ChinaRxiv

AI translation · View original & related papers at
chinarxiv.org/items/chinaxiv-202204.00007

Comparative Analysis of Chloroplast Genomes of Three Hibiscus mutabilis Cultivars and Closely Related Species: A Postprint

Authors: Li Zhenbing, Ren Ting, Deng Jiaojiao, Chen Junpei, Zhou Songdong, Zeng Xinmei, Ma Jiao, Li Fangwen

Date: 2022-03-30T21:00:37Z

Abstract

Hibiscus mutabilis has a long cultivation history and is an ancient garden tree species and medicinal plant native to China. To investigate the evolutionary characteristics of H. mutabilis cultivars and their relatives, clarify the phylogenetic relationships among H. mutabilis cultivars and between them and their close relatives, and explore the inheritance pattern of the H. mutabilis chloroplast genome (chloroplast DNA, cpDNA), this study selected three H. mutabilis cultivars ('Danban Bai' , 'Jinqiu Song' , and 'Mudan Fen') from a hybrid combination and performed the first sequencing of their chloroplast genomes using the high-throughput Illumina NovaSeq platform. Following assembly and annotation, three complete cpDNA sequences were obtained. Combined with the cpDNA of the close relative H. taiwanensis already completed by our team and the cpDNA of H. syriacus and H. rosa-sinensis obtained from the gene bank, this study conducted a comparative analysis of cpDNA composition and structural characteristics for four Hibiscus species and three cultivars within H. mutabilis, and reconstructed their phylogenetic tree. The results showed that: (1) The chloroplast genome sequence lengths of 'Danban Bai' , 'Jinqiu Song' , and 'Mudan Fen' were 160,880, 160,879, and 160,920 bp, respectively, each containing 130 genes, including 85 protein-coding genes, 8 ribosomal RNAs, and 37 transfer RNAs. (2) Comparative analysis revealed that the three intraspecific cultivars of H. mutabilis and its close relative H. taiwanensis were highly conserved in chloroplast genome structure, with inverted repeat (IR) regions of 26,300 bp; H. syriacus and H. rosa-sinensis exhibited IR contraction, with lengths of 25,745 and 25,598 bp, respectively. (3) Phylogenetic analysis indicated that the three intraspecific cultivars clustered into a monophyletic clade, which then grouped with H. taiwanensis into a highly supported branch, demonstrating that H. mutabilis and H. taiwanensis are the closest relatives; compared with H. syriacus and H. rosa-sinensis, H. mutabilis and H. taiwanensis are more closely related to

H. hamabo, H. tiliaceus, and H. cannabinus. (4) The three H. mutabilis cultivars could be distinguished from each other through chloroplast genome sequences; in terms of large/small single-copy (LSC/SSC) region lengths, 'Danban Bai' , 'Jinqiu Song' , and 'Mudan Fen' were 89,355 bp/18,925 bp, 89,353 bp/18,926 bp, and 89,400 bp/18,920 bp, respectively, and candidate molecular markers and DNA barcodes were developed from repeat sequence and nucleotide diversity analyses, which can serve as molecular barcodes for cultivar identification. (5) The chloroplast genomes of H. mutabilis cultivars 'Danban Bai' and 'Jinqiu Song' showed the smallest differences and closest relationship, and based on their maternal-parent and offspring relationship, the maternal inheritance characteristic of the H. mutabilis chloroplast genome was demonstrated. This study contributes to a better understanding of the evolutionary characteristics of the chloroplast genomes of the three H. mutabilis cultivars and H. taiwanensis, as well as the phylogenetic relationships among species, and provides fundamental chloroplast genome data for accurate identification and superior cultivar breeding of H. mutabilis.

Full Text

Preamble

Comparative Analysis of Chloroplast Genomes of Three Hibiscus mutabilis Cultivars and Related Species

Authors: LI Zhenbing[1], REN Ting[1], DENG Jiaojiao[1], CHEN Junpei[1], ZHOU Songdong[1]*, ZENG Xinmei[2], MA Jiao[2], LI Fangwen[2]
[1]College of Life Sciences, Sichuan University, Key Laboratory of Bio-Resource and Eco-Environment of Ministry of Education, Chengdu 610065, China
[2]Chengdu Institute of Landscape Architecture, Chengdu 610083, China

Abstract: Hibiscus mutabilis, native to China with a long cultivation history, is an ancient garden tree species and medicinal plant. To investigate the evolutionary characteristics among Hibiscus cultivars and related species, clarify phylogenetic relationships between Hibiscus mutabilis cultivars and their close relatives, and explore the inheritance pattern of H. mutabilis chloroplast DNA (cpDNA), we selected three cultivars from a hybrid combination ('Danbanbai' , 'Jinqiusong' , and 'Mudanfen') and sequenced their chloroplast genomes for the first time using the Illumina NovaSeq high-throughput sequencing platform.

After assembly and annotation, three complete cpDNA sequences were obtained. Combined with the cpDNA of the related species H. taiwanensis from our research group and H. syriacus and H. rosa-sinensis from public databases, we conducted comparative analyses of cpDNA composition and structural characteristics across four Hibiscus species (including three cultivars of H. mutabilis) and reconstructed their phylogenetic tree. The results showed: (1) The chloroplast genomes of 'Danbanbai' , 'Jinqiusong' , and 'Mudanfen' were 160,880 bp, 160,879 bp, and 160,920 bp in length, respectively, each containing 130

