BMC Plant Biology

RESEARCH

Open Access

Comparative chloroplast genomics of *Hibiscus* (Malvaceae) and its phylogenetic implications

Abdullah[1,2*], Chengyu Chen[1,2], Rushan Yan[1,2], Abdul Sammad[1,2], Parviz Heidari[3], Tanveer Hussain[4], Ibrahim Hadiza Mohammed[5], Mohammad Tahir Waheed[6], Ibrar Ahmed[7,8,9], Atinga Vallada[10*] and Xiaoxuan Tian[1,2,11*]

Abstract

Hibiscus L. (Hibisceae, Malvoideae, Malvaceae) is globally distributed and valued for its ornamental, agricultural, and medicinal significance, yet its taxonomic problems remain unresolved. Here, we *de novo* assembled 17 complete chloroplast (cp.) genomes representing 16 *Hibiscus* species, as one species was represented by two varieties, and compared them with 13 previously published genomes of this genus. These genomes ranged from 159,125 to 163,360 base pairs and exhibited a conserved quadripartite structure, encoding 112 unique genes (78 protein-coding genes, 30 transfer RNAs, and 4 ribosomal RNAs). While codon usage, amino acid frequency, substitution patterns, and the expansion and contraction of inverted repeats revealed high similarity, the number of simple sequence repeats (SSRs) varied among species. Comparative analyses identified five non-coding polymorphic loci (*trnH-psbA*, *ndhF-rpl32*, *rps19-rpl22*, *trnR-atpA*, and *ndhD-ccsA*) and six coding loci (*rpl22*, *ycf1*, *matK*, *rpl32*, *rbcL*, and *ndhF*) as suitable candidates for DNA barcoding and phylogenetic reconstruction. Phylogenomic analyses based on the cp. genome confirmed the polyphyly of *Hibiscus*, with *Urena* Dill. ex L., *Malvaviscus* Fabr., *Kosteletzkya* C. Presl, and *Kydia* Roxb. nested within it. Three major clades were recovered (/Phylloglandula, /Trionum, and /Euhibiscus). *Hibiscus volkensii* Gürke showed a long branch, suggesting a distinct evolutionary trajectory. *Thepparatia scandens* (Roxb. ex G.Don) Phuph. appeared as a sister to *Hibiscus verdcourtii* Craven. Congruence was observed between the cp. genome phylogeny and recent sectional classifications, although *Hibiscus* sect. *Bombicella* was confirmed to be polyphyletic. Overall, this study provides insights into the cp. genome evolution of *Hibiscus* and its implications for the phylogeny of this genus. These findings highlight the need for nuclear genomic data to resolve remaining uncertainties and clarify the taxonomy of this complex genus and other genera within the tribe Hibisceae.

Keywords Taxonomy, Polymorphic loci, Evolution, Phylogenetic, *Bombicella*, *Lilibiscus*

*Correspondence:
Abdullah
abdullah@tjutcm.edu.cn; abd.ullah30@yahoo.com
Atinga Vallada
pharmvalat484@gsu.edu.ng
Xiaoxuan Tian
tian_xiaoxuan@tjutcm.edu.cn
[1]State Key Laboratory of Chinese Medicine Modernization, Tianjin University of Traditional Chinese Medicine, Tianjin 301617, China
[2]Haihe Laboratory of Modern Chinese Medicine, Tianjin 301617, China
[3]Faculty of Agriculture, Shahrood University of Technology, Shahrood, Iran
[4]Department of Biological Sciences, Virtual University of Pakistan, Islamabad, Pakistan
[5]Department of Pharmacognosy and Drug Development, Faculty of Pharmacy, Ahmadu Bello University Zaria, Kaduna 810107, Nigeria
[6]Department of Biochemistry, Faculty of Biological Sciences, Quaid-i-Azam University, Islamabad 45320, Pakistan
[7]Alpha Genomics Private Limited, Islamabad 45710, Pakistan
[8]Microbiological Analysis Team, Group for Biometrology, Korea Research Institute of Standards and Science (KRISS), Daejeon 34113, Republic of Korea
[9]Alpha Genomics Korea Co., Ltd. , Daejeon 34427, Republic of Korea
[10]Department of Pharmacognosy and Drug Development, Gombe State University, Gombe 760214, Nigeria
[11]State Key Laboratory of Dao-di Herbs, Beijing 100700, P. R. China

Introduction

The genus *Hibiscus* L. (Hibisceae, Malvoideae, Malvaceae) comprises approximately 438 species [1] and is widely distributed across tropical and temperate regions, occurring as herbs, shrubs, or trees [1–5]. *Hibiscus* species possess considerable economic and cultural significance, with prominent roles in horticulture, agriculture, and the food and pharmaceutical industries [5–7].

Taxonomic classification within *Hibiscus* has remained problematic across multiple hierarchical levels. At the species level, high morphological variation and occasional hybridization in certain lineages complicate species delimitation [8, 9]. At the sectional level, morphology-based subdivisions conflict with molecular phylogenies, resulting in inconsistent circumscriptions among studies [4–6, 10]. At the generic level, classification is challenging due to both complex morphological variation and molecular phylogenetic evidence: *Hibiscus* is polyphyletic, with several allied genera (e.g., *Abelmoschus* Medik., *Malvaviscus* Fabr., *Kosteletzkya* C. Presl, *Urena* Dill. ex L.) nested within it, indicating unresolved intergeneric relationships within tribe Hibisceae [4–6, 10, 11].

Earlier molecular phylogenetic work by Pfeil and Crisp [6] identified four deeply divergent clades—/Phylloglandula, /Trionum, /Euhibiscus, and /Calyphylli—revealing substantial phylogenetic structuring within *Hibiscus* that did not align with traditional classification. To address this polyphyly, Pfeil and Crisp [6] proposed several taxonomic solutions: (i) broadly expanding *Hibiscus* to include all embedded genera, (ii) splitting *Hibiscus* into multiple smaller, monophyletic genera, or (iii) accepting a paraphyletic *Hibiscus* as a practical compromise, also discussing a hybrid approach incorporating embedded genera within an enlarged *Hibiscus*. These proposals highlighted the complexity of reconciling morphological and molecular evidence and framed subsequent taxonomic debate. Recently, Hanes et al. [5] recognized 43 genera within tribe Hibisceae and substantially restricted *Hibiscus* largely to the /Euhibiscus lineage, reassigning many species to other genera and marking a significant shift toward monophyly. Following this revision, Barrett et al. [4] reinstated the genus *Sabdariffa* (DC.) Kostel., segregating it from *Hibiscus* sect. *Furcaria* DC. (a section within the /Phylloglandula clade). They recognized 117 species and 6 subspecies, further supporting a fully monophyletic circumscription of *Hibiscus*. These cascading taxonomic challenges—spanning species, sectional, and generic levels—underscore the continued need for broad phylogenomic evidence to stabilize the classification of *Hibiscus* and its allied genera.

The chloroplast (cp.) is a double-membrane-bound, self-replicating organelle responsible for photosynthesis. In most angiosperms, cp. genomes are maternally inherited, although paternal inheritance has been observed in certain gymnosperms [12–14]. The cp. genome typically consists of circular, double-stranded DNA ranging from 75 to 250 kb and encodes approximately 120 genes, including those for transfer RNAs (tRNAs), ribosomal RNAs (rRNAs), and protein-coding genes (CDS) [15]. It is generally organized into a conserved quadripartite structure composed of a large single-copy (LSC) region and a small single-copy (SSC) region, separated by a pair of inverted repeats (IRa and IRb) [12, 15–17]. In some lineages, the loss of one inverted repeat results in a single-copy genome structure [18].

Chloroplast genomes are evolutionarily dynamic and undergo a range of mutational events, including structural rearrangements, inversions, nucleotide substitutions, and insertions/deletions [12, 19]. These inherent polymorphisms make the cp. genome a valuable tool for phylogenetic and taxonomic research, species barcoding, population genetics, and conservation genetics [12, 20, 21].

Due to its predominantly uniparental inheritance and lack of meiotic recombination, the cp. genome evolves more slowly than its nuclear counterpart [12, 22]. This reduced evolutionary rate, coupled with high sequence conservation in coding regions, makes the cp. genome particularly suitable for resolving both deep phylogenetic relationships and fine-scale population-level divergence [15]. As a result, markers derived from whole cp. genome sequences are considered reliable, cost-effective, and informative for species identification and evolutionary inference [23, 24].

Although cp. genome studies exist for some *Hibiscus* species [25–27], most have focused on one or a few species (2–9 species), or on cultivar-level variation within a single species [19, 26–31]. Genus- and tribal-level phylogenomic analyses using complete cp. genomes are largely absent, with previous studies relying primarily on a few selected cp. or nuclear markers [4–6, 10, 11, 32]. To date, complete cp. genomes of only 13 *Hibiscus* species have been reported in the National Center for Biotechnology Information (NCBI). This limited sampling has prevented comprehensive analysis of evolutionary patterns, structural variations, and the identification of informative loci needed to resolve taxonomic problems in the genus. Broader sampling is particularly important because recent discoveries of synapomorphic inversions in *Triumfetta* Plum. ex L. (Malvaceae) [33, 34] demonstrate that broad cp. genomic comparisons can uncover previously undetected structural features with phylogenetic significance. Whether similar structural variations exist within *Hibiscus*, potentially reflecting its complex evolutionary history, or whether the genus instead maintains a conserved cp. genome architecture despite its

taxonomic complexity, remains unknown due to insufficient sampling.

The 29 *Hibiscus* species analyzed in this study address key systematic gaps. Our sampling (i) spans major clades of *Hibiscus* (e.g., /Phylloglandula, /Trionum, /Euhibiscus), with emphasis on poorly sampled /Euhibiscus lineages of *H.* sect. *Bombicella*, where sectional boundaries require further taxonomic treatment [4, 5]; (ii) includes 11 species with available whole-genome sequencing data, enabling *de novo* assembly of their cp. genomes for the first time and providing novel cp. genome resources, mostly from Australian *H.* sect. *Bombicella*; and (iii) comprises five species sequenced directly from herbarium specimens, including four members of *H.* sect. *Bombicella—Hibiscus aponeurus*, *Hibiscus denudatus*, *Hibiscus fuscus*, and *Hibiscus micranthus*—in which previous studies had detected polyphyly, allowing us to test whether these patterns reflect biological processes or limitations of prior molecular markers, as well as *Hibiscus volkensii*, which had not been included in earlier phylogenetic analyses. *De novo* assembly was employed to avoid potentially masking structural rearrangements, gene losses, or boundary shifts, thereby ensuring accurate reconstruction of cp. genomes across diverse lineages. The dataset is complemented with 13 previously published cp. genomes representing economically important and widely distributed species.

To address these gaps, we conducted a comparative analysis of cp. genomes from 29 *Hibiscus* species. We *de novo* assembled 17 complete cp. genomes representing 16 species (one species was represented by two varieties) and incorporated 13 previously published cp. genomes. The primary objectives were to: (i) investigate cp. genome evolution within *Hibiscus* through comparative analysis; (ii) identify polymorphic loci suitable for species barcoding and phylogenetic inference; and (iii) reconstruct phylogenetic relationships within *Hibiscus* and clarify its relationships with allied genera within tribe Hibisceae using complete cp. genome sequences.

