their high gene content and conserved genomic structure, are valuable for studying maternal inheritance in flowering plants, particularly polyploids (Birky, 1995; Soltis & Soltis, 2000). Chloroplast DNA is typically uniparentally inherited and, due to

its self-replicating mechanism and relatively independent evolution, is frequently used to explore relationships among closely related species and within species (Yu et al., 2019). High-throughput sequencing of complete chloroplast genomes now provides abundant phylogenetic information, and comparative analysis of genomic structure, gene variation, and repeat sequence arrangement facilitates reconstruction of population genetic structure, historical dynamics, and lineage differentiation (Sun, 2012; Wang, 2015). This study employed high-throughput sequencing of *C. tientaiensis* chloroplast genomes to analyze SNPs, assess population genetic diversity, and infer genealogical structure and differentiation, thereby informing conservation and restoration strategies for this endangered species.

1.1 Field Survey and Sampling

Natural populations of *C. tientaiensis* occur at similar maximum elevations above 890 m, within a subtropical mountain humid climate zone, with acidic soil pH and high forest vegetation coverage characterized by subtropical evergreen broad-leaved forests (Chen et al., 2020). The species has a narrow distribution, with wild communities found only at Huading Mountain in Tiantai County, Dapan Mountain in Pan' an County, Gaomu Mountain and Qingmingjian in Pan' an County, Yangtianhu in Qingtian County, and Shangshantou in Jingning She Autonomous County. Specifically, 19 individuals coexist at Huading Mountain in Tiantai County, only one individual remains at Yangtianhu in Qingtian County, 18 individuals with diameter at breast height (DBH) >20 cm occur at Shangshantou in Jingning County, five individuals coexist at Dapan Mountain in Pan' an County, and one individual each occurs at Gaomu Mountain and Qingmingjian in Pan' an County. The distance between populations in Pan' an County is within 20 km, while the straight-line distance between the Tiantai County and Jingning County populations is approximately 250 km (Table 1).

Table 1. Basic conditions of *Carpinus tientaiensis* sampling sites

Sample ID	Distribution region	Latitude and longitude	Altitude	Sample size
JST_1-10	Shangshantou, Jingning County	119°37 E, 27°46 N	—	10
PDS_1-3	Dapan Mountain, Pan' an County	120°31 E, 28°58 N	—	3
PGS_1	Gaomu Mountain, Pan' an County	120°32 E, 28°54 N	—	1
PQJ_1	Qingmingjian, Pan' an County	120°28 E, 28°49 N	—	1
QYH_1	Yangtianhu, Qingtian County	119°59 E, 28°12 N	—	1

Sample ID	Distribution region	Latitude and longitude	Altitude	Sample size
THS_1-10	Huading Mountain, Tiantai County	121°05' E, 29°15' N	—	10

Materials were collected from 26 individuals across six natural populations (including mother plants from all populations) (Table 1). At the Jingning County Shangshantou plot (50 m × 15 m), nine individuals were sampled at 5–6 m intervals (average DBH 25.2 cm), plus one isolated tree with a maximum branch DBH of 11.1 cm. All five individuals from Dapan Mountain, Gaomu Mountain, and Qingmingjian in Pan'an County were collected (average DBH 20.41 cm). One individual was collected from Yangtianhu in Qingtian County (DBH 28.66 cm). At Huading Mountain in Tiantai County, seven individuals were collected from Ximaopeng (average DBH 43.08 cm) and three individuals near the highway (average DBH 20.07 cm). Fresh leaves were collected in the field, frozen in dry ice, cleaned, and stored at -80°C for subsequent experiments.

1.2 Chloroplast Genome Sequencing, Assembly, and Annotation

Chloroplast genomic DNA was extracted using the TIANGEN DNAsecure Plant Genomic DNA Kit (DP320). High-throughput sequencing was performed on the Illumina NovaSeq 6000 platform with 150 bp paired-end reads, generating at least 6 Gb of raw sequence data per sample. After filtering raw data to eliminate low-quality reads (Phred score cutoff = 30), high-quality data were obtained. Complete chloroplast genome sequences were determined for all 26 *C. tientaiensis* individuals. The published chloroplast genome sequence (accession: KY174338) was used as a reference (Yang et al., 2017). The DOGMA software was employed to annotate complete chloroplast genomes (Wyman et al., 2004), and the online program OGDRAW (<https://chlorobox.mpimp-golm.mpg.de/OGDraw.html>) was used to construct chloroplast genome maps. Basic information for each cp genome sequence was compiled, including genome size, gene features, and GC content.