genes (85 protein-coding genes, 8 rRNA genes, and 37 tRNA genes). (2) Comparative analysis revealed that the three H. mutabilis cultivars and the related species H. taiwanensis were highly conserved in chloroplast genome structure, with inverted repeat (IR) regions of 26,300 bp; H. syriacus and H. rosa-sinensis showed IR contraction at 25,745 bp and 25,598 bp, respectively. (3) Phylogenetic analysis indicated that the three cultivars formed a monophyletic clade that clustered with H. taiwanensis with high support, demonstrating the closest relationship between H. mutabilis and H. taiwanensis. Compared with H. syriacus and H. rosa-sinensis, H. mutabilis and H. taiwanensis showed closer affinity to H. hamabo, H. tiliaceum, and H. cannabinus. (4) The three H. mutabilis cultivars could be distinguished by chloroplast genome sequences. The LSC/SSC region lengths were 89,355 bp/18,925 bp for 'Danbanbai' , 89,353 bp/18,926 bp for 'Jinqiusong' , and 89,400 bp/18,920 bp for 'Mudanfen' . Candidate molecular markers and DNA barcodes developed from repeat sequence and nucleotide diversity analyses can serve as molecular tools for cultivar identification. (5) The chloroplast genomes of 'Danbanbai' and 'Jinqiusong' showed minimal divergence and the closest phylogenetic relationship. Based on their maternal-offspring relationship, we confirmed the maternal inheritance pattern of the Hibiscus chloroplast genome. This study enhances our understanding of chloroplast genome evolutionary characteristics and phylogenetic relationships among three H. mutabilis cultivars and H. taiwanensis, providing fundamental chloroplast genome data for accurate cultivar identification and breeding of superior varieties.

Keywords: Hibiscus mutabilis, Hibiscus taiwanensis, chloroplast genome, molecular marker, phylogenetic relationship

Funding: National Natural Science Foundation of China (32170209); Tissue Culture and Molecular Breeding of Hibiscus mutabilis in Chengdu Botanical Garden (18H0567); National Plant Herbarium Resource Bank—Digital and Sharing Platform of the Herbarium of Sichuan University (00204054A9016)

Corresponding Author: ZHOU Songdong, Ph.D., Associate Professor, Research on Plant Phylogeny and Molecular Evolution, (E-mail) zsd@scu.edu.cn

Introduction

Hibiscus mutabilis, also known as "Jushuanghua," belongs to the Malvaceae family and is native to China, with distribution across East and Southeast Asia. The species produces large, showy flowers with an extended blooming period and exhibits strong carbon sequestration, oxygen release, and cooling capacities, making it valuable for urban landscaping. The Chengdu Botanical Garden has long been dedicated to research and application of H. mutabilis as the city flower, developing numerous disease-resistant and pest-resistant varieties with different flowering periods. However, extensive artificial selection and natural hybridization have resulted in complex genetic relationships, unclear phyloge-

netic affinities, and ambiguous classification and evolutionary patterns among cultivars. Clarifying cultivar classification and phylogenetic relationships is crucial for facilitating inter-cultivar hybridization and new variety breeding.

Hibiscus taiwanensis, also called "Goutou Furong" or "Shan Furong," is another Hibiscus species endemic to Alishan, Taiwan. According to the Flora of China, H. taiwanensis closely resembles H. mutabilis, differing only in having coarse, stiff hairs and rough pubescence throughout the plant versus stellate tomentum in H. mutabilis. Some scholars have proposed that they may not be distinct species. Furthermore, H. mutabilis and H. taiwanensis hybridize readily with a fruit set rate of 57.89%, comparable to the 62.5% rate among H. mutabilis cultivars but substantially higher than the 8.33% rate for H. mutabilis × H. rosa-sinensis crosses and even lower success with H. syriacus, reflecting their close genetic affinity. Insufficient genetic information often limits our understanding of cultivated plants and their wild relatives, yet identifying genetic variation between them is essential for introducing favorable traits from wild relatives into cultivated varieties.

Chloroplast genomes (cpDNA) range from 75–250 kb and typically contain approximately 120 genes, including tRNA, rRNA, and protein-coding genes. Due to their simple structure, high conservation, and high copy number, cpDNA has been widely applied in molecular marker development and phylogenetic studies. Increasing evidence demonstrates that chloroplast genetic engineering offers significant advantages for plant genetic improvement, making chloroplasts a novel tool for plant transformation. Previous research on H. mutabilis has focused primarily on cultivation and breeding or chemical composition, with few studies addressing phylogenetic relationships among cultivars or between cultivars and related species. Recent progress includes chloroplast genome analysis of H. mutabilis and H. taiwanensis, though these studies have limitations. Abdullah et al. (2021) first sequenced and compared cpDNA of three Malvaceae species across different genera, while Xu et al. (2020) from our group reported the H. taiwanensis chloroplast genome. Zhang et al. (2021) examined pollen microstructure of some H. mutabilis cultivars using scanning electron microscopy and discussed its taxonomic significance. Although pollen morphology is highly stable and can reflect evolutionary relationships, genomic-level molecular data provide vastly more genetic information and, when combined with morphological studies, enable more precise and comprehensive phylogenetic and identification research. To date, limited genetic information is available for H. mutabilis cultivars, and no cpDNA studies have compared among cultivars or between cultivars and H. taiwanensis, severely hindering phylogenetic and trait improvement research. Since cpDNA is not always maternally inherited—being predominantly maternal in most angiosperms but paternal in most gymnosperms, with occasional biparental inheritance in some angiosperms—we selected three cultivars from a hybrid combination to conduct the first comparative chloroplast genome and phylogenetic analysis of H. mutabilis cultivars and their close relative H. taiwanensis.

This study addresses four scientific questions: (1) What are the evolutionary characteristics of chloroplast genomes among three H. mutabilis cultivars and H. taiwanensis? (2) What are the phylogenetic relationships between H. mutabilis and its related species? (3) Can molecular markers or DNA barcodes for cultivar identification be developed based on chloroplast genome composition and structure? (4) What is the inheritance pattern of H. mutabilis cpDNA? This research provides crucial firsthand genetic data for cultivar identification, evolutionary studies, genetic improvement, and breeding of superior H. mutabilis varieties.