Materials and methods

Materials collection, DNA extraction, and sequencing

Leaf tissue samples from five *Hibiscus* species were obtained from herbarium specimens housed at the Finnish Museum of Natural History (LUOMUS), University of Helsinki, Finland (Table 1). The corresponding voucher numbers were: *H. aponeurus* (C.192869), *H. denudatus* (C.192902), *H. fuscus* (C.192937), *H. micranthus* (C.192998), and *H. volkensii* (C.193229). Detailed specimen information is available through the LUOMUS herbarium database (e.g., http://id.luomus.fi/C.192869, with the voucher number changed accordingly for each species). Total genomic DNA was extracted using the DNeasy Plant Mini Kit (Qiagen, Germany), following the manufacturer's instructions with some modifications: samples were incubated in a water bath for 1 h (instead of 10 min), chilled on ice for 30 min (instead of 10 min), and eluted with 50 µL of elution buffer after a 30-minute

Table 1 Whole genome sequencing data and mean coverage depth for the *de novo* assembled chloroplast genomes of *Hibiscus* species

Species	Short Read Archive (accession)	Sequencing Data (GB)	WGS reads (millions)	Chloroplast reads (millions)	Coverage depth
Hibiscus brachysiphonius F.Muell.	SRR8666754	1.87	5.1	0.17	122
Hibiscus burtonii F.M.Bailey.	SRR8666554	1.7	4.68	0.15	120
Hibiscus campanulatus A.J.Perkins.	SRR8591790	1.95	5.32	0.25	194
Hibiscus goldsworthii F.Muell.	SRR8666785	0.98	2.7	0.23	176
Hibiscus leptocladus Benth.	SRR8666295	1.81	4.96	0.33	260
Hibiscus panduriformis Burm.f.	SRR8666286	1.97	5.36	0.06	46
Hibiscus sp. Gardneri	SRR8666190	1.76	4.84	0.27	208
Hibiscus sturtii var. campylochlamys Benth.[£]	SRR8666291	1.95	5.32	0.23	181
Hibiscus sturtii var. platychlamys Benth.[£]	SRR8666288	1.75	4.78	0.17	135
Hibiscus coatesii F.Muell.	SRR8666784	1.13	3.11	0.04	33
Hibiscus tridactylites Lindl.	ERR11606326	132	366	2.35	2193
Hibiscus richardsonii Sweet ex Lindl.	ERR12945461	144.5	398	1.09	1010
Hibiscus aponeurus Sprague & Hutch.	SRR33120790	8.96	21.54	0.49	461
Hibiscus denudatus Benth.	SRR33120789	9	21.62	0.58	539
Hibiscus fuscus Garcke	SRR33120788	9.53	22.90	0.52	486
Hibiscus volkensii Gürke	SRR33561777	5.89	14.16	0.19	174
Hibiscus micranthus L.f.	SRR33561776	8.36	20.08	0.58	543

The species shown in bold are *de novo* assembled from herbarium specimens in the current study

WGS whole genome sequencing, *GB* gigabytes

[£]This species appears twice because it includes two distinct varieties, each represented by an independent raw sequencing dataset

incubation at room temperature (instead of 5 min). The extracted DNA was sent to Novogene (Tianjin, China) for library preparation and sequencing. Libraries with an average insert size of 350 base pairs were constructed, and 150 base-pair (bp) paired-end sequencing was performed on an Illumina NovaSeq 6000 platform.

Chloroplast genome assembly and annotation

Whole-genome paired-end Illumina sequencing data (150 bp reads) were generated for five newly sequenced *Hibiscus* species. In addition, publicly available datasets for 12 further *Hibiscus* taxa were retrieved from the National Center for Biotechnology Information (NCBI) and the European Nucleotide Archive (ENA) (Table 1); these 12 taxa represent 11 species, as *H. sturtii* is represented by two varieties. Read quality filtering was performed using fastp [35] with the following parameters: -g -q 5 -u 50 -n 15 -l 150 --overlap_diff_limit 1 --overlap_diff_percent_limit 10. Paired-end reads were removed if either mate contained > 10% ambiguous bases (*N*), > 50% low-quality bases (Q ≤ 5), or showed adapter contamination as described by Huang et al. [36]. Only high-quality clean reads were retained for downstream analyses, with overall PHRED quality scores reaching 98% at Q20 and 96% at Q30. Chloroplast genomes of all taxa were *de novo* assembled from the filtered reads using GetOrganelle v1.7.5+ [37] with the parameters -R 20 embplant_pt -k 21,45,65,85,105. All assemblies were subsequently validated, and coverage depth was assessed by mapping the raw reads of each taxon to its respective *de novo* assembled cp. genome using Geneious R8.1 [38] with low-sensitivity, fast-mapping settings. Chloroplast genome annotation was carried out using GeSeq v2.0.3 [39] with default settings for circular cp. genomes, including annotation of the IR regions and the *rps12* gene. Transfer RNA genes were further validated using tRNAscan-SE v2.0.7 [40] and ARAGORN v1.2.38 [41], both also run under default settings. Protein-coding gene boundaries were manually visualized and curated by aligning each genome through MAFFT v7 [42] with *Hibiscus rosa-sinensis* and *Hibiscus syriacus* cp. genome sequences in Geneious R8.1 [38]. Final annotated genomes were prepared for submission to NCBI using GB2Sequin v1 [43], which implements tbl2asn v25.8 for Sequin file generation. The raw sequencing data were submitted to NCBI under BioProject PRJNA1196377 for reproducibility of results.

Comparative genomic analysis of *Hibiscus* species

The comparative dataset included 17 de novo assembled cp. genomes of *Hibiscus* taxa (representing 16 species) and 13 previously published cp. genomes retrieved from NCBI. Accession numbers for raw sequencing reads, either generated in this study or retrieved from NCBI and ENA, are provided in Table 1, and accession numbers for the assembled cp. genomes, along with those retrieved from NCBI, are listed in Table 2. All newly assembled genomes are available in Supplementary Data 1. The genetic features of the cp. genomes were analyzed using Geneious R8.1, and circular genome maps were generated using Chloroplot v.1 [44]. Gene arrangement patterns were visualized through progressive Mauve alignment v.2.4.0 integrated in Geneious R8.1 [45]. IR boundary dynamics were examined using IRscope [46]. Codon usage and amino acid frequencies were calculated with custom Python scripts (Scripts 1 and 2). These are now part of CGAS (Chloroplast Genome Analysis Suite) [47]. The microsatellite repeats were identified using MIcroSAtellite (MISA) [48], applying minimum thresholds of 10 repeat units for mononucleotides, 5 for dinucleotides, 4 for trinucleotides, and 3 for tetra-, penta-, and hexanucleotides.

Synonymous and non-synonymous substitution analyses were performed using the KaKs_Calculator v2 [49] with *H. rosa-sinensis* as the reference. The Ka/Ks ratios were interpreted based on standard evolutionary models: values < 1 indicate purifying selection, = 1 suggest neutral evolution, and > 1 reflect positive selection. Substitution patterns and transition-transversion ratios were determined by pairwise alignment of each genome to *H. rosa-sinensis* using MAFFT v7 [42] integrated in Geneious R8.1, followed by analysis with a custom Python script (Script 3). Genetic divergence among *Hibiscus* species was assessed through nucleotide diversity (π) calculations using CPStool v1 [36].

Phylogenetic analysis

To infer phylogenetic relationships among *Hibiscus* species and related genera within the tribe Hibisceae, we used the cp. genome of *Bombax ceiba* L. (NCBI accession: OR832744) from the subfamily Bombacoideae as an outgroup, following previous studies [5, 26]. This species was chosen because Bombacoideae is sister to Malvoideae [19, 26, 50], providing a stable root for the Malvoideae clade. After several taxonomic revisions, Hibisceae has been resolved as monophyletic, but discrepancies remain regarding the number and boundaries of genera within the tribe. Therefore, using *B. ceiba* as an outgroup helps minimize potential rooting issues. Details about Hibisceae phylogeny can be found in Hanes et al. [5]. The species names of related genera and their accessions are given in Table S1. Coding sequences were extracted and concatenated using Geneious R8.1. Multiple sequence alignment was performed using MAFFT v7 [42] and manually curated, resulting in a final dataset of 69,636 nucleotide sites, including 2,120 parsimony-informative sites and 1,773 distinct site patterns. Maximum likelihood (ML) phylogenies were reconstructed using IQ-TREE v3 [51]. The optimal substitution model