1.3 Analysis of Repeat Sequences, IR Region Boundary Expansion/Contraction, and Codon Usage Bias

Dispersed repeats were analyzed using Reputer software, designated as cpLTRs. Simple sequence repeats (SSRs) were identified using the MISA tool (<http://pgrc.ipk-gatersleben.de/misa/>) with parameters: 1-10, 2-5, 3-4, 4-3, 5-3, 6-3. Comparative analysis of LSC, SSC, and IR lengths and boundaries was performed using the online tool IRscope (<https://irscope.shinyapps.io/irapp/>) to examine IR region expansion and contraction. Codon usage bias was analyzed and visualized using R software.

1.4 Phylogenetic Tree Construction

To examine regional phylogenetic relationships, phylogenetic analysis was conducted based on complete cpDNA sequences. Maximum likelihood (ML) trees were constructed using RAxML 8.0 software, and Bayesian inference trees were built using MrBayes 3.3. Haplotypes were developed from chloroplast genomes, and haplotype networks were constructed using PopART software (Kimura, 1980; Leigh & Bryant, 2015). Nucleotide diversity parameters were analyzed using DnaSP v6 (<http://www.ub.edu/dnasp/>). Molecular variance (AMOVA) was performed to assess molecular variation between lineages and calculate the genetic differentiation index (F_{ST}) to determine divergence levels.

Results

2.1 Basic Characteristics of *Carpinus tientaiensis* Chloroplast Genomes

Using the Illumina NovaSeq 6000 platform with 150 bp paired-end reads, an average of 30 million paired reads per sample were generated, from which complete cp genomes were determined for 26 *C. tientaiensis* individuals. The complete chloroplast genome consists of four parts, similar to other *Carpinus* species (Yang, 2019; Zhao et al., 2021). Genome length ranged from 159,281 to 159,841 bp (average: 159,616.2 bp), comprising a large single-copy (LSC) region (88,360-88,711 bp; average: 88,522.35 bp), a small single-copy (SSC) region (18,420-18,794 bp; average: 18,634.92 bp), and a pair of inverted repeats (IR) (each 26,067-26,451 bp; average: 26,229.46 bp) (Table 2, Figure 1).

Table 2. Characteristics of complete chloroplast genomes of *Carpinus tientaiensis*

Genome characteristic	Amount
Genome size (bp)	159,616.2 (average)
LSC length (bp)	88,522.35 (average)
SSC length (bp)	18,634.92 (average)
IR length (bp)	26,229.46 (average)
Protein-coding genes	86
tRNA genes	37
rRNA genes	8
GC content (%)	36.41 (average)

After data filtering, Q20 quality scores ranged from 96.46% to 97.21% (average: 96.98%), and Q30 scores ranged from 90.81% to 92.46% (average: 91.94%). The overall GC content averaged 36.41%, with LSC, SSC, and IR regions showing 34.20%, 30.09%, and 42.37% respectively. The IR region exhibited higher GC content than LSC and SSC regions. Among the 131 genes in the chloroplast genome, 86 were protein-coding genes, 37 were tRNA genes, and 8 were rRNA

genes. The LSC region contained 60 protein-coding and 22 tRNA genes; the SSC region contained 12 protein-coding and 1 tRNA gene; and the IR region contained 7 protein-coding genes, 7 tRNA genes, and all 4 rRNA genes in duplicate (Table 2, Figure 1).

2.3 Repeat Sequence Analysis in Chloroplast Genomes

Slipped-strand mispairing and inappropriate recombination of repeat sequences can cause sequence variation and DNA rearrangement (Wicke et al., 2011). Repeat sequences are important genetic markers closely related to species origin and evolution, generally classified as dispersed repeats and simple sequence repeats.