Sample Collection and Morphological Characteristics

Three H. mutabilis cultivars were collected from the Chengdu Botanical Garden (104°8 11 E, 30°45 52 N): 'Danbanbai'(H. mutabilis cv. Danbanbai), 'Jinqiusong' (H. mutabilis cv. Jinqiusong), and 'Mudanfen' (H. mutabilis cv. Mudanfen). These three cultivars represent a hybrid combination, with 'Danbanbai' as the female parent, 'Mudanfen' as the male parent, and 'Jinqiusong' as the F1 generation, providing an ideal system for investigating chloroplast inheritance patterns. Ecologically, 'Danbanbai' and 'Mudanfen' are early-flowering cultivars (June–September), while 'Jinqiusong' is a mid-season cultivar (September–October). Morphologically, 'Jinqiusong' produces double flowers with degenerated stamens and sterile ovaries, whereas the other two cultivars are fertile. 'Danbanbai' has single white flowers, while 'Mudanfen' and 'Jinqiusong' are pink and red, respectively. 'Mudanfen' exhibits a peony-like flower form with a larger average diameter than the other two cultivars. The related species H. taiwanensis flowers from late October, produces fertile single flowers with white-pink petals, and shows minimal morphological differences from H. mutabilis, primarily in trichome type. Floral morphology is illustrated in Plate I.

Plate I. Morphological characters of flowers of three H. mutabilis cultivars and H. taiwanensis.
a. H. mutabilis cv. Danbanbai; b. H. mutabilis cv. Mudanfen; c. H. mutabilis cv. Jinqiusong; d. H. taiwanensis.

1.2 Genomic DNA Extraction and Sequencing

Fresh, healthy leaves from the three H. mutabilis cultivars were collected. Total genomic DNA was extracted from leaf tissue using a modified CTAB method. DNA quality and concentration were assessed using 1% agarose gel electrophoresis and a fluorescent dye assay (Quant-iT PicoGreen dsDNA Assay Kit). DNA libraries with 400 bp insert sizes were constructed following the Illumina TruSeq Nano DNA LT protocol and sequenced on the Illumina NovaSeq platform (paired-end, 2×150 bp). DNA extraction and sequencing were performed at Nanjing Personal Gene Technology Co., Ltd.

1.3 Chloroplast Genome Assembly and Annotation

Each species yielded at least 5 Gb of raw data. After quality filtering, NOVOPlasty software (k-mer = 39) was used for de novo assembly, with the rbcL gene of H. taiwanensis as the seed sequence. Chloroplast genome sequences were annotated using the online tool GeSeq and manually corrected with Geneious v9.0.2. The three annotated sequences were deposited in NCBI under accession numbers MZ846191 ('Danbanbai'), MZ846192 ('Jinqiusong'), and MZ855502 ('Mudanfen'). Physical maps were generated using Organellar Genome DRAW (OGDRAW). The H. taiwanensis sequence (MK937807) was also generated by our group (Xu et al., 2019). Additional chloroplast genome sequences used in this study were downloaded from NCBI (Table 1).

1.4 Chloroplast Genome Comparative Analysis

MISA (MicroSatellite identification tool) was used to identify simple sequence repeats (SSRs) in the three H. mutabilis cultivars, H. taiwanensis, H. syriacus, and H. rosa-sinensis. Parameters were set as follows: minimum repeat numbers of 10, 5, 4, 3, 3, and 3 for mono-, di-, tri-, tetra-, penta-, and hexa-nucleotide SSRs, respectively. REPuter was employed to analyze repeat types in the three H. mutabilis cultivars and H. taiwanensis (Hamming distance = 3, minimum repeat length = 30 bp). IRscope was used to visualize IR boundary expansion and contraction across the three H. mutabilis cultivars, H. taiwanensis, H. syriacus, and H. rosa-sinensis. Global sequence alignments were performed using mVISTA (Shuffle-LANGAN mode) to assess similarity, and nucleotide polymorphism was calculated using DnaSP6. Protein-coding sequences (CDS) were extracted using Geneious v9.02 (retaining only one copy for duplicated genes), and relative synonymous codon usage (RSCU) was calculated using CodonW.

1.5 Phylogenetic Analysis

Phylosuite software was used to extract CDSs from 17 chloroplast genomes. After removing duplicate and non-shared genes, sequences were aligned with MAFFT, optimized with MACSE, trimmed with Gblocks, and concatenated. Modeltest identified JC+I+G as the best-fit nucleotide substitution model. Bayesian inference (BI) was conducted using MrBayes with Markov chain Monte Carlo (MCMC) algorithm for 10 million generations, sampling every 1,000 generations and discarding the first 20% as burn-in. Maximum likelihood (ML) analysis was performed using IQtree with 1,000 bootstrap replicates. The final tree was visualized and annotated using iTOL.

Table 1. Names and GenBank accession numbers of plant species selected in this study

Species	Accession Number	Species	Accession Number
Hibiscus syriacus	KR259989	Thespesia populnea	NC053702

Species	Accession Number	Species	Accession Number
H. rosa-sinensis	NC048518	Abutilon theophrasti	NC045873
H. cannabinus	NC042239	Alcea rosea	NC053839
H. tiliaceum	NC053702	Malva verticillata	MT644160
Sida szechuensis	NC030195	Gossypium barbadense	NC051877
Ceiba speciosa (outgroup)	MK820674	G. thurberi	GU907100
H. hamabo	NC008641		

2.1 Basic Features of Chloroplast Genomes

As shown in Figure 1, 'Danbanbai' and 'Jinqiusong' showed minimal length difference (160,880 bp vs. 160,879 bp), differing by only one base pair, while 'Mudanfen' exhibited a more substantial difference at 160,920 bp. All three cpDNAs displayed the typical quadripartite circular structure, comprising a pair of inverted repeats (IRa and IRb, each 26,300 bp), a large single-copy region (LSC; 89,355 bp, 89,353 bp, and 89,400 bp, respectively), and a small single-copy region (SSC; 18,925 bp, 18,926 bp, and 18,920 bp, respectively). Table 2 shows that H. taiwanensis had a notably longer LSC region than the three H. mutabilis cultivars. Within cultivars, 'Danbanbai' and 'Jinqiusong' differed by only 1–2 bp in both LSC and SSC regions, while 'Mudanfen' showed greater divergence.