Table 2 Chloroplast genome features of *Hibiscus*

Species	Length (base pairs)				Guanine-cytosine content (%)				Guanine-cytosine content (%)			
	Complete	LSC	SSC	IR	Complete	LSC	SSC	IR	tRNA	rRNA	CDS	Accession Number
Hibiscus aponeurus	159,125	88,252	20,011	25,431	37.13	35.04	31.6	42.92	52.9	55.49	38.15	PV666612
Hibiscus brachysiphonius	159,875	89,377	19,972	25,263	36.96	34.77	31.51	42.98	52.94	55.45	38.12	BK064974
Hibiscus burtonii	160,326	89,690	20,132	25,252	36.89	34.7	31.27	43.01	52.97	55.43	38.1	BK064982
Hibiscus campanulatus	160,245	89,478	20,255	25,256	36.87	34.71	31.17	43	52.97	55.43	38.11	BK064981
Hibiscus cannabinus L.	163,360	90,800	19,534	26,513	36.57	34.27	30.87	42.61	53.15	55.45	38.03	MW446503
Hibiscus coatesii	160,148	89,486	20,154	25,254	36.92	34.76	31.33	42.99	52.97	55.45	38.13	BK068671
Hibiscus coccineus Walter	160,280	89,121	18,673	26,243	36.92	34.71	31.58	42.58	53.17	55.45	38.11	OK336487
Hibiscus denudatus	159,952	89,382	19,982	25,294	36.95	34.81	31.45	42.91	52.88	55.45	38.16	PV666613
Hibiscus fuscus	159,242	88,319	20,005	25,459	37.11	34.98	31.65	42.95	52.81	55.49	38.2	PV666614
Hibiscus goldsworthii	160,374	89,522	20,300	25,276	36.83	34.67	31.15	42.93	52.81	55.45	38.11	BK064980
Hibiscus hamabo Siebold & Zucc.	161,729	89,217	19,570	26,471	36.9	34.79	30.89	42.68	53.24	55.47	38.11	NC_030195
Hibiscus leptocladus	159,904	89,433	19,971	25,250	36.93	34.75	31.33	42.99	52.97	55.45	38.13	BK064978
Hibiscus micranthus	159,661	88,477	19,974	25,605	37.14	35.01	31.76	42.91	52.84	55.49	38.22	PV666615
Hibiscus moscheutos L.	160,208	89,087	18,647	26,237	36.93	34.72	31.57	42.58	53.25	55.46	38.1	NC_068824
Hibiscus mutabilis L.	160,879	89,353	18,926	26,300	36.92	34.73	31.46	42.6	53.2	55.4	38.1	MK820657
Hibiscus panduriformis	161,188	89,728	18,870	26,295	36.87	34.65	31.47	42.61	52.84	55.45	38.11	BK064979
Hibiscus richardsonii	160,988	89,672	19,014	26,151	36.83	34.58	31.39	42.68	52.84	55.4	38.06	BK071661
Hibiscus rosa-sinensis L.	160,751	89,277	20,246	25,614	37.03	34.93	31.34	42.95	53.21	55.47	38.21	ON641322
Hibiscus sabdariffa L.	162,428	90,327	19,617	26,242	36.74	34.46	31.37	42.66	53.13	55.47	38.06	MZ522720
Hibiscus sinosyriacus L.H.Bailey	160,892	89,747	19,661	25,742	36.85	34.66	31.32	42.77	53.09	55.45	38.14	MZ367751
Hibiscus sp. Gardneri	160,234	89,552	20,174	25,254	36.9	34.74	31.31	42.99	52.97	55.45	38.12	BK064977
Hibiscus sturtii var. campylochlamys	160,239	89,432	20,227	25,290	36.8	34.63	31.12	42.91	52.84	55.45	38.08	BK064976
Hibiscus sturtii var. platychlamys	160,569	89,773	20,214	25,291	36.77	34.57	31.16	42.92	52.84	55.45	38.1	BK064975
Hibiscus syriacus L.	161,022	89,835	19,831	25,678	36.83	34.64	31.11	42.87	52.84	55.45	38.13	KM103366
Hibiscus taiwanensis S.Y.Hu	161,056	89,538	20,680	25,419	36.89	34.67	31.68	42.92	53.19	55.4	38.11	MK937807
Hibiscus tiliaceus L.	161,863	89,331	19,572	26,480	36.89	34.8	30.88	42.64	53.18	55.47	38.06	NC_053627
Hibiscus tridactylites	160,535	89,277	18,954	26,152	36.9	34.67	31.42	42.68	52.84	55.4	38.07	BK071662
Hibiscus trionum L.	160,530	89,272	18,954	26,152	36.9	34.68	31.42	42.68	53.17	55.41	38.08	OL628829
Hibiscus verdcourtii Craven	160,990	89,195	19,029	26,383	36.78	34.57	31.33	42.47	52.88	55.43	38.04	OZ224470
Hibiscus volkensii	161,217	89,020	20,771	25,713	36.93	34.89	31.07	42.82	52.84	55.5	38.12	PV666616

(TVM+F+I+R3) was selected by ModelFinder [52] under the Bayesian Information Criterion. Branch support was assessed using 10,000 ultrafast bootstrap replicates (-bb 10,000) and 10,000 Shimodaira–Hasegawa approximate likelihood ratio test replicates (SH-aLRT; -alrt 10,000) [53]. Topological accuracy was further improved with bootstrap-guided nearest-neighbor interchange optimization (-bnni), and computational efficiency was enhanced with automatic thread detection (-nt AUTO). Final trees were visualized and annotated using iTOL v4 [54].

Results

Sequencing and *de novo* assembly of cp. genomes

For five newly sequenced *Hibiscus* species, Illumina NovaSeq 6000 generated 5.89–9.53 Gb (gigabytes) of raw data per species, containing 14.16–22.9 million 150-bp paired-end reads. High-quality cp. assemblies were achieved with 174–543× average coverage depths, corresponding to 0.19–0.58 million cp. genome reads.

For 12 additional taxa (representing 11 species), publicly available whole-genome sequencing (WGS) datasets (0.98–144.5 Gb; corresponding to 2.7–398 million paired-end reads) enabled successful cp. genome assembly. The resulting assemblies showed 33–2193× coverage depths, supported by 0.04–2.35 million cp-derived reads (Table 1). The coverage depth maps of these cp. genomes are provided in Fig. S1.

Chloroplast genome features of *Hibiscus* species

The cp. genomes of 29 *Hibiscus* species exhibited a conserved quadripartite organization and identical gene arrangement (Fig. 1; Fig. S2). Genome sizes ranged from 159,125 to 163,360 bp, comprising LSC (88,252–90,800 bp) and SSC (18,647–20,771 bp) regions separated by IRs (25,250–26,513 bp) (Table 2).

Each genome encoded 112 unique genes: 78 CDS (excluding the *infA* pseudogene), 30 tRNAs, and 4 rRNAs (Table S2). Sixteen genes were duplicated in the IR regions, including 4 rRNAs, 7 tRNAs, and 5 CDS (*rpl23*, *rps7*, *ycf2*, *ndhB*, *rpl2*), excluding the trans-spliced *rps12*. The truncated copy of *ycf1*, located at the IRa/SSC boundary, was also excluded. Eighteen genes contained introns—6 tRNAs (*trnV-UAC*, *trnI-GAU*, *trnL-UAA*, *trnK-UUU*, *trnG-UCC*, *trnA-UGC*) and 12 CDS (*ndhB*, *petB*, *petD*, *rpl2*, *rpoC1*, *rpl16*, *ndhA*, *rps16*, *rps12*, *ycf3*, *clpP*, *atpF*). Sixteen genes had single introns; *clpP* and *ycf3* had two (Fig. 1; Table S2).

Overall GC content ranged from 36.57% to 37.14%. IRs exhibited the highest GC content (42.47%–43.01%), followed by LSC (34.27%–35.04%) and SSC (30.87%–31.76%). Among gene categories, rRNAs had the highest GC content (up to 55.5%), followed by tRNAs (52.81%–53.25%) and CDS (38.03%–38.22%), reflecting the GC-rich nature of IRs and rRNA loci. Details are provided in Table 2.

Inverted repeats contraction and expansion

Inverted repeat contraction and expansion analysis revealed highly conserved junction patterns across the 29 *Hibiscus* species (Fig. 2; Fig. S3). The *ycf1* gene consistently spanned the JSB (IRb/SSC) junction, initiating in IRb and terminating in the SSC region. This arrangement generated a truncated *ycf1* pseudogene at the JSA (SSC/IRa) junction, ranging from 101 bp (*H. volkensii*) to 1101 bp (*H. trionum*) (Fig. S3). The *ndhF* gene maintained a consistent position at the JSA boundary while remaining entirely within the SSC. At the JLA (IRa/LSC) junction, *trnH-GUG* was fully contained within the LSC. The *rps19* and *rpl2* genes were positioned at the JLB (LSC/IRb) junction.

Analysis of codon usage, amino acid frequency, and repeats

Codon usage and amino acid frequencies were highly conserved across all analyzed *Hibiscus* species. Relative synonymous codon usage (RSCU) analysis revealed a strong bias towards codons ending in A or T over those ending in C or G at the third nucleotide position (Table S3). All CDS began with the standard ATG start codon (encoding methionine), except for *ndhD*, which consistently used ACG. This deviation suggests a post-transcriptional RNA editing event—likely a C-to-U conversion—restoring the canonical AUG start codon. Leucine was the most abundant amino acid, followed by isoleucine, whereas cysteine was the least frequent (Table S4).

Analysis of simple sequence repeats (SSRs) revealed notable interspecific variation, with total repeat counts ranging from 50 to 117 (Table S5). Mononucleotide repeats were the most abundant, predominantly comprising A/T motifs. Dinucleotide repeats—mainly AT/TA—were generally the second most frequent, though surpassed by tetranucleotide repeats in some species. Among trinucleotide repeats, AAT/TTA motifs were most prevalent.

Synonymous/non-synonymous substitutions and transition/transversion patterns

Transversions outnumbered transitions in all species. Transition counts ranged from 650 to 1,327, while transversions ranged from 897 to 1,626 (Table S6). The transition-to-transversion (Ts/Tv) ratio varied from 0.65 to 0.82, with *Hibiscus cannabinus* showing the highest ratio, indicating a relative transition bias in that species. These patterns reflect species-specific divergence in cp. genome substitution dynamics within the genus *Hibiscus*.

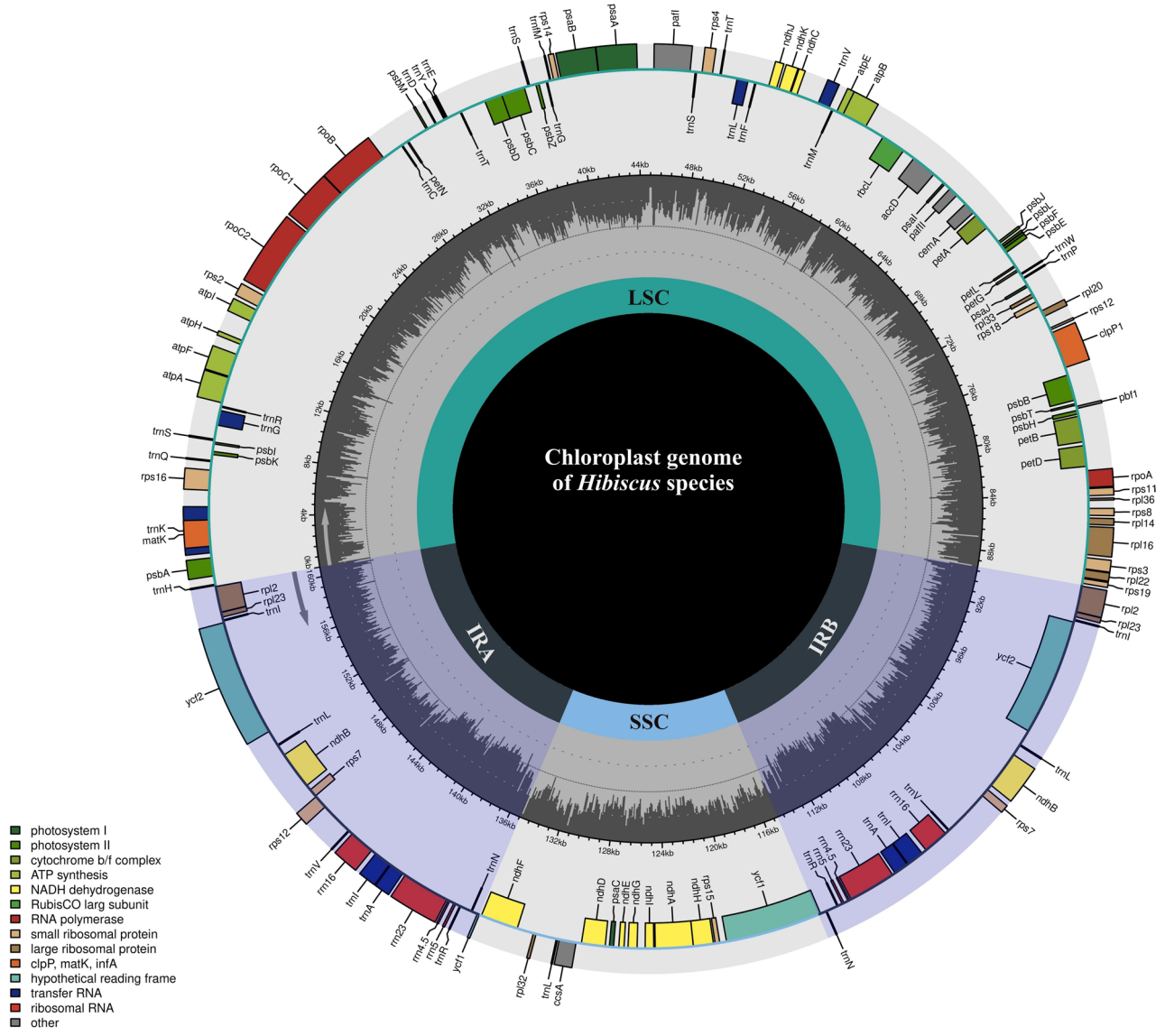

Fig. 1 Circular map of the *Hibiscus* chloroplast genomes, divided into four major regions: the Large Single-Copy (LSC), Small Single-Copy (SSC), and two inverted repeats (IRa and IRb). Genes displayed on the outer circle are transcribed clockwise, while those on the inner circle are transcribed counterclockwise. Functional gene categories are color-coded as indicated in the accompanying legend. The gray histogram in the inner circle represents the GC content distribution across the entire genome