Dispersed repeats were analyzed using Reputer software, designated here as cpLTR long terminal repeats. In *C. tientaiensis* cp genomes, an average of 32 forward repeats, 25 palindromic repeats, and 22 reverse repeats were detected (Figure 2), with most repeats ranging from 10 to 38 bp in length. Simple sequence repeats (SSRs) are short tandem repeats of 1-6 bp widely distributed in chloroplast genomes. These cpSSRs are uniparentally inherited and play important roles in genome recombination and rearrangement, serving as effective molecular markers in population genetics and evolutionary studies (Zhou et al., 2018). Based on SSR distribution across chloroplast regions, *C. tientaiensis* averaged 48 SSRs in protein-coding genes, 3 in tRNA genes, and 41 in non-coding regions. Among protein-coding genes, six genes (*matK*, *atpA*, *rpoB*, *atpB*, *cemA*, *rpl2*) contained one tandem repeat, one gene (*rpoC2*) contained four tandem repeats, and one gene (*ycf1*) contained two tandem repeats. The LSC and SSC regions shared 25 *rps12* gene tandem repeats, while the IR region contained 10 *rps12* tandem repeats.

A total of 87 distinct SSR types were identified, repeated more than 10 times. Mononucleotide repeats were most abundant (average: 50, 57.47% of total), followed by tetranucleotide (14), dinucleotide (12), trinucleotide and compound nucleotides (4 each), pentanucleotide (3), with no hexanucleotide repeats detected (Figure 3). Most cpSSRs consisted of A/T repeats, rarely containing G/C tandem repeats. Mononucleotide repeats comprised 92% A/T bases, enriching the AT content of the cp genome. Among cpSSR base repeats, 10 repeats accounted for 31.20%, 11 repeats for 16.95%, 12 repeats for 34.44%, 13 repeats for 5.65%, and >14 repeats for 11.76% (Figure 4). SSR distribution across regions was 68.59% in LSC, 17.96% in SSC, and 6.58% each in IRA and IRB, with LSC containing the most repeats (Figure 5). The abundant base repeats and copy numbers observed in *C. tientaiensis* cpDNA provide valuable genetic information for population evolutionary studies.

2.4 Codon Usage Bias Analysis

Codons play a crucial role in genetic information transfer, linking nucleic acids and proteins. Statistical analysis of all protein-coding chloroplast DNA and

amino acid sequences revealed that protein-coding genes accounted for 65.65% of the *C. tientaiensis* chloroplast genome sequence. Codon frequency analysis showed higher frequencies for AUG, UUA, AGA, GCU, and UCU, and lower frequencies for CUG, GUG, AGC, CUC, and CUG (Figure 6). Among these, the most common amino acid in protein-coding genes was isoleucine (Ile), occurring 1,146 times. Relative synonymous codon usage (RSCU) analysis indicated that tryptophan (Trp) had an RSCU value of 1, showing no codon preference. Among all codons, 47.62% had $RSCU > 1$, with most (28/30, 93.33%) ending in A or T (U), while 50.79% had $RSCU < 1$, with most (30/32, 93.75%) ending in C or G (Figure 6).

In protein-coding chloroplast genes of *C. tientaiensis*, 20 amino acids were encoded by 63 codons, with most amino acids showing codon preference except aspartic acid (Asp). A total of 40 preferred codons were identified, involving 19 amino acids. Among preferred codons, 63.49% showed high preference (Figure 7). These results further demonstrate the relative conservation of the *C. tientaiensis* chloroplast genome, as high codon preference is common in higher plants (Yu & Han, 2021).

2.5 Chloroplast Genome Variation Detection

Single nucleotide polymorphisms (SNPs) and structural variation are crucial in evolution. Chloroplast genome structural variation includes insertions/deletions, transitions, transversions, and genomic rearrangements. A total of 314 SNPs were identified across 26 *C. tientaiensis* cpDNA sequences. Average SNP numbers were 132 in JST population, 126 in QYH, 18 in PDS, 18 in PGS, 8 in PQJ, and 12 in THS. When comparing population individuals to the THS_1 chloroplast genome, average base substitution numbers (Tv) were 95 in JST, 90 in QYH, 16 in PDS, 15 in PGS, 6 in PQJ, and 12 in THS (Table 3). All pairwise sequence comparisons showed more transversions than transitions, a pattern also observed in other taxa (Stoltzfus & Norris, 2016).