The H. mutabilis chloroplast genome contained 130 genes, including 85 protein-coding genes (PCGs), 37 tRNA genes, and 8 rRNA genes. Most genes occurred as single copies in either the LSC or SSC region. The SSC region housed 13 genes (12 PCGs: ndhF, rpl32, ccsA, ndhD, psaC, ndhE, ndhG, ndhI, ndhA, ndhH, rps15, ycf1; and 1 tRNA: trnL-UAG), while the LSC region contained 85 genes (63 PCGs and 22 tRNAs). Seventeen genes were duplicated in the IR regions: 6 PCGs (rpl2, rpl23, ycf2, ndhB, rps7, rps12), 7 tRNAs (trnI-CAU, trnL-CAA, trnV-GAC, trnI-GAU, trnA-UGC, trnR-ACG, trnN-GUU), and 4 rRNAs (rrn16, rrn23, rrn4.5, rrn5). The ycf1 gene spanned the SSC/IR boundary, while the rps12 gene had its first exon in the LSC region and the other two exons in the IR regions. Seventeen genes contained one intron, and ycf3 and clpP each contained two introns. All genes were categorized into three functional groups: expression-related, photosynthesis-related, and other genes (Table 3).

Figure 1. Physical map of the Hibiscus mutabilis chloroplast genome.
1. Photosystem I; 2. Photosystem II; 3. Cytochrome b/f complex; 4. ATP synthase; 5. NADH dehydrogenase; 6. RubisCO large subunit; 7. RNA polymerase; 8. Ribosomal proteins (SSU); 9. Ribosomal proteins (LSU); 10. Transfer RNAs; 11. Ribosomal RNAs; 12. clpP, matK; 13. Other genes; 14. Hypothetical chloroplast reading frames (ycf).

Table 2. Comparison of chloroplast genome characteristics among four Hibiscus species (including three H. mutabilis cultivars)

Genome Feature	Total Size (bp)/GC Content (%)	SSC Region (bp)/GC Content (%)	LSC Region (bp)/GC Content (%)	IR Region (bp)/GC Content (%)	Number of Genes
H. mutabilis cv. Danbanbai	160,880/36.9	18,925/31.5	89,355/34.7	26,300/42.6	130
H. mutabilis cv. Jinqiusong	160,879/36.9	18,926/31.5	89,353/34.7	26,300/42.6	130
H. mutabilis cv. Mudanfen	160,920/36.9	18,920/31.5	89,400/34.7	26,300/42.6	130
H. taiwanensis	161,056/36.9	18,918/31.5	89,538/34.7	26,300/42.6	130
H. rosa-sinensis	160,951/37.0	20,246/31.3	89,509/34.9	25,598/42.9	130
H. syriacus	161,022/36.8	19,831/31.1	89,701/34.7	25,745/42.8	130

Table 3. Gene content of the Hibiscus mutabilis chloroplast genome

Gene Category	Group	Gene Names	Number
Expression-related genes	Ribosomal large subunit	rpl2*, 23, 32, 22, 16*, 14, 36, 20, 33	9
	Ribosomal small subunit	rps7*, 15, 19, 3, 8, 11, 12*, 18, 4, 14, 2, 16*	12
	RNA polymerase	rpoA, rpoB, rpoC1*, rpoC2	4
	Ribosomal RNAs	rrn16*, rrn23*, rrn4.5*, *rrn5*	4

Gene Category	Group	Gene Names	Number
	Transfer RNAs	trnI-CAU, *trnL-CAA*, trnV-GAC, *trnA-UGC*, trnR-ACG, *trnI-GAU*, trnN-GUU, *trnL-UAG*, *trnP-UGG*, *trnW-CCA*, *trnM-CAU*, *trnV-UAC*, trnF-GAA, trnL-UAA, *trnT-UGU*, *trnS-GGA*, *trnfM-CAU*, *trnG-GCC*, *trnS-UGA*, *trnT-GGU*, *trnE-UUC*, *trnY-GUA*, *trnD-GUC*, *trnC-GCA*, *trnR-UCU*, *trnG-UCC*, trnS-GCU, trnQ-UUG, trnK-UUU*, trnH-GUG	30
Photosynthesis-related genes	Photosystem I	psaC, J, I, A, B	5
	Photosystem II	psbH, N, T, B, E, F, L, J, Z, C, D, M, I, K, A	15
	NADH dehydroge-nase	ndhB, *H, A*, I, G, E, D, F, C	9
	Cytochrome b/f complex	petD, *B*, G, L, A, N	6
	ATP synthase	atpB, E, I, H, F*, A	6

Gene Category	Group	Gene Names	Number
	RubisCO large subunit	rbcL	1
Other genes	Translation initiation factor	infA	1
	ATP-dependent protease	clpP**	1
	Maturase	matK	1
	Envelope membrane protein	cemA	1
	Acetyl-CoA carboxylase subunit	accD	1
	C-type cytochrome synthesis	ccsA	1
	Hypothetical reading frames (ycf)	ycf2*, ycf1, ycf4, ycf3**	4
Total			**130**

*One intron; **Two introns; *Duplicated gene

2.2 IR Boundary Analysis

Comparison of gene distribution at IR/LSC and IR/SSC boundaries among three H. mutabilis cultivars, H. taiwanensis, H. rosa-sinensis, and H. syriacus revealed IR expansion and contraction patterns. As shown in Figure 2, genes near these boundaries included rps19, rpl2, ycf1, ndhF, and trnH. The three H. mutabilis cultivars and H. taiwanensis showed identical IR boundary positions. The SSC/IRa boundary was located within the ycf1 gene across all six sequences, with ycf1 spanning 4,026 bp in the SSC region of H. mutabilis and H. taiwanensis, but 5,599 bp and 5,083 bp in H. rosa-sinensis and H. syriacus, respectively. The ndhF gene was located in the SSC region in all sequences, positioned 32 bp from the SSC/IRb boundary in H. mutabilis and H. taiwanensis, but 150 bp from the boundary in the other two species. Similarly, the rpl2 gene was located in the IRb region in all sequences, positioned 103 bp from the LSC/IRb boundary in H. mutabilis and H. taiwanensis, 67 bp in H. rosa-sinensis, and 113 bp in H. syriacus.