Analysis of synonymous (*Ks*) and non-synonymous (*Ka*) substitution rates indicated strong purifying selection across most *Hibiscus* cp. CDS, with Ka/Ks ratios below 1 (Table S7). Only four genes—*ndhK*, *rpl22*, *rpl23*, and *ycf2*—exhibited signatures of positive selection (Ka/Ks > 1). Specifically, *ndhK* was under positive selection in *H. micranthus*; *rpl22* in *H. aponeurus*, *H. fuscus*, *Hibiscus hamabo*, *Hibiscus sabdariffa*, and *Hibiscus tiliaceus*; *ycf2* in *H. volkensii*; and *rpl23* in *Hibiscus coccineus*, *Hibiscus moscheutos*, *Hibiscus mutabilis*, *Hibiscus panduriformis*, *Hibiscus richardsonii*, *Hibiscus taiwanensis*, *Hibiscus tridactylites*, *Hibiscus trionum*, and *Hibiscus verdcourtii*. These patterns suggest potential lineage-specific adaptive evolution within the genus.

Identification of suitable polymorphic loci

Analysis of nucleotide diversity (π) revealed that intergenic and intronic regions exhibited the highest polymorphism, while CDS were comparatively less variable, and tRNA and rRNA genes remained highly conserved (Fig. 3). The most variable loci were selected based on the top π values, without applying a strict numerical threshold. Five loci were identified from non-coding regions, including *trnH-psbA* (π = 0.091), *ndhF-rpl32* (π = 0.075), *rps19-rpl22* (π = 0.071), *trnR-atpA* (π = 0.057), and

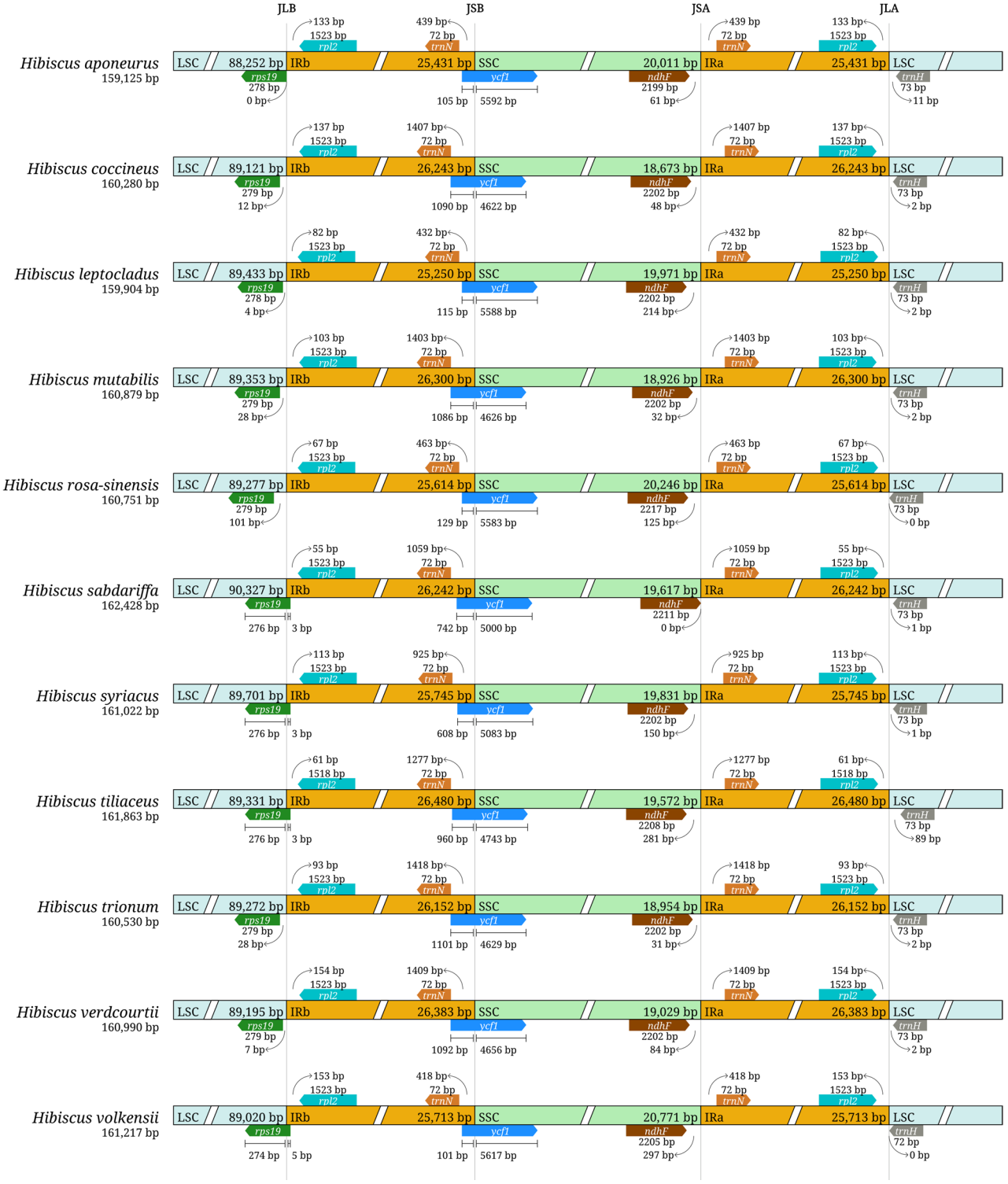

Fig. 2 Comparison of Large Single-Copy (LSC), Small Single-Copy (SSC), and Inverted Repeat (IR) boundary regions across chloroplast genomes of *Hibiscus* species. The positions of genes at the four junctions—JLA (LSC/IRa), JSA (IRa/SSC), JSB (SSC/IRb), and JLB (IRb/LSC)—are shown. Each horizontal bar represents the cp. genome structure, with gene annotations indicated at boundary regions. The comparison of all 29 species is provided in Figure S3

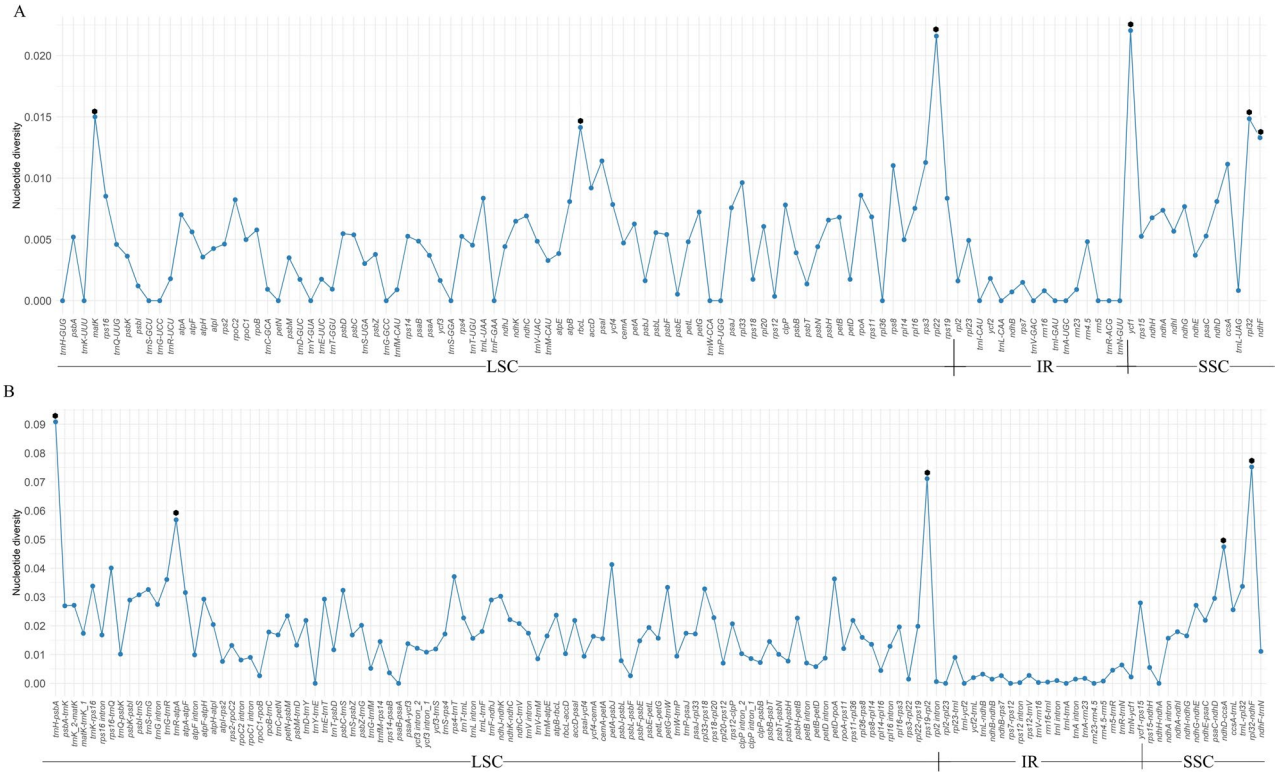

Fig. 3 Nucleotide diversity across chloroplast genome regions in *Hibiscus* species. **A** Protein-coding, tRNA, and rRNA genes. **B** Intergenic spacers and intronic regions. The x-axis denotes specific chloroplast regions; the y-axis represents nucleotide diversity. A schematic line beneath the plots indicates the location of each region within the chloroplast genome's structural regions: LSC (large single-copy), SSC (small single-copy), and IR (inverted repeat). Black markers highlight 11 (six coding and five non-coding) suitable polymorphic loci

ndhD-ccsA (π = 0.047). Six loci were identified from coding regions, including *rpl22* (π = 0.022), *ycf1* (π = 0.022), *matK* (π = 0.015), *rpl32* (π = 0.014), *rbcL* (π = 0.014), and *ndhF* (π = 0.013). Although the π value of *ndhF* was slightly lower, it was included because the difference was minimal, and this gene has been widely employed in previous phylogenetic studies of *Hibiscus* and related taxa.