Table 3. Chloroplast comparative genome statistics for *Carpinus tientaiensis*

Population	SNPs	Single nucleotide polymorphism		Insertion (INS)	Deletion (DEL)	Total
		Transition	Transversion			
JST	132	—	95	—	—	—
QYH	126	—	90	—	—	—
PDS	18	—	16	—	—	—
PGS	18	—	15	—	—	—
PQJ	8	—	6	—	—	—
THS	12	—	12	—	—	—

2.6 IR Region Boundary Expansion and Contraction

Angiosperm cp genomes are highly conserved, with expansion and contraction of IR regions and single-copy (SC) boundary regions being the primary mechanisms causing length variation in higher plant cp genomes (Saina et al., 2018). Comparison of IR/SC boundary regions across 26 complete *C. tientaiensis* cp genomes revealed distinct differences at junction positions. The *rpl2* gene length was consistently 1,513 bp, with conserved endpoints 71 bp from both JLA and JLB intergenic regions. The *trnN-GUU* gene (72 bp) showed variable distances to JSB and JSA intergenic regions (1,516-1,864 bp) but did not exceed the endpoints of the following intergenic regions. The *rps19* gene extended 3 bp from JSB into IRb at the JSB/IRb boundary. The *ycf1* gene spanned the JSA/IRa region, with 4,211-4,553 bp located in the SSC region and extending 1,192-1,540 bp into IRb (Figure 8). Variation at these IR/SC boundaries contributed to overall cp genome length differences, a phenomenon also observed in other plants (Yin et al., 2018).

2.7.1 Haplotype Network Construction

Haplotype networks were constructed using variation in genetically linked nucleotide sequences (SNPs) from *C. tientaiensis* cpDNA. In phylogenetic analysis, each insertion or deletion was coded as a single evolutionary event. Most haplotypes were population-specific, with little sharing observed between populations. Geographically proximate populations showed identical haplotypes within groups: Tiantai County (THS) H1-H3, Jingning County (JST) H4-H6, and Pan' an County Dapan Mountain (PDS) H7-H9. Private haplotypes found in single specimens were H10 from Pan' an Qingmingjian (PQJ), H11 from Pan' an Gaomu Mountain (PGS), and H12 from Qingtian County (QYH). The haplotype network revealed two major groups: the eastern Zhejiang THS population and the western Zhejiang JST population, separated by 95 mutational steps. Analysis of polymorphic sites identified three nucleotide substitutions in the *rps16-trnQ* sequence that distinguished the groups: T→C (position 275), A→G (position 299), and T→G (position 1254). The network suggests a possible population history characterized by star-like central radiation, with most haplotypes differing by only 2-6 base mutations. THS haplotypes gave rise to PDS, PQJ, and PGS subgroups, with substitution links between THS and PDS haplotypes suggesting possible homology. JST haplotypes gave rise to the QYH subtype, with minimal differences (1-2 base mutations) between the two haplotypes, indicating local population expansion after a bottleneck event.

2.7.2 Phylogenetic Tree Reconstruction

Chloroplast genomes contain rich phylogenetic information and have been widely used for phylogenetic reconstruction at both inter- and intraspecific levels. Complete chloroplast genome data can resolve long-standing phylogenetic controversies. To assess phylogenetic relationships in *C. tientaiensis*, we performed phylogenetic analysis using complete chloroplast sequences from 26

individuals. Phylogenetic trees were constructed using maximum likelihood (RAxML) and Bayesian inference (MrBayes) methods, with node values representing ML bootstrap support (BS) and Bayesian posterior probability (PP) (Stamatakis, 2014; Xie et al., 2018).

Analysis of complete chloroplast genome sequences divided the six natural populations into THS and JST clades (100BS/100PP). The THS clade included: first branch THS_{10} and THS_2-3 (65BS/83PP); second branch PQJ_1 (61BS/60PP), PDS_1 (59BS/73PP), PGS_1 and PDS_2-3 (70BS/100PP), and THS_1 and THS_4-9 (70BS/100PP). The JST clade included: first branch JST_7 and JST_9-10 (89BS/100PP); second branch JST_1-6, JST_8, and QYH_1 (85BS/100PP) (Figure 10). Topologies from LSC and SSC datasets were consistent with the species chloroplast genome tree, showing only minor topological differences within intraspecific clades, supporting the formation of monophyletic groups.