Figure 2. Comparison of junctions between LSC, SSC, and IR regions among Hibiscus taiwanensis, H. syriacus, H. rosa-sinensis, and three H. mutabilis cultivars.

2.3 Chloroplast Microsatellite and Repeat Sequence Analysis

Microsatellites (simple sequence repeats, SSRs) are 1–6 bp tandem repeats widely distributed in chloroplast genomes. SSRs exhibit high polymorphism and specificity, making them valuable markers for studying gene flow, population genetics, and genetic mapping. We analyzed six SSR types (mono-, di-, tri-, tetra-, penta-, and hexa-nucleotide) across six chloroplast genomes. As shown in Figure 3, mononucleotide repeats were most abundant (66.67–74.55% of total SSRs), exceeding the combined total of all other repeat types. The three H. mutabilis cultivars each contained 95 SSRs, nearly identical to H. taiwanensis (96 SSRs). By contrast, H. rosa-sinensis and H. syriacus showed substantial differences with 63 and 110 SSRs, respectively. All six SSR types were present in H. mutabilis cultivars and H. taiwanensis, while hexa-nucleotide repeats were absent in H. rosa-sinensis and H. syriacus. The only consistent difference between H. mutabilis cultivars and H. taiwanensis was the presence of one pentanucleotide repeat in the former versus two in the latter. Among the three cultivars, 'Mudanfen' was clearly distinct, possessing one additional mononucleotide repeat but one fewer hexa-nucleotide repeat than the other two cultivars.

Previous studies indicate that repeat regions represent hotspots for genome rearrangement and are valuable for developing genetic markers in population genetics. Four repeat types were detected: forward, reverse, complement, and palindromic. As shown in Figure 4, all three H. mutabilis cultivars contained all four repeat types, while H. taiwanensis lacked complement repeats. The three cultivars each had 22 forward and 3 reverse repeats, whereas H. taiwanensis had 26 forward and 1 reverse repeat. Although the proportions of repeat types were similar among the three cultivars, distinct differences existed. While all had 22 forward repeats, their composition varied: 'Mudanfen' had 16 repeats of 30–39 bp and 6 of 40–49 bp, while the other two cultivars had 17 and 5, respectively. 'Danbanbai' and 'Jinqiusong' each had 12 palindromic repeats versus 11 in 'Mudanfen' , and 'Danbanbai' and 'Mudanfen' each had 2 complement repeats versus 1 in 'Jinqiusong' .

Figure 3. SSR analysis of chloroplast genomes in Hibiscus taiwanensis, H. rosa-sinensis, H. syriacus, and three H. mutabilis cultivars.

Figure 4. Analysis of repeated sequences in chloroplast genomes of Hibiscus taiwanensis and three H. mutabilis cultivars.
A. Number of four repeat types; B. Number of repeat sequences by length.

2.4 Nucleotide Polymorphism Analysis

mVISTA was used to align chloroplast genomes of three H. mutabilis cultivars, H. taiwanensis, H. rosa-sinensis, and H. syriacus, with 'Danbanbai' as the reference and the latter two species as outgroups. Figure 5 clearly shows that H. rosa-sinensis and H. syriacus differed substantially from the reference, while 'Mudanfen' , 'Jinqiusong' , and H. taiwanensis showed much smaller differences. Highly divergent regions among the five cpDNAs were primarily located in intergenic spacers, though some protein-coding regions such as ycf1 also showed high variability. The four rRNA genes were most conserved, being identical across all three H. mutabilis cultivars and H. taiwanensis.

To precisely characterize nucleotide polymorphism among the three H. mutabilis cultivars and H. taiwanensis, we calculated nucleotide diversity values (Pi) for CDS and intergenic regions relative to the 'Danbanbai' reference. As shown in Figure 6, CDS sequences were highly conserved with generally low Pi values. For CDS regions, 'Jinqiusong' showed polymorphism only in ycf1, 'Mudanfen' only in ndhB, and H. taiwanensis in accD, atpA, clpP, ndhB, ndhD, matK, and ycf1, though maximum Pi values did not exceed 0.0017. In intergenic regions, H. taiwanensis exhibited variation in 14 spacers, with trnR-UCU~atpA, psbZ~trnG-GCC, infA~rps8, and ndhE~ndhG identified as hypervariable hotspots (Pi 0.00418). Three of these four highly variable regions were located in the LSC region and one in the SSC region, while IR regions showed low Pi values (<0.0209). 'Mudanfen' showed polymorphism in three intergenic spacers, with psbZ~trnG-GCC and ycf4~cemA as hotspot regions, both in the LSC region. 'Jinqiusong' showed no intergenic region differences. These highly variable regions can be used to design specific DNA barcodes.

Figure 5. Sequence identity plot of six chloroplast genomes (using Hibiscus mutabilis cv. Danbanbai as reference). The Y-axis represents percent identity ranging from 50% to 100%.

Figure 6. Comparative analysis of nucleotide variability among Hibiscus taiwanensis, H. mutabilis cv. Mudanfen, H. mutabilis cv. Jinqiusong, and H. mutabilis cv. Danbanbai. The X-axis represents protein-coding genes and intergenic regions; the Y-axis represents nucleotide diversity.