Phylogeny of the genus *Hibiscus*

The phylogenetic analysis of complete cp. coding sequences from 42 Hibisceae species yielded a strongly supported and well-resolved tree (Fig. 4), rooted with *B. ceiba* as the outgroup. The results confirmed the polyphyletic nature of *Hibiscus*, as species from *Abelmoschus*, *Malvaviscus*, *Kosteletzkya*, *Urena*, and *Cenocentrum* were nested within the *Hibiscus* clade. At a broader level, the recovered topology aligns with the three major *Hibiscus* lineages proposed by Pfeil and Crisp [6]—/Euhibiscus, /Phylloglandula, and /Trionum—all of which received strong support (backbone nodes = 95–100% BS).

The /Trionum clade (BS = 98–100%, with one exception) comprised three primary components: the *Abelmoschus* lineage forming a tightly grouped cluster suggestive of recent divergence; a lineage containing *H. mutabilis* and *H. taiwanensis* (*Hibiscus* sect. *Venusti*) together with *H. panduriformis* (*Hibiscus* sect. *Panduriformes*, now recognized as *Cravenia* McLay & R.L.Barrett); and a clade of species from *Hibiscus* sect. *Trionum*, recovered as sister to *Malvaviscus penduliflorus* Moc. & Sessé ex DC. Another supported lineage (BS = 88) included *Hibiscus coccineus* Walter and *Hibiscus moscheutos* L. (*Hibiscus* sect. *Muenchhusia*, now reinstated as a separate genus *Muenchhusia*), which grouped with *H. verdcourtii* and *Thepparatia scandens*, forming a distinct sister lineage despite belonging to different genera.

The /Phylloglandula clade (BS = 96–100%) comprised *H. sabdariffa* and *H. cannabinus* (*H.* sect. *Furcaria*, now reinstated as a separate genus *Sabdariffa*) sister to two *Urena* species, all showing long branches. *H. hamabo* and *H. tiliaceus* were recovered as sister to *Kydia* Roxb., also with extended branch lengths. *H. volkensii* formed an isolated, highly divergent lineage within this clade.

The /Euhibiscus clade was strongly supported (BS = 100% at the principal node), with most internal nodes also exhibiting maximal support, except two nodes within *H.* sect. *Bombicella* clade I (BS = 50 and 89). *H.* sect. *Bombicella* clade I displayed complex relationships, potentially reflecting rapid diversification. Within this clade, *H. syriacus*, *Hibiscus sinosyriacus* (*Hibiscus*

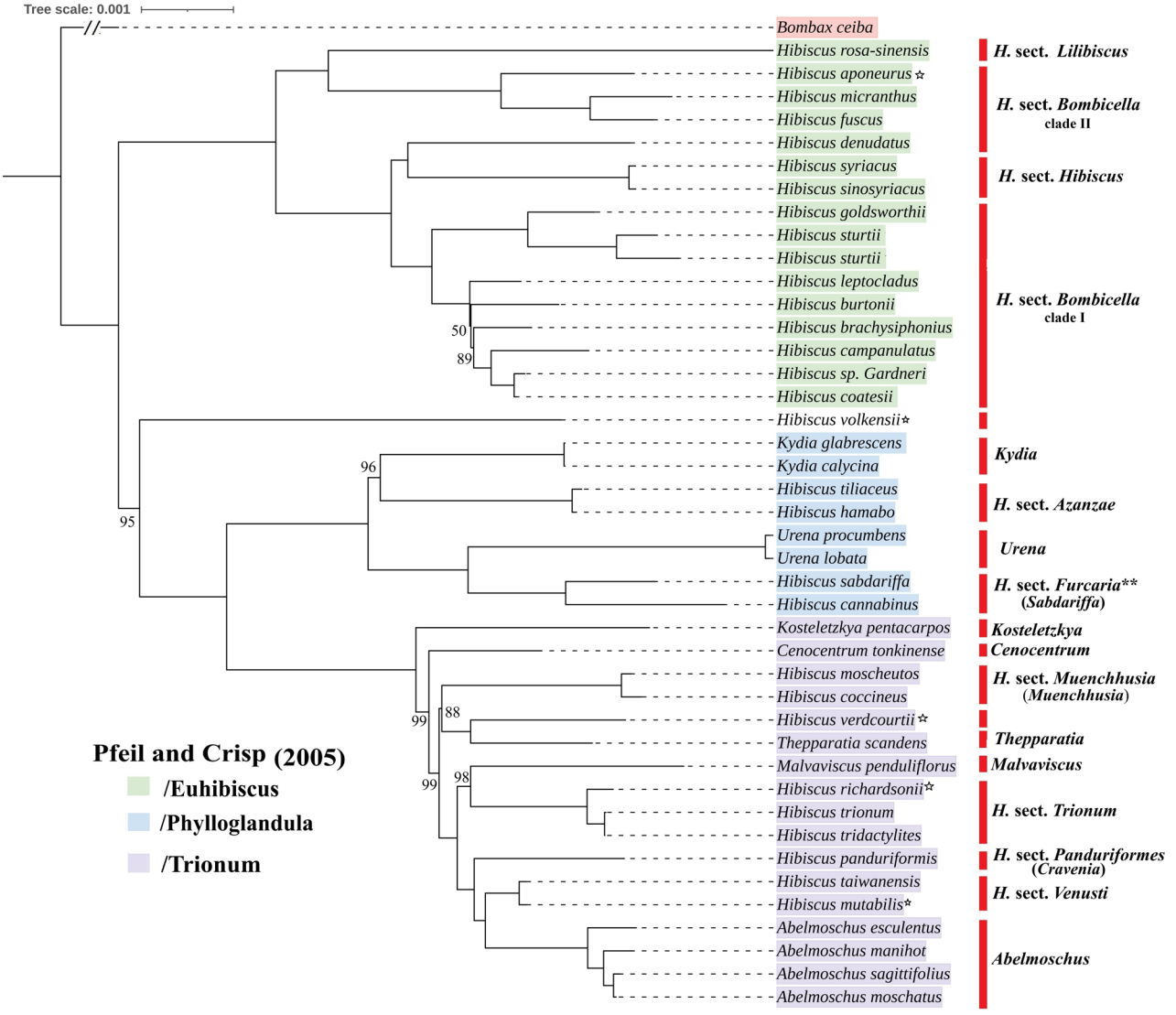

Fig. 4 Maximum likelihood phylogeny of *Hibiscus* and related genera inferred from complete chloroplast genomes. Bootstrap values are shown for branches with < 100% support. The scale bar represents 0.001 nucleotide substitutions per site. Background shading denotes major clades following Pfeil and Crisp (2005) [6]: /Euhibiscus (light green), /Trionum (lavender), and /Phylloglandula (light blue). Sectional or generic affiliations are indicated by red vertical bars, based primarily on Hanes et al. [5] and supplemented by other recent studies. *Hibiscus* sect. *Panduriformes* is now recognized as *Cravenia* McLay & R.L.Barrett, and *Hibiscus panduriformis* is now named *Craveniapanduriformis* (Burm.f.) McLay & R.L.Barrett. *Hibiscus* sect. *Furcaria* has been reinstated as genus *Sabdariffa* by Barrett et al. [4] and H. *coccineus* and *H. moscheutos* (sect. *Muenchhusia*) have been reinstated as genus *Muenchhusia* by Hanes and Barrett [55]. The genera *Kydia*, *Urena*, *Kosteletzkya*, *Cenocentrum*, *Thepparatia*, and *Malvaviscus* are nested within *Hibiscus*, highlighting the polyphyly of the genus. Species marked with ☆ were not included in Hanes et al. [5]

sect. *Hibiscus*), and *H. denudatus* formed a basal lineage. Other species of *H.* sect. *Bombicella* (clade II), including *H. aponeurus*, *H. fuscus*, and *H. micranthus*, were placed with *H. rosa-sinensis* (*Hibiscus* sect. *Lilibiscus*) at the base. These findings demonstrate that *H.* sect. *Bombicella* is polyphyletic, with its species distributed across several well-supported lineages.

Overall, the cp. phylogeny highlights extensive diversification within *Hibiscus*, reinforces its polyphyletic nature, and reveals that several species are genetically more closely allied to other genera of Hibisceae than to *Hibiscus* itself.

Discussion

We performed *de novo* assembly of the complete cp. genomes of 16 *Hibiscus* species and conducted a comparative genomic analysis with 13 previously published *Hibiscus* genomes. This comprehensive dataset provides insights into the evolutionary dynamics of *Hibiscus* cp. genomes and enhances our understanding of the taxonomy within this complex genus.

Conserved structure of the *Hibiscus* cp. genome and patterns of inverted repeat expansion and contraction

Comparative genomic analysis of 29 *Hibiscus* species revealed strong conservation in genome structure, gene organization, and gene content, with no evidence of inversions or major structural rearrangements. These findings are consistent with studies of Malvaceae [19, 26, 29, 50, 56] and other angiosperm families—including Solanaceae [57, 58], Asteraceae [59], and Araceae [20, 60]—where cp. genomes typically maintain stable gene content and arrangement. Such conservation likely results from molecular constraints on cp. genome evolution, including operon-based gene organization, uniparental inheritance, efficient DNA repair mechanisms, and the rarity of cp. fusion events [22]. However, cp. genome rearrangements and gene losses have been documented in certain angiosperm lineages such as Fabaceae, Thymelaeaceae, and Cyperaceae [12, 17, 61, 62], underscoring exceptions to this general pattern. Recently, a synapomorphic inversion has been reported in *Triumfetta* (Malvaceae) [33, 34]. Despite this finding, species within all three major clades—/Phylloglandula, /Trionum, and /Euhibiscus—of *Hibiscus* showed high similarity in gene content and arrangement. Our analysis also confirmed that the GC content of *Hibiscus* cp. genomes aligns with previous reports from Malvaceae [19, 26, 50, 56, 63] and other plant families [16, 20, 58, 59, 64–67], with elevated GC content in the IR regions—a trend attributed to the high GC content (up to 55%) of rRNA genes.

Chloroplast genome size can vary even within the same angiosperm family or genus due to contraction and expansion of IR regions and length variations in intergenic spacer regions caused by insertions and deletions [26, 68]. These structural changes can convert single-copy genes into duplicated genes or vice versa, often resulting in pseudogene formation at the junctions between the LSC, SSC, and IR regions [26, 69]. In the current study, a pseudogenized copy of *ycf1*$^{\Psi}$ was observed at the JSA junction.