2.7.3 Lineage Structure Analysis

Genetic diversity parameters were calculated using DnaSP, and AMOVA was performed to assess lineage differentiation. The fixation index (Fst) was analyzed to determine divergence levels. Nucleotide substitution and insertion/deletion variation revealed 12 cpDNA haplotypes (Nh) across six *C. tientaiensis* distribution areas, with 25% of these haplotypes found in single individuals and minimal sharing between populations. Lineage divergence was analyzed through haplotype diversity index (Hd) and nucleotide diversity index (Pi), where higher values indicate greater genetic diversity (Zhou, 2014; Nikulin et al., 2020). The PGS, PQJ, and YHS subgroups consisted of single individuals with only private haplotypes. Nucleotide diversity (Pi) was 0.00001 for JST, 0.00002 for THS, and 0.00003 for PDS populations, with all populations showing low variation (Pi < 0.005). Haplotype diversity (Hd) was 0.6 for JST and 0.511 for THS populations, indicating relatively low haplotype diversity. The PDS subgroup showed Hd = 1, likely due to its small sample size of only three individuals (Table 4).

Table 4. Statistical table of genealogical diversity parameters

Population	Number of samples	Number of haplotypes	Haplotype diversity (Hd)	Average	
				Polymorphic sites (S)	Nucleotide differences (Pi)

AMOVA of six *C. tientaiensis* populations revealed an Fst of 0.9709, indicating strong genetic differentiation between populations. This shows that 97.09% of total genetic variation occurred among populations, while only 2.91% occurred within populations, demonstrating greater genetic variation between than within

populations (Table 5). This pattern primarily results from habitat fragmentation and geographic isolation, which hinder gene flow between populations (Sun, 2012; Zheng, 2015; Nikulin et al., 2020).

Table 5. Statistical table of genealogical molecular variance test

Source of variation	Sum of squares	Variance squares	Percentage (%)
Among populations	—	—	97.09
Within populations	—	—	2.91
Total	—	—	100
Fixation index (Fst)	0.9709		

Discussion and Conclusion

Complete chloroplast genome sequences provide a rich source of phylogenetic information. The *C. tientaiensis* cpDNA averaged 159,616.2 bp in length with a GC content of 36.41%, indicating high conservation of the cp genome. Repeat analysis revealed an average of 32 forward repeats, 25 palindromic repeats, and 22 reverse repeats as LTRs, plus 87 SSR types. Most repeats were located in protein-coding regions, non-coding regions, and tRNA genes. LTR repeats are common in algal and angiosperm genomes, representing a major factor promoting cp genome rearrangement, with many rearrangement endpoints associated with such repeats (Pombert et al., 2005; Zhang et al., 2020). Across all individuals, SSRs typically consisted of A/T repeats, with mononucleotide repeats comprising 92% A/T bases. Most protein-coding genes showed strong codon preference, with 63.49% of preferred codons exhibiting high preference and the third position of chloroplast codons showing high A/T preference. Studies have shown that genomic AT content is related to repeat sequence dynamics and codon bias in chloroplast protein-coding genes (Yu et al., 2019; Wu et al., 2020).

SNP numbers and base substitutions in chloroplast genomes provide valuable markers for phylogenetic resolution among species (Zheng, 2015; Nikulin et al., 2020). According to coalescent theory, haplotype networks combined with geographic information can infer population origin and dispersal history (Huang et al., 2014). Analysis of SNP polymorphic sites and nucleotide variation revealed that *C. tientaiensis* is divided into Tiantai County (THS) and Jingning County (JST) populations, with geographically close subgroups clustering together. Except for the distant relationship between THS and JST haplotypes (95 base mutations), most haplotypes differed by only 2–6 base mutations. The clustering of haplotypes from the same or nearby geographic locations may result from the species' specific environmental requirements for subtropical humid or warm temperate climates (Chen, 1994). *Carpinus tientaiensis* exhibits population-specific haplotypes without sharing between populations, likely due to geographic isolation limiting gene flow over short periods (Nikulin et al., 2020). Low nucleotide diversity ($P_i < 0.005$) and low haplotype diversity in JST and THS populations ($H_d = 0.5\text{--}0.6$) suggest recent bottleneck effects (Sun, 2012; Zhou, 2014). Due

to climatic fluctuations and glacial cycles in the late Tertiary and Quaternary periods, *C. tientaiensis* likely retreated to small refugia or expanded locally into humid forest areas (Chen, 1994; Qi et al., 2012).