2.5 Selection Pressure and Codon Usage Bias Analysis

Relative synonymous codon usage (RSCU) values assess codon usage preferences in coding sequences, with higher values indicating stronger bias. We analyzed codon composition and RSCU across chloroplast genomes of three H. mutabilis cultivars, H. taiwanensis, and 13 related species. Leucine was most abundant, followed by isoleucine and glycine, while cysteine was least abundant, followed by tryptophan and serine. Except for tryptophan and methionine, all amino acids were encoded by two or more synonymous codons, with arginine and leucine each using six synonymous codons. Most codons with high RSCU values ended with A or U, consistent with previous findings of A/T-ending codon bias in

plants. Malvaceae species showed high conservation in codon preference, though some intergeneric differences existed. The three H. mutabilis cultivars and H. taiwanensis clustered tightly together, with other species grouping roughly by genus and tribe.

Figure 7. RSCU values of all protein-coding genes from chloroplast genomes of 15 species (including three cultivars). Red and white indicate higher and lower RSCU values, respectively. The right side shows the phylogenetic relationship among species.

2.6 Phylogenetic Relationships

We selected 16 Malvaceae chloroplast genomes to investigate phylogenetic relationships among H. mutabilis and related species, using Ceiba speciosa (Bombacaceae) as outgroup. ML and BI analyses were performed using 76 shared CDSs from 17 chloroplast genomes. Both methods produced identical topologies with high posterior probabilities and bootstrap values for most clades. As shown in Figure 8, the three H. mutabilis cultivars and H. taiwanensis formed a highly supported monophyletic group within Hibiscus, with maximum support values.

Figure 8. Phylogenetic relationships of 15 species (including three cultivars) inferred from ML and BI analyses based on 76 shared protein-coding genes. Numbers on branches represent bootstrap support values and posterior probabilities; * indicates maximum support in both analyses.

Discussion

3.1 Comparative Analysis of Hibiscus Chloroplast Genomes

Comparative analysis revealed that the three H. mutabilis cultivars and their close relative H. taiwanensis were highly conserved in cpDNA structure, with identical IR regions of 26,300 bp. In contrast, H. syriacus and H. rosa-sinensis showed IR contraction to 25,745 bp and 25,598 bp, respectively. The three H. mutabilis cultivars exhibited length variation in LSC/SSC regions: 89,355 bp/18,925 bp ('Danbanbai'), 89,353 bp/18,926 bp ('Jinqiusong'), and 89,400 bp/18,920 bp ('Mudanfen'), with 'Danbanbai' and 'Jinqiusong' showing the smallest differences. GC content is an important indicator of phylogenetic relationships, and the three H. mutabilis cultivars and H. taiwanensis showed identical total GC content and GC content in IR and SC regions, whereas H. syriacus and H. rosa-sinensis differed markedly. Additionally, the IR boundary genes (ycf1, trnH, ndhF, rpl2, rps19) showed no expansion or contraction among H. mutabilis cultivars and H. taiwanensis, while H. syriacus and H. rosa-sinensis exhibited obvious fluctuations. IR boundary variation is a primary cause of interspecific differences in chloroplast genomes. Thus, GC content and IR boundary patterns strongly support the close relationship between H. mutabilis and H. taiwanensis, while distinguishing them from H. syriacus and H. rosa-sinensis.

Synonymous codons arise through mutation, and evolutionary pressures alter their usage frequencies. Codon usage bias reflects long-term adaptation to base composition, tRNA abundance, and environmental selective pressures, influencing translation initiation, elongation, accuracy, mRNA splicing, and protein folding. Consequently, codon preference can reflect phylogenetic relationships. Our RSCU clustering grouped H. taiwanensis with the three H. mutabilis cultivars, with no difference between 'Jinqiusong'and 'Danbanbai'. 'Mudanfen'and H. taiwanensis showed slightly higher preference for leucine codon UUA compared to the other two cultivars, and H. taiwanensis additionally differed in valine and serine codon preferences. These patterns confirm the close relationship between H. mutabilis and H. taiwanensis, with 'Danbanbai' and 'Jinqiusong' being most similar. Studies of other terrestrial plants such as peanut, cherry, and radish have shown that chloroplast genome size, structure, gene content, and order are highly conserved between cultivated species and wild relatives. The high conservation observed between H. mutabilis cultivars and H. taiwanensis corroborates these findings. The minimal morphological differences between H. mutabilis and H. taiwanensis (primarily trichome type) contrast with greater morphological divergence from H. syriacus and H. rosa-sinensis. The high fruit set rate from H. mutabilis × H. taiwanensis crosses and low success with H. rosa-sinensis and H. syriacus further support these phylogenetic relationships. Pollen morphology studies by Zhang et al. (2021) also grouped 'Danbanbai' with 'Jinqiusong' separately from 'Mudanfen' . Our genomic results validate these morphological and breeding observations, demonstrating their utility for evolutionary and phylogenetic studies. Furthermore, since the three cultivars represent a hybrid combination with 'Danbanbai' as the maternal parent of 'Jinqiusong' , their high similarity confirms maternal inheritance of Hibiscus cpDNA, providing a theoretical foundation for breeding and genetic research.

3.2 Phylogenetic Relationships Among Hibiscus Species

Phylogenetic analysis strongly supported Hibiscus as a monophyletic genus, confirming its status as a natural taxonomic group. The three H. mutabilis cultivars formed a monophyletic clade with 99% bootstrap support, which then clustered with H. taiwanensis in a highly supported clade (100%), indicating their closest relationship. H. hamabo, H. tiliaceum, and H. cannabinus formed a monophyletic group that subsequently clustered with the H. mutabilis/H. taiwanensis clade, showing closer affinity than H. syriacus and H. rosa-sinensis. Beyond structural cpDNA features, phylogenetic analysis provides direct visualization of these relationships. In addition to confirming close relationships among H. mutabilis cultivars and with H. taiwanensis, the phylogeny showed that Hibiscus forms a monophyletic genus. In the taxonomic treatment by Feng (1984), Gossypium and Thespesia were placed in Hibisceae based on macromorphology. However, palynological studies by Peng et al. (2018) revealed significant pollen differences between these genera and Hibiscus, showing greater similarity to some Malveae species. Our phylogeny groups Gossypium and Thespesia with Malveae rather than Hibisceae, supporting the palynological evidence

and demonstrating the effectiveness of chloroplast genome data for phylogenetic inference.