The dynamic nature of IR boundary shifts provides valuable insights into evolutionary patterns at both the population genetic and phylogenetic levels [24, 70–72]. Closely related species typically exhibit high similarity in cp. genome junctions [73], whereas more distantly related taxa show greater divergence [74]. In the current study, *Hibiscus* species displayed conserved junction structures at the genus level, consistent with reports for Malvaceae and other angiosperm genera [21, 73]. However, some angiosperm lineages, such as Araceae and Apocynaceae [72, 74, 75], exhibit significant variation even among closely related species.

Identification of high-resolution markers in the cp. genome

Not all cp. genes are equally informative for robust phylogenetic resolution due to heterogeneous evolutionary rates across the genome [24, 26]. Previous phylogenetic studies of *Hibiscus* and tribe Hibisceae have relied on established cp. markers (e.g., *rbcL*, *matK*, *trnH-psbA*, *trnQ-rps16*, *ycf1*, *ndhF*, *rpl16*) alongside nuclear markers such as the internal transcribed spacer (ITS) regions and the coding gene *GBSSI* (granule-bound starch synthase) [4–6, 10, 32, 76]. However, these markers offer limited resolution at shallow taxonomic levels, highlighting the need for more variable loci to resolve recent divergences and support DNA barcoding in seed plants [67, 77].

Our whole-genome comparative approach enabled systematic screening of sequence variation across both coding and non-coding regions, revealing loci with potentially higher phylogenetic utility within *Hibiscus*. Based on analyses of 29 *Hibiscus* species, we identified five polymorphic non-coding loci (*trnH-psbA*, *ndhF-rpl32*, *rps19-rpl22*, *trnR-atpA*, and *ndhD-ccsA*) and six coding loci (*rpl22*, *ycf1*, *matK*, *rpl32*, *rbcL*, and *ndhF*). While partially overlapping with traditionally used markers [77], these loci include novel intergenic regions and exhibit elevated nucleotide diversity in lineage-specific comparisons, suggesting greater discriminatory power for resolving recent divergences. Despite inherent limitations of the cp. genome, such as uniparental inheritance and slower substitution rates compared to nuclear DNA [12, 22], cp. markers remain valuable due to cp. genome structural conservation, ease of amplification, and high copy number. For studies prioritizing cp. genome data, the loci identified here may improve topological resolution and support values in *Hibiscus* phylogenies.

Phylogenetic analysis and taxonomic implications in *Hibiscus*

Our results highlight the polyphyly of *Hibiscus* based on complete cp. coding sequences, as species of some related genera (*Malvaviscus*, *Kosteletzkya*, *Urena*, *Kydia*, and *Cenocentrum*, etc.) are nested within *Hibiscus*. This finding is consistent with earlier studies that identified similar patterns using both morphological and molecular data [5, 10, 11]. In response to such polyphyly, Pfeil and Crisp [6] proposed three potential taxonomic solutions: (i) broadly circumscribing *Hibiscus* to include all embedded genera, (ii) splitting *Hibiscus* into multiple smaller, monophyletic genera, or (iii) accepting *Hibiscus* as paraphyletic. They also suggested a hybrid approach, incorporating embedded genera within an expanded concept of *Hibiscus*.

Additionally, our results support the classification of *Hibiscus* into four main clades: /Phylloglandula, /Trionum, /Euhibiscus, and /Calyphylli, as originally proposed by Pfeil and Crisp [6]. However, our sampling included

species from only the first three clades, with /Calyphylli unrepresented in this study. These clades were defined primarily by molecular phylogenetic data because, as Pfeil and Crisp [6] demonstrated, few reliable morphological synapomorphies exist to delimit monophyletic groups within *Hibiscus* and tribe Hibisceae. The difficulty in finding diagnostic morphological characters suggests that many floral and vegetative features exhibit homoplasy—evolving convergently in distantly related lineages or being retained as ancestral traits—making morphologically defined groups incongruent with phylogenetic relationships [6, 10]. Consequently, attempts to circumscribe *Hibiscus* or its sections using traditional morphological taxonomy have consistently failed to produce monophyletic groups. The complete cp. genome phylogeny reinforces this conclusion: monophyly within *Hibiscus* and Hibisceae may be achieved through phylogeny-based generic delimitation rather than morphology-based classification, regardless of whether this requires broadly circumscribing *Hibiscus* to include nested genera or splitting it into multiple smaller genera, as also suggested previously [6]. Our phylogeny is generally congruent with recent classifications that have reassigned certain species across sections or genera and have suggested the reinstatement of *Sabdariffa* as a separate genus [4, 5]. Moreover, the strong bootstrap support for /Euhibiscus and the recent treatment of its species as genus *Hibiscus* [5] indicate that dividing *Hibiscus* into smaller genera may be taxonomically feasible; however, such decisions should be evaluated using integrative phylogenomic evidence.

Recently, Hanes et al. [5] described the species of /Euhibiscus as *Hibiscus*. However, they also noted the paraphyly of *H.* sect. *Bombicella*. In our phylogeny, *H. syriacus* and *H. sinosyriacus* (*H.* sect. *Hibiscus*) share a clade with *H. denudatus* (*H.* sect. *Bombicella* clade II), and these species form a clade that is sister to other species of *H.* sect. *Bombicella* clade I. This clade is itself sister to another distinct group comprising *H. rosa-sinensis* (*H.* sect. *Lilibiscus*), *H. aponeurus* (not sampled by Hanes et al. [5]), *H. fuscus*, and *H. micranthus* (both *H.* sect. *Bombicella* clade II). These findings indicate that the current section-level classification remains inconclusive despite the use of the complete cp. genome and requires further taxonomic revision using nuclear markers, as this paraphyly may result from incomplete lineage sorting (ILS) or frequent hybridization, as discussed below.

Notably, *H. coccineus* and *H. moscheutos* (sect. *Muenchhusia*) formed a distinct clade in our phylogeny. This result aligns with their recognition as a separate lineage within Hibisceae and is congruent with the recent reinstatement of *Muenchhusia* as a distinct genus [5, 55]. The independent congruence between our plastome data and this subsequent taxonomic revision provides additional support for the phylogenetic distinctiveness of this lineage.

Furthermore, *H. verdcourtii* (*H.* sect. *Trionum*) and *T. scandens* formed a strongly supported sister relationship (100% bootstrap), separated by a long branch indicating substantial phylogenetic divergence. Notably, *H. verdcourtii* did not group with other species of *H.* sect. *Trionum* but instead occupied an isolated position, suggesting that it may align more closely with another lineage or even warrant recognition as a separate genus. This deep divergence also supports the treatment of *Thepparatia* as a distinct genus by Hanes et al. [5]. However, resolving the taxonomy of *H. verdcourtii* requires further investigation, as the *H. verdcourtii* cp. genome used here was obtained from NCBI; therefore, sequencing of verified voucher specimens is essential to confirm its phylogenetic placement. Once validated, these results would provide important evidence for a broader taxonomic revision of *H.* sect. *Trionum*, which remains in need of comprehensive reassessment [5]. Likewise, *H. volkensii* forms a highly divergent single-species clade. In contrast, *H. cannabinus* and *H. sabdariffa*, although recently reinstated as the separate genus *Sabdariffa* [4] and forming a monophyletic clade sister to *Urena*, exhibit relatively shallow divergence from other *Hibiscus* species. Given the substantially greater phylogenetic distance exhibited by *H. volkensii*, it presents an even stronger case for reinstatement as a distinct genus, pending more comprehensive morphological and molecular investigations.

The main limitation of this study is the inclusion of only 29 *Hibiscus* species (25 when considering the recent reinstatement of *Sabdariffa* and the reassignment of some *Hibiscus* species to other genera by Hanes et al. [5], Barrett et al. [4], and Hanes and Barrett [55]) out of more than 400 species (or approximately 220 following recent taxonomic revisions). Although the cp. genome is maternally inherited and has been widely used for phylogenetic studies of *Hibiscus* [4, 5, 10, 11], several biological processes can lead to discordance between cp. and nuclear phylogenies. Hybridization is well documented in *Hibiscus* and has been reported in several natural populations, suggesting that gene flow across species boundaries is not uncommon [78]. Frequent hybridization may result in the introgression of cp. genomes across species boundaries, producing cp-based trees that do not fully reflect nuclear evolutionary relationships [79, 80]. Incomplete lineage sorting may also contribute, especially in rapidly radiating lineages where ancestral polymorphisms persist and become randomly fixed in descendant species [81, 82]. These processes likely influenced some unexpected placements in our cp. phylogeny, including the highly divergent *H. volkensii* and the mixed relationships of species within *H.* sect. *Bombicella*. Such discordances may also explain why some species of *H.* sect. *Hibiscus*

(e.g., *H. syriacus* and *H. sinosyriacus*) group with species of other sections rather than forming cohesive, section-specific clades.

Consequently, while our cp. phylogeny provides strong evidence for polyphyly and reveals the major clade structure within *Hibiscus*, it may not fully capture the evolutionary history of the genus. Integrating nuclear genomic data—using multi-locus or genome-wide approaches—alongside broader taxon sampling will be critical to resolving these conflicts. Such analyses can disentangle the effects of hybridization and ILS, providing a more comprehensive understanding of evolutionary relationships. These considerations directly inform the perspectives outlined in the conclusion, which emphasize the need for nuclear genomic approaches to clarify the taxonomy and evolutionary history of *Hibiscus* and other genera within the tribe Hibisceae.

Conclusion and future prospects

Our comparative analysis of cp. genomes reveals a generally conserved structure of *Hibiscus* species but identifies localized regions of significant nucleotide diversity. We identified specific candidate markers exhibiting substantially higher polymorphism than traditional loci, offering strong potential for enhanced phylogenetic resolution; however, their broader utility requires further validation. Phylogenetic reconstructions confirm the polyphyly of *Hibiscus*, with related genera consistently nested within the clade, aligning with previous studies. However, persistent incongruences at the sectional level suggest unresolved complexities, likely due to hybridization, incomplete lineage sorting, or the limitations of cp-only data. To resolve these relationships, future studies should incorporate nuclear genomic information, particularly using multi-locus approaches such as Angiosperms353 or transcriptomic datasets, which have demonstrated success in addressing complex phylogenetic patterns [83–86]. Overall, the findings of this study offer significant insights into cp. genome evolution in *Hibiscus* and establish a valuable basis for future phylogenetic and taxonomic research on *Hibiscus* and other genera within the tribe Hibisceae.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12870-026-08791-5.

Supplementary Material 1: Fig S1. Mapping results of Hibiscus species. The name of the species is mentioned with each graph. These graphs also show statistics on the mapped reads, including their quality, mean, minimum, and maximum coverage depth.

Supplementary Material 2: Fig S2. Mauve alignment showing the structural organization of 29 Hibiscus chloroplast genomes based on collinear blocks. The alignment reveals a high degree of similarity among all chloroplast genomes. Colored blocks represent gene regions: black = transfer RNA, red = ribosomal RNA, white = protein-coding genes, and green = intron-containing tRNA genes.