This study of *C. tientaiensis* genealogical structure and differentiation informs conservation and domestication strategies for populations with higher genetic diversity. Unique haplotypes were detected in all six natural populations, particularly in THS, JST, and PDS subgroups, which showed relatively high haplotype diversity. These populations occupy larger habitats with more stable internal environments and require enhanced habitat protection to maintain high genetic diversity. With small population sizes and high isolation levels, *C. tientaiensis* shows substantial genetic differentiation between populations and declining germplasm resources, making it an urgently endangered species requiring protection. For populations showing reproductive decline, genetic rescue should be implemented through pollen-mediated gene flow experiments, introducing new individuals or genotypes to mitigate genetic erosion and improve population viability.

References

AGUILAR R, QUESADA M, ASHWORTH L, et al., 2008. Genetic consequences of habitat fragmentation in plant populations: susceptible signals in plant traits and methodological approaches. *Mol Ecol*, 17(24): 5177-5188.

BIRKY CW, 1995. Uniparental inheritance of mitochondrial and chloroplast genes: mechanisms and evolution. *Proc Natl Acad Sci USA*, 92(25): 11331-11338.

BRITTON RJ, ROWEN L, WILLIAMS J, et al., 2003. Majority of divergence between closely related DNA samples is due to indels. *Proc Natl Acad Sci USA*, 100(8): 4661-4665.

CHEN MS, KE SS, 2013. Acclimation of anatomical structure and photosynthesis characteristics in leaves of *Carpinus tientaiensis* to irradiance. *Sci Silv Sin*, 49(2): 46-53.

CHEN MS, KE SS, JIN ZX, et al., 2020. *Conservation Biology of Carpinus tientaiensis*. Beijing: China Forestry Publishing House: 196-212.

CHEN ZD, 1994. Phylogeny and phytogeography of the Betulaceae (II). *Acta Phytotax Sin*, 32(2): 101-153.

Editorial Committee of the Flora of China of Chinese Academy of Science, 1979. *Flora Reipublicae Popularis Sinicae* (Vol. 21). Beijing: Science Press: 65-67.

FU LK, 2003. *Higher Plants of China: Vol. 4*. Qingdao: Qingdao Publishing House: 260-270.

HUANG DQ, LI QQ, ZHOU CJ, et al., 2014. Intraspecific differentiation of *Allium wallichii* (Amaryllidaceae) inferred from chloroplast DNA and internal

transcribed spacer fragments. *J Syst Evol*, 52(3): 341-354.

KIMURA M, 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J Mol Evol*, 16(2): 111-120.

LEIGH JW, BRYANT D, 2015. PopART: Full-feature software for haplotype network construction. *Methods Ecol Evol*, 6(9): 1110-1116.

NIKULIN AY, NIKULIN VY, GONTCHAROV AA, 2020. *Orostachys spinosa* (Crassulaceae) origin and diversification: East Asia or South Siberian Mountains? Chloroplast DNA data. *Plant Syst Evol*, 306(5): 84.

POMBERT JF, OTIS C, LEMIEUX C, et al., 2005. The chloroplast genome sequence of the green alga *Pseudendoclonium akinetum* (Ulvophyceae) reveals unusual structural features and new insights into the branching order of chlorophyte lineages. *Mol Biol Evol*, 22(9): 1903-1918.

QI XS, CHEN C, COMES HP, et al., 2012. Molecular data and ecological niche modelling reveal a highly dynamic evolutionary history of the East Asian Tertiary relict *Cercidiphyllum* (Cercidiphyllaceae). *New Phytol*, 196(2): 617-630.

SAINA JK, LI ZZ, GICHIRA AW, et al., 2018. The complete chloroplast genome sequence of tree of heaven *Ailanthus altissima* (Mill.) (Sapindales: Simaroubaceae), an important pantropical tree. *Int J Mol Sci*, 19(4): 929.

SOLTIS PM, SOLTIS DE, 2000. The role of genetic and genomic attributes in the success of polyploids. *Proc Natl Acad Sci USA*, 97(13): 7051-7057.

STAMATAKIS A, 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, 30(9): 1312-1313.