3.3 Molecular Barcodes for Hibiscus Species and Cultivars

Candidate molecular markers and DNA barcodes developed from repeat sequence and nucleotide diversity analyses can serve as molecular tools for cultivar identification. We identified candidate species-level and cultivar-level cpDNA markers for Hibiscus that can distinguish species at the interspecific level and differentiate the three ornamental cultivars 'Danbanbai' , 'Jinqiusong' , and 'Mudanfen' .

Microsatellites and repeat sequences are abundant in cpDNA and useful for population genetics and marker development. The SSR profiles of H. mutabilis differed markedly from H. syriacus and H. rosa-sinensis but were similar to H. taiwanensis, and could not distinguish 'Danbanbai'from 'Jinqiusong'. In contrast, repeat sequence analysis clearly differentiated H. taiwanensis from H. mutabilis and distinguished among the three cultivars, making it more effective for cultivar identification. DNA barcodes are short, highly variable DNA sequences that enable rapid and accurate species identification. Through nucleotide polymorphism analysis, we identified chloroplast DNA barcodes between H. taiwanensis and H. mutabilis (trnR-UCU~atpA, psbZ~trnG-GCC, infA~rps8, ndhE~ndhG) and among H. mutabilis cultivars (psbZ~trnG-GCC, ycf4~cemA, ndhB, ycf1). Most hotspot regions were intergenic spacers, which evolve faster than coding regions under weaker selective pressure and are thus more suitable for low-level phylogenetic and evolutionary studies.

H. taiwanensis differed most from 'Danbanbai' , with variations in 7 genes and 14 intergenic spacers. 'Mudanfen' differed from 'Danbanbai' by only one gene (ndhB) and three intergenic spacers. The ndhB gene encodes an NADH dehydrogenase subunit, and its deletion severely reduces photosynthetic carbon assimilation capacity, though plants can maintain photoregulation through RNA editing. Overexpression of soybean ndhB in rice enhanced photosynthetic efficiency and altered agronomic traits, suggesting that variation in this gene may contribute to the larger flower diameter, greater flower number, and taller stature of 'Mudanfen' . 'Jinqiusong' showed the smallest difference from 'Danbanbai' , with only a single substitution in the ycf1 gene. Although ycf1 encodes a product of unknown function, knockout experiments in tobacco demonstrated its essential role in cell survival. With its rapid evolutionary rate, ycf1 serves as a useful molecular marker for resolving low-level phylogenetic relationships. Thus, Hibiscus cultivars and related species can be rapidly and accurately identified using specific intergenic spacers and particular genes (e.g., ycf1, ndhB).

References

ABDULLAH, MEHMOOD TW, FURRUKH M, et al., 2021. Correlations among oligonucleotide repeats, nucleotide substitutions, and insertion-deletion

mutations in chloroplast genomes of plant family Malvaceae[J]. J Syst Evol, 59(2): 388-402.

AMAR MH, 2020. ycf1-ndhF genes, the most promising plastid genomic barcode, sheds light on phylogeny at low taxonomic levels in Prunus persica[J]. J Genet Eng Biotechnol, 18(1): 531-547.

AMAR MH, MAGDY M, WANG L, et al., 2019. Peach chloroplast genome variation architecture and phylogenomic signatures of cp DNA introgression in Prunus[J]. Can J Plant Sci, 9(6): 885–896.

AMIRYOUSEFI A, HYVONEN J, POCZAI P, et al., 2018. IRscope: An online program to visualize the junction sites of chloroplast genomes[J]. Bioinformatics, 34(17): 3030–3031.

BEHURA SK, SEVERSON DW, 2012. Codon usage bias: Causative factors, quantification methods and genome-wide patterns: With emphasis on insect genomes[J]. Biol Rev, 88(1): 49–61.

BEIER S, THIEL T, MUNCH T, et al., 2017. MISA-web: a web server for microsatellite prediction[J]. Bioinformatics, 33(16): 2583–2585.

BULL LN, PABON-PENA CR, FREIMER NB, 1999. Compound microsatellite repeats: practical and theoretical features[J]. Genome Res, 9(9): 830–838.

CAI L, ZENG XM, WANG X, et al., 2021. Analysis and assessment of amino acid component in flowers of Hibiscus mutabilis L. among different cultivars[J]. Sci Technol Food Ind, 42(20): 279—285.

CHEN JH, HAO ZD, XU HB, et al., 2015. The complete chloroplast genome sequence of the relict woody plant Metasequoia glyptostroboides Hu et Cheng[J]. Front Plant Sci, 6(June): 447.

CHEN SL, SONG JY, YAO H, et al., 2009. Strategy and key technique of identification of Chinese herbal medicine using DNA barcoding[J]. Chin J Nat Med, 7(5): 322–327.

CHENG FT, LI ZH, LIU CY, et al., 2015. DNA barcoding of the genus Rehmannia (Scrophulariaceae)[J]. Plant Sci J, 33(1): 25-32.

CHO MS, YOON HS, KIM SC, 2018. Complete chloroplast genome of cultivated flowering cherry, Prunus x yedoensis 'Somei-yoshino' in comparison with wild Prunus yedoensis Matsum. (Rosaceae)[J]. Mol Breed, 38(9): 112.

DIERCKXSENS N, MARDULYN P, SMITS G, 2017. NOVOPlasty: De novo assembly of organelle genomes from whole genome data[J]. Nucl Acid Res, 45(4): e18.

DRESCHER A, RUF S, CALSA T, et al., 2000. The two largest chloroplast genome-encoded open reading frames of higher plants are essential genes[J]. Plant J, 22(2): 97–104.

FENG GM, 1984. Flora of China [M]. Beijing: Science Press & St. Louis: Missouri Botanical Garden Press, 49(2): 14-16.

GREINER S, LEHWARK P, BOCK R, 2019. Organellar Genome DRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes[J]. Nucl Acid Res, 47(W1): W59-W64.