Supplementary Material 3: Fig S3. Comparison of large single-copy (LSC), small single-copy (SSC), and inverted repeat (IR) boundary regions across chloroplast genomes of Hibiscusspecies. The positions of genes at the four junctions—JLA (LSC/IRa), JSA (IRa/SSC), JSB (SSC/IRb), and JLB (IRb/LSC)—are shown. Each horizontal bar represents the cp genome structure, with gene annotations indicated at boundary regions.

Supplementary Material 4. Table S1 and Table S2

Supplementary Material 5. Table S3

Supplementary Material 6. Table S4

Supplementary Material 7. Table S5

Supplementary Material 8. Table S6

Supplementary Material 9. Table S7

Supplementary Material 10. Codon usage python script

Supplementary Material 11. Amino acid frequency python script

Supplementary Material 12. SNPs analysis python script

Supplementary Material 13. Hibiscus chloroplast genomes

Acknowledgements

We gratefully acknowledge all researchers who submitted raw data or de novo assembled cp. genomes to NCBI and other public databases, which enabled the comparative genomics and phylogenomics analyses of Hibiscus conducted in this study. Accession numbers for each species utilized are provided in Tables 1 and 2 of the manuscript. We also extend our sincere thanks to Dr. Peter Poczai (University of Helsinki) and the Finnish Biodiversity Information Facility (FinnBIF), which is maintained by LUOMUS at the University of Helsinki, for providing the Hibiscus leaf samples used for DNA extraction and sequencing. Author Ibrar Ahmed is supported by the Brain Pool Fellowship, the National Research Foundation (NRF) of Korea through Grant Number RS-2023-00223245 for his postdoctoral research.

Authors' contributions

Writing original draft: A, CC, AV; Data analysis: A, CC, RY, HIM; Data curation: A, RY, AS; Data interpretation: A, AV, AS, TH; Software: A, AV, PH; Conceptualization: A, AV, IA, XT, MTW; Review and editing of manuscript: A, IA, PH, MTW, TA, XT. Figures 1, 2, 3 and 4 were drawn by A and AV.

Funding

This study did not receive any funding.

Data availability

The raw sequencing data generated in this study are available under BioProject PRJNA1196377, and the de novo assembled cp. genomes of 5 newly sequenced species have been deposited in NCBI under accession numbers PV666612–PV666616. In addition, we assembled 12 cp. genomes from publicly available raw sequencing data from NCBI and ENA, representing 11 species (including two genomes for the two varieties of ***H. sturtii***). The accession numbers of the raw datasets are listed in Table 1, and the corresponding cp. genome accession numbers for the assembled genomes are provided in Table 2. All de novo assembled cp. genomes are also available in Supplementary Data 1.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

All the authors approved the content of the manuscript and agreed to their authorship.

Competing interests

The authors declare no competing interests.

Received: 9 July 2025 / Accepted: 15 April 2026
Published online: 18 April 2026

References

1. POWO (Plants of the World online). Facilitated by the Royal Botanic Gardens, Kew. 2025. Available at: https://powo.science.kew.org/. Accessed 15 Jun 2025.
2. Bayer C, Fay MF, De Bruijn AY, Savolainen V, Morton CM, Kubitzki K, et al. Support for an expanded family concept of Malvaceae within a recircumscribed order Malvales: A combined analysis of plastid *atpB* and *rbcL* DNA sequences. Bot J Linn Soc. 1999;129:267–303.
3. Bayer C, Kubitzki K. Malvaceae. Flowering Plants · Dicotyledons. Berlin, Heidelberg: Springer Berlin Heidelberg; 2003. pp. 225–311.
4. Barrett RL, Yoshikawa VN, McLay TGB, Duarte MC, Mwachala G, Hanes MM. Reinstatement of *Sabdariffa* and new combinations to support a monophyletic concept of *Hibiscus* (Malvaceae: Hibisceae). Aust Syst Bot. 2025;38:SB24013.
5. Hanes MM, Blanchard OJ, Valencia-D J, McLay T, Abbott JR, McDaniel SF, et al. Phylogenetic relationships within tribe Hibisceae (Malvaceae) reveal complex patterns of polyphyly in *Hibiscus* and *Pavonia*. Syst Bot. 2024;49:77–116.
6. Pfeil BE, Crisp M. What to do with Hibiscus ? A proposed nomenclatural resolution for a large and well known genus of Malvaceae and comments on paraphyly. Aust Syst Bot. 2005;18:49–60.
7. Vasudeva N, Sharma SK. Biologically active compounds from the genus *Hibiscus*. Pharm Biol. 2008;46:145–53.
8. Slamet A. The Diversity of *Hibiscus rosa-sinensis* based on Morphological Approach. Sci Educ J Pendidik Sains. 2018;7:33–41.
9. Abdelfattah HME, Hussein HA, Teleb SS, El-Demerdash MM, George NM. Chemotaxonomy compared to morphological and anatomical taxonomy of five *Hibiscus* species. J Plant Res. 2024;137:967–84.
10. Pfeil BE, Brubaker CL, Craven La, Crisp MD. Phylogeny of *Hibiscus* and the tribe Hibisceae (Malvaceae) using chloroplast DNA sequences of *ndhF* and the *rpl16* intron. Syst Bot. 2002;27:333–50.
11. Koopman MM, Baum DA. Phylogeny and Biogeography of Tribe Hibisceae (Malvaceae) on Madagascar. Syst Bot. 2008;33:364–74.
12. Daniell H, Lin C-S, Yu M, Chang W-J. Chloroplast genomes: diversity, evolution, and applications in genetic engineering. Genome Biol. 2016;17:134.
13. Daniell H. Transgene containment by maternal inheritance: effective or elusive? Proc Natl Acad Sci U S A. 2007;104:6879–80.
14. Neale DB, Sederoff RR. Paternal inheritance of chloroplast DNA and maternal inheritance of mitochondrial DNA in *Loblolly pine*. Theor Appl Genet. 1989;77:212–6.
15. Palmer JD. Comparative organization of chloroplast genomes. Annu Rev Genet. 1985;19:325–54.
16. Abdullah RA, Chen C, Yan R, Naz I, Heidari P, et al. Comparative chloroplast genomics of *Erigeron* (Asteroideae, Asteraceae). BMC Plant Biol. 2025;25:1356.
17. Abdullah, Li H, Yan R, Chen C, Islam M, Majid A et al. Evolutionary dynamics of the chloroplast genome in *Daphne* (Thymelaeaceae): comparative analysis with related genera and insights into phylogenetics. FEBS Open Biol. 202616:503–19.
18. Wu C-S, Wang Y-N, Hsu C-Y, Lin C-P, Chaw S-M. Loss of different inverted repeat copies from the chloroplast genomes of Pinaceae and Cupressophytes and influence of heterotachy on the evaluation of gymnosperm phylogeny. Genome Biol Evol. 2011;3:1284–95.
19. Abdullah MF, Shahzadi I, Ali Z, Islam M, Naeem M, et al. Correlations among oligonucleotide repeats, nucleotide substitutions, and insertion–deletion mutations in chloroplast genomes of plant family Malvaceae. J Syst Evol. 2021;59:388–402.
20. Henriquez CL, Abdullah, Ahmed I, Carlsen MM, Zuluaga A, Croat TB, et al. Molecular evolution of chloroplast genomes in Monsteroideae (Araceae). Planta. 2020;251:72.
21. Abdullah FA, Chen C, Yan R, Naz I, Li H. The chloroplast genome of *Trollius acaulis*: insights into comparative genomics and phylogenetic relationships. J Asia-Pacific Biodivers. 2026;19:1–8. https://doi.org/10.1016/j.japb.2025.03.008.
22. Wicke S, Schneeweiss GM, dePamphilis CW, Müller KF, Quandt D. The evolution of the plastid chromosome in land plants: Gene content, gene order, gene function. Plant Mol Biol. 2011;76:273–97.
23. Nguyen VB, Linh Giang VN, Waminal NE, Park HS, Kim NH, Jang W, et al. Comprehensive comparative analysis of chloroplast genomes from seven *Panax* species and development of an authentication system based on species-unique single nucleotide polymorphism markers. J Ginseng Res. 2020;44:135–44.
24. Li H, Abdullah, Yang H, Guo H, Yuan Y, Ahmed I, et al. Chloroplast genome evolution of *Berberis* (Berberidaceae): Implications for phylogeny and metabarcoding. Gene. 2025;933:148959.
25. Cvetković T, Areces-Berazain F, Hinsinger DD, Thomas DC, Wieringa JJ, Ganesan SK, Strijk JS. Phylogenomics resolves deep subfamilial relationships in Malvaceae sl. G3. 2021;11(7):jkab136. https://doi.org/10.1093/g3journal/jkab136.
26. Abdullah MF, Shahzadi I, Waseem S, Mirza B, Ahmed I et al. Chloroplast genome of *Hibiscus rosa-sinensis* (Malvaceae): Comparative analyses and identification of mutational hotspots. Genomics. 2020;112:581–91.
27. Kwon SH, Kwon HY, Choi YI, Shin H. Comprehensive analysis of chloroplast genome of *Hibiscus sinosyriacus*: evolutionary studies in related species and genera. Forests. 2023;14:2221.
28. Kim Y, Oh YJ, Han KY, Kim GH, Ko J, Park J. The complete chloroplast genome sequence of *Hibiscus syriacus* L. 'Mamonde' (Malvaceae). Mitochondrial DNA Part B. 2019;4:558–9.
29. Cheng Y, Zhang L, Qi J, Zhang L. Complete chloroplast genome sequence of *Hibiscus cannabinus* and comparative analysis of the Malvaceae family. Front Genet. 2020;11:512826.
30. Tang D, Wei F, Zhou R. Comparative analysis of chloroplast genomes of Kenaf cytoplasmic male sterile line and its maintainer line. Sci Rep. 2021;11:5301.
31. Kwon H-Y, Kim J-H, Kim S-H, Park J-M, Lee H. The complete chloroplast genome sequence of *Hibiscus syriacus*. Mitochondrial DNA Part A. 2016;27:3668–9.
32. Small RL. Phylogeny of *Hibiscus* sect. *Muenchhusia* (Malvaceae) based on chloroplast *rpL16* and *ndhF*, and nuclear ITS and GBSSI sequences. Syst Bot. 2004;29:385–92.
33. Yan R, Zeng Y, Chen C, Rab A, Li H, Ullah R et al. Molecular evolution of chloroplast genome in *Triumfetta* (Grewioideae, Malvaceae). Planta. 2025;262:82.
34. Rushan Y, Abdullah, Li H, Sammad A, Fu M, Cui Y, et al. Complete chloroplast genomes of four *Triumfetta* species from herbarium specimens: comparative analysis and identification of a synapomorphic inversion. Authorea (Preprint). 2025. https://doi.org/10.22541/AU.176081979.91848030/V1.
35. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics. 2018;34:i884–90.
36. Huang L, Yu |, Huanxi, Wang Z, Xu W. CPStools: A package for analyzing chloroplast genome sequences. iMetaOmics. 2024;1:e25.
37. Jin J-J, Yu W-B, Yang J-B, Song Y, DePamphilis CW, Yi T-S, et al. GetOrganelle: a simple and fast pipeline for de novo assembly of a complete circular chloroplast genome using genome skimming data. Genome Biol. 2020;21:241.
38. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, et al. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics. 2012;28:1647–9.
39. Tillich M, Lehwark P, Pellizzer T, Ulbricht-Jones ES, Fischer A, Bock R, et al. GeSeq – versatile and accurate annotation of organelle genomes. Nucleic Acids Res. 2017;45:W6–11.
40. Chan PP, Lowe TM. tRNAscan-SE: Searching for tRNA genes in genomic sequences. Methods in Molecular Biology. Humana Press Inc. 2019. pp. 1–14.
41. Laslett D, Canback B. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. Nucleic Acids Res. 2004;32:11–6.
42. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. Mol Biol Evol. 2013;30:772–80.
43. Lehwark P, Greiner S. GB2sequin - A file converter preparing custom GenBank files for database submission. Genomics. 2019;111:759–61.
44. Zheng S, Poczai P, Hyvönen J, Tang J, Amiryousefi A. Chloroplot: An online program for the versatile plotting of organelle genomes. Front Genet. 2020;11:25.
45. Darling AE, Mau B, Perna NT, progressiveMauve. Multiple genome alignment with gene gain, loss and rearrangement. PLoS ONE. 2010;5:e11147.
46. Amiryousefi A, Hyvönen J, Poczai P. IRscope: an online program to visualize the junction sites of chloroplast genomes. Bioinformatics. 2018;34:3030–1.
47. Abdullah YR, Tian X. CGAS (Chloroplast Genome Analysis Suite): an automated python pipeline for comprehensive comparative chloroplast genomics. iMetaOmics. 2026;e70093 (online first). https://doi.org/10.1002/imo2.70093.
48. Thiel T, Michalek W, Varshney R, Graner A. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L). Theor Appl Genet. 2003;106:411–22.