STOLTZFUS A, NORRIS RW, 2016. On the causes of evolutionary transition: transversion bias. *Mol Biol Evol*, 33(3): 595-602.

SUN Y, 2012. *Phylogeography and Population Genetics of the East Asia Endemic Genus Kirengeshoma* (Hydrangeaceae). Hangzhou: Zhejiang University.

WANG CT, YE CL, 2007. Endangering causes of endemic rare wild plants and conservation methods in Zhejiang Province. *J Fujian For Sci Technol*, 34(2): 202-204.

WANG JH, 2015. *Preliminary Study on Growth Performance and Molecular Phylogeography of Natural Populations in an Endangered Maple, Acer griseum* (Dicotyledoneae: Sapindaceae), Endemic to China. Beijing: Chinese Academy of Forestry Sciences.

WEI XZ, IANG MX, 2012. Limited genetic impacts of habitat fragmentation in an “old rare” relict tree, *Euptelea pleiospermum* (Eupteleaceae). *Plant Ecol*, 213(6): 909-917.

WICKE S, SCHNEEWEISS GM, DEPAMPHILIS CW, et al., 2011. The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. *Plant Mol Biol*, 76(3-5): 273-297.

WU ZH, LIAO R, YANG TG, et al., 2020. Analysis of six chloroplast genomes provides insight into the evolution of *Chrysosplenium* (Saxifragaceae). *BMC Genomics*, 21(1): 621.

WYMAN SK, JANSEN RK, BOORE JL, 2004. Automatic annotation of organelar genomes with DOGMA. *Bioinformatics*, 20(17): 3252-3255.

XIE DF, YU Y, DENG YQ, et al., 2018. Comparative analysis of the chloroplast genomes of the Chinese endemic genus *Urophysa* and their contribution to chloroplast phylogeny and adaptive evolution. *Int J Mol Sci*, 19(7): 1847.

YANG XY, 2019. *Phylogenetic Analysis of Betulaceae Plastomes*. Lanzhou: Lanzhou University: 47-52.

YANG YZ, WANG MC, LU ZQ, et al., 2017. Characterization of the complete chloroplast genome of *Carpinus tientaiensis*. *Conserv Genet Resour*, 9(2): 339-341.

YIN KQ, ZHANG Y, LI YJ, et al., 2018. Different natural selection pressures on the *atpF* gene in evergreen sclerophyllous and deciduous oak species: evidence from comparative analysis of the complete chloroplast genome of *Quercus aquifoloides* with other oak species. *Int J Mol Sci*, 19(4): 1042.

YU F, HAN M, 2021. Analysis of codon usage bias in the chloroplast genome of alfalfa (*Medicago sativa*). *Guizhou Sci*, 41(12): 2069-2076.

YU T, ZHANG YY, GAO J, et al., 2019. Complete chloroplast genome sequence of *Betula halophila*, a plant species with extremely small populations. *Sci Silv Sin*, 55(2): 41-49.

ZHANG QJ, LI W, LI K, et al., 2020. The chromosome-level reference genome of tea tree unveils recent bursts of non-autonomous LTR retrotransposons in driving genome size evolution. *Mol Plant*, 13(7): 935-938.

ZHANG SY, DING BY, 1993. *Flora of Zhejiang* (Volume General). Hangzhou: Zhejiang Science and Technology Publishing House: 245-251.

ZHAO RN, CHU XJ, LIU W, et al., 2021. Structure and variation analysis of chloroplast genomes in *Carpinus*. *J Nanjing For Univ (Nat Sci Ed)*, 45(2): 25-34.

ZHENG X, 2015. *Origin Area and Glacial Refugia: Chloroplast DNA Diversity in the Arctic-Alpine Plant Oxyria digyna* (Polygonaceae). Hefei: Anhui University.

ZHOU T, WANG J, JIA Y, et al., 2018. Comparative chloroplast genome analyses of species in *Gentiana* section *Cruciata* (Gentianaceae) and the development of authentication markers. *Int J Mol Sci*, 19(7): 1962.

ZHOU WY, 2014. *MtDNA Cytb Diversity of Siniperca scherzeri from Seven Water Systems*. Guangzhou: Ji'nan University: 41-46.

Note: Figure translations are in progress. See original paper for figures.

Source: ChinaXiv –Machine translation. Verify with original.