HANSON G, COLLER J, 2017. Codon optimality, bias and usage in translation and mRNA decay[J]. Nat Rev Mol Cell Biol, 19(1): 20–30.

IVICA L, PEER B, 2021. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation[J]. Nucl Acid Res, 49(W1): 293-296.

KEARSE M, MOIR R, WILSON A, et al., 2012. Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data[J]. Bioinformatics, 28(12): 1647-1649.

KURTZ S, CHOUDHURI JV, OHLEBUSCH E, et al., 2001. REPuter: the manifold applications of repeat analysis on a genomic scale[J]. Nucl Acid Res, 29(22): 4633-4642.

LI HE, GUO QQ, ZHENG WL, 2018. Characterization of the complete chloroplast genomes of two sister species of Paeonia: genome structure and evolution[J]. Conserv Genet Resour, 10(2): 209-212.

LIBRADO P, ROZAS J, 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data[J], Bioinformatics, 25(11): 1451-1452.

LIM TK, 2014. Hibiscus taiwanensis[M]//LIM TK. Edible medicinal and non-medicinal plants. New Delhi: Springer Netherlands, 8: 381–384.

LIU J, SONG YJ, SUN JY, et al., 2019. Identification and analysis of RNA editing sites of ndhA, ndhB, ycf3, atpA and rps8 genes in sweet potato and its two wild species[J]. J Jiangsu Norm Univ(Nat Sci Ed), 37(2): 15-20.

MAYOR C, BRUDNO M, SCHWARTZ JR, et al., 2000. VISTA: visualizing global DNA sequence alignments of arbitrary length[J]. Bioinformatics, 16(11): 1046-1047.

MU LS, HE Y, LUO A, et al., 2017. Progress on application of plastid genetic engineering in plant breeding [J]. J Henan Agric Sci, 46(6): 1-12.

NEALE DB, SEDEROFF RR, 1989. Paternal inheritance of chloroplast DNA and maternal inheritance of mitochondrial DNA in loblolly pine[J]. Theor Appl Genet, 77(2): 212–216.

NOVOA EM, DE POUPLANA LR, 2012. Speeding with control: Codon usage, tRNAs, and ribosomes[J]. Trends Genet, 28(11): 574–581.

PENG HW, ZHOU SD, HE XJ, 2018. Pollen morphology of 26 Taxa from 15 genera of Malvaceae in China and its systematic significance[J]. Acta Bot Boreal-Occident Sin, 38(10): 1832-1845.

PENG J, ZHAO YL, DONG M, et al., 2021. Exploring the evolutionary characteristics between cultivated tea and its wild relateds using complete chloroplast genomes[J]. BMC Ecol Evol, 21(1): 71.

RAVI V, KHURANA JP, TYAGI AK, et al., 2008. An update on chloroplast genomes[J]. Plant Syst Evol, 271(1-2): 101-122.

REN GP, DONG YY, DANG YK, 2019. Codon codes: Codon usage bias influences many levels of gene expression[J]. Sci Sin Vit, 49(7): 839–847.

SHARP PM, LI WH, 1987. The codon adaptation index–a measure of directional synonymous codon usage bias, and its potential applications[J]. Nucl Acid Res, 15(3): 1281-1295.

SHI XQ, LI FW, LIU XL, et al., 2021. Study on Adaptability of Hibiscus mutabilis germplasm resources in Chengdu region[J]. J Sichuan For Sci Technol, 42(4): 90-93.

TILLICH M, LEHWARK P, PELLIZZER T, et al., 2017. GeSeq –versatile and accurate annotation of organelle genomes[J]. Nucl Acid Res, 45: W6-W11.

WANG J, LI Y, LI CJ, et al., 2019. Twelve complete chloroplast genomes of wild peanuts: great genetic resources and a better understanding of Arachis phylogeny[J]. BMC Plant Biol, 19(1): 504.

WANG Y, 2017. Research status and prospects of Hibiscus mutabilis[J]. J Sichuan For Sci Technol, 38(5): 124-127.

WANG Y, FENG LP, HUANG LL, et al., 2021. Rapid identification on chemical constituents of Hibiscus mutabilis flowers by UPLC-Q-Orbitrap HRMS[J]. Nat Prod Res Dev, 33(12): 2042-2052.

WANG XM, 2016. Agronomic traits and salt tolerance of soybean-NDHB subunit-overexpressed rice line[D]. Hangzhou: Zhejiang University.

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[J]. Int J Mol Sci, 19(7): 1847.

XU XR, ZHOU SD, SHI XQ, 2019. The complete chloroplast genome of Hibiscus taiwanensis (Malvaceae)[J]. Mitochondrial DNA Part B, 4(2): 2532-2533.

YANG QQ, JIANG M, WANG LQ, et al., 2019. Complete chloroplast genome of Allium chinense: comparative genomic and phylogenetic analysis[J]. Acta Pharm Sin, 54(1): 173-181.

YANG YZ, ZENG XM, MA J, et al., 2019. Observation and analysis of flowering characteristics of different early flowering cultivars of Hibiscus mutabilis Linn[J]. Mod Agric Sci Technol, (17):144-145.

ZHANG D, LI WX, JAKOVLIĆ I, et al., 2020. PhyloSuite: an integrated and scalable desktop platform for streamlined molecular sequence data management and evolutionary phylogenetics studies[J]. Mol Ecol Resour, 20(1): 348-355.

ZHANG L, ZHANG MY, ZHEN XM, et al., 2021. Pollen morphology of 19 cultivars of Hibiscus mutabilis in Chengdu and its taxonomic significance[J]. J Trop Subtrop Bot, 29(4): 421-429.

ZHENG P, SHI HW, DENG HB, et al., 2012. Study on the ecological functions of sixty-five garden species in Wuhan City, China[J]. Plant Sci J, 30(5): 468-475.

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

Source: ChinaXiv —Machine translation. Verify with original.