49. Zhang Z, Li J, Zhao X-Q, Wang J, Wong GK-S, Yu J. KaKs_Calculator: Calculating Ka and Ks Through Model Selection and Model Averaging. Genomics Proteom Bioinf. 2006;4:259–63.
50. Wang JH, Moore MJ, Wang H, Zhu ZX, Wang HF. Plastome evolution and phylogenetic relationships among Malvaceae subfamilies. Gene. 2021;765:145103.
51. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, Von Haeseler A, et al. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. Mol Biol Evol. 2020;37:1530–4.
52. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods. 2017;14:587–9.
53. Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. UFBoot2: Improving the ultrafast bootstrap approximation. Mol Biol Evol. 2018;35:518–22.
54. Letunic I, Bork P. Interactive Tree of Life (iTOL) v4: Recent updates and new developments. Nucleic Acids Res. 2019;47:W256–9.
55. Hanes MM, Barrett RL. The hardy Hibiscus—still hardy, but no longer *Hibiscus*: Reinstatement of *Muenchhusia* (Malvaceae: Hibisceae). J Bot Res Inst Tex. 2025;19:375–87.
56. Abdullah SI, Mehmood F, Ali Z, Malik MS, Waseem S, et al. Comparative analyses of chloroplast genomes among three *Firmiana* species: Identification of mutational hotspots and phylogenetic relationship with other species of Malvaceae. Plant Gene. 2019;19:100199.
57. Mehmood F, Abdullah, Shahzadi I, Ahmed I, Waheed MT, Mirza B. Characterization of Withania somnifera chloroplast genome and its comparison with other selected species of Solanaceae. Genomics. 2020;112:1522–30.
58. Amiryousefi A, Hyvönen J, Poczai P. The chloroplast genome sequence of bittersweet (*Solanum dulcamara*): Plastid genome structure evolution in Solanaceae. PLoS ONE. 2018;13:e0196069.
59. Abdullah MF, Rahim A, Heidari P, Ahmed I, Poczai P. Comparative plastome analysis of *Blumea*, with implications for genome evolution and phylogeny of Asteroideae. Ecol Evol. 2021;11:7810–26.
60. Henriquez CL, Abdullah, Ahmed I, Carlsen MM, Zuluaga A, Croat TB, et al. Evolutionary dynamics of chloroplast genomes in subfamily Aroideae (Araceae). Genomics. 2020;112:2349–60.
61. Keller J, Rousseau-Gueutin M, Martin GE, Morice J, Boutte J, Coissac E, et al. The evolutionary fate of the chloroplast and nuclear *rps16* genes as revealed through the sequencing and comparative analyses of four novel legume chloroplast genomes from *Lupinus*. DNA Res. 2017;24:343–58.
62. Lee C, Ruhlman TA, Jansen RK. Unprecedented intraindividual structural heteroplasmy in *Eleocharis* (Cyperaceae, Poales) plastomes. Genome Biol Evol. 2020;12:641–55.
63. Abdullah WS, Mirza B, Ahmed I, Waheed MT. Comparative analyses of chloroplast genomes of *Theobroma cacao* and *Theobroma grandiflorum*. Biol (Bratisl). 2020;75:761–71.
64. Zia A, Zahoor M, Abdullah, Nisar A, Batool N, Bibi A, et al. De novo assembly and characterisation of chloroplast genomes of broccoli cvs. Marathon and Green sprout using next generation sequencing. Turkish J Agric For. 2022;46:525–35.
65. Chen C, Guo H, Abdullah, Li T, Li G, Liu J, et al. The complete chloroplast genome of *Phyllostachys aureosulcata* McClure (Bambusoideae: Poaceae): Comparative genomics and phylogenetic analysis. J Asia-Pacific Biodivers. 2025;18:190–196. https://doi.org/10.1016/J.JAPB.2024.10.003.
66. Nhat Nam N, Ngoc Trai N, Phuong Thuy N, Quoc Duy L, Nguyen Tuong Van P, Nguyen TT et al. The complete chloroplast genome of *Erythrina variegata* L. (Papilionoideae, Fabaceae). Ecol Evol. 2025;15:e70838.
67. Teshome GE, Mekbib Y, Hu G, Li Z-Z, Chen J. Comparative analyses of 32 complete plastomes of Tef (*Eragrostis tef*) accessions from Ethiopia: phylogenetic relationships and mutational hotspots. PeerJ. 2020;8:e9314.
68. Yu X, Zuo L, Lu D, Lu B, Yang M, Wang J. Comparative analysis of chloroplast genomes of five *Robinia* species: Genome comparative and evolution analysis. Gene. 2019;689:141–51.
69. Yu X, Tan W, Zhang H, Gao H, Wang W, Tian X. Complete chloroplast genomes of *Ampelopsis humulifolia* and *Ampelopsis japonica*: molecular structure, comparative analysis, and phylogenetic analysis. Plants. 2019;8:410.
70. Choi IS, Jansen R, Ruhlman T. Caught in the Act: Variation in plastid genome inverted repeat expansion within and between populations of *Medicago minima*. Ecol Evol. 2020;10:12129–37.
71. Zhu A, Guo W, Gupta S, Fan W, Mower JP. Evolutionary dynamics of the plastid inverted repeat: the effects of expansion, contraction, and loss on substitution rates. New Phytol. 2016;209:1747–56.
72. Abdullah HD, Yan R, Ma M, Naz I, Ullah R, et al. Inverted repeats dynamics shape Asclepiadoideae (Apocynaceae) chloroplast genomes: effects on genome size, gene content, structural arrangement, and mutation rate. BMC Genomics. 2025;26:697.
73. Iram S, Hayat MQ, Tahir M, Gul A, Abdullah, Ahmed I. Chloroplast genome sequence of *Artemisia scoparia*: Comparative analyses and screening of mutational hotspots. Plants. 2019;8:476.
74. Wang Y, Zhang CF, Ochieng Odago W, Jiang H, Yang JX, Hu GW, et al. Evolution of 101 Apocynaceae plastomes and phylogenetic implications. Mol Phylogenet Evol. 2023;180:107688.
75. Abdullah HCL, Mehmood F, Carlsen MM, Islam M, Waheed MT, et al. Complete chloroplast genomes of Anthurium huixtlense and Pothos scandens (Pothoideae, Araceae): unique inverted repeat expansion and contraction affect rate of evolution. J Mol Evol. 2020;88:562–74.
76. Poovitha S, Stalin N, Balaji R, Parani M. Multi-locus DNA barcoding identifies *matK* as suitable marker for species identification in *Hibiscus* L. Genome. 2016;59:1150–6.
77. Li X, Yang Y, Henry RJ, Rossetto M, Wang Y, Chen S. Plant DNA barcoding: From gene to genome. Biol Rev. 2014;90:157–66.
78. Thomson LAJ, Butaud J, Braglia L, Mabberley DJ. Pacific *Hibiscus* Species (Malvaceae) in sect. *Lilibiscus*. 2. *Hibiscus kaute* sp. nov., a Missing Link from Eastern Polynesia 1. Pacific Science. 2022;76:175–96.
79. Gardner EM, Bruun-Lund S, Niissalo M, Chantarasuwan B, Clement WL, Geri C, et al. Echoes of ancient introgression punctuate stable genomic lineages in the evolution of figs. Proc Natl Acad Sci USA. 2023;120:e2222035120.
80. Lee-Yaw JA, Grassa CJ, Joly S, Andrew RL, Rieseberg LH. An evaluation of alternative explanations for widespread cytonuclear discordance in annual sunflowers (*Helianthus*). New Phytol. 2019;221:515–26.
81. Meleshko O, Martin MD, Korneliussen TS, Schröck C, Lamkowski P, Schmutz J, et al. Extensive genome-wide phylogenetic discordance is due to incomplete lineage sorting and not ongoing introgression in a rapidly radiated bryophyte genus. Mol Biol Evol. 2021;38:2750–66.
82. Feng S, Bai M, Rivas-González I, Li C, Liu S, Tong Y, et al. Incomplete lineage sorting and phenotypic evolution in marsupials. Cell. 2022;185:1646–60.
83. Herrando-Moraira S, Calleja JA, Galbany-Casals M, Garcia-Jacas N, Liu JQ, López-Alvarado J, et al. Nuclear and plastid DNA phylogeny of tribe Cardueae (Compositae) with Hyb-Seq data: A new subtribal classification and a temporal diversification framework. Mol Phylogenet Evol. 2019;137:313–32.
84. Mandel JR, Dikow RB, Siniscalchi CM, Thapa R, Watson LE, Funk VA. A fully resolved backbone phylogeny reveals numerous dispersals and explosive diversifications throughout the history of Asteraceae. Proc Natl Acad Sci U S A. 2019;116:14083–8.
85. Zhang G, Yang J, Zhang C, Jiao B, Panero JL, Cai J, et al. Nuclear phylogenomics of Asteraceae with increased sampling provides new insights into convergent morphological and molecular evolution. Plant Commun. 2024;5:100851.
86. Johnson MG, Pokorny L, Dodsworth S, Botigué LR, Cowan RS, Devault A, et al. A universal probe set for targeted sequencing of 353 nuclear genes from any flowering plant designed using k-medoids clustering. Syst Biol. 2019;68:594–606.

Publisher's Note