Scientia Horticulturae 313 (2023) 111909

Contents lists available at ScienceDirect

Scientia Horticulturae

journal homepage: www.elsevier.com/locate/scihorti

Comparative chloroplast genome analysis of *Citrus* (Rutaceae) species: Insights into genomic characterization, phylogenetic relationships, and discrimination of subgenera

Wenbo Shi [a], Weicai Song [a], Jin Liu [b], Chao Shi [a,c,*], Shuo Wang [a,*]

[a] College of Marine Science and Biological Engineering, Qingdao University of Science and Technology, Qingdao 266042, China
[b] Yunnan Institute of Tropical Crops, Xishuangbanna 666100, China
[c] Plant Germplasm and Genomics Center, Germplasm Bank of Wild Species in Southwest China, Kunming Institute of Botany, the Chinese Academy of Sciences, Kunming 650204, China

ARTICLE INFO

Keywords:
Citrus
Plastid genomics
Genomic structure
Sequence variation
Molecular markers
Phylogeny
Taxonomy

ABSTRACT

Citrus species are fruit trees with great economic value worldwide and have been domesticated and cultivated for several years. Although numerous phylogenetic taxonomy studies have been conducted since the establishment of the genus *Citrus*, the determination of phylogenetic relationships of *Citrus* species and the discrimination of *Citrus* subgenera face some challenges due to species hybridization, polyembryony, and asexual reproduction. In this study, the complete chloroplast genomes of 16 *Citrus* species were *de novo* assembled, then several unique characters were identified. For example, *infA* and *rpl22* gene losses occurred in *Citrus* species; the *trnQ*-UUG gene may have played a significant role in the evolution of *Citrus* species; the *rpoC1* gene in chloroplast genomes of *Citrus* species was under positive selection; and several highly variable loci (*trnK*-UUU–*trnQ*-UUG, *atpF–atpH*, *trnG*-GCC–*trnfM*-CAU, *accD–psaI*, *petD–rpoA*, and *rpl32–trnL*-UAG) were found in the chloroplast genomes of the *Citrus* species. These variable loci could serve as potential markers for phylogenetic studies. The genus *Citrus* was subdivided into seven subgenera based on phylogenetic analyses of the complete chloroplast genomes of *Citrus* species: *Poncirus, Fortunella, Papeda, Citron, Cephalocitrus, Aurantium*, and *Sinocitrus*. This study may help with the identification, taxonomy, use, and evolution of *Citrus* species.

1. Introduction

The *Citrus* species are fruit trees with great economic value globally since they are well known for providing high-quality fruits and spices (Sun et al., 2015). The *Citrus* species have been domesticated and cultivated for a long time and are primarily distributed in the tropical and subtropical regions (Kumar et al., 2014; Su et al., 2014; Wu et al., 2018). Though there are various *Citrus* species, most of them have a narrow geographical distribution. This has contributed to the genetic divergence between lineages being both significant and intrinsically related (Gmitter and Hu, 1990; Wali et al., 2013; Wang et al., 2018). Furthermore, numerous studies have been performed to assess the phylogenetic taxonomy of the genus *Citrus* since it was established in 1753 by botanist Linnaeus. However, the taxonomy of the genus *Citrus* is still unclear (Lu et al., 2011; Nicolosi et al., 2000; Sun et al., 2015) due to apomixis and hybrid speciation of *Citrus* species, the high frequency of bud mutation, the extensive history of cultivation, and the dearth of remaining wild *Citrus* woods (Agouillal et al., 2017; Ding et al., 2015; Froelicher et al., 2011; Herrero et al., 1996; Moore, 2001; Wu et al., 2021). The greatest disagreement in the classification of *Citrus* species lies in the delimitation of subgenera and species within the genus *Citrus* (Barrett and Rhodes, 1976). The genus *Citrus* was first subdivided into the subgenera *Papeda* and *Eucitrus* by Swingle in 1967 (Swingle, 1967). However, Tanaka divided the genus *Citrus* into the subgenera *Archicitrus* and *Metacitrus* (Tanaka, 1977), while Zeng Mian, a Chinese expert on *Citrus*, classified the genus *Citrus* into the subgenera *Papeda, Citron, Cephalocitrus, Aurantium*, and *Sinocitrus* (Zeng, 1962). In addition, Scora, Barrett, and Rhodes believed that *C. medica, C. maxima*, and *C. reticulata* are the key species of the genus *Citrus* based on the essential oil composition of *Citrus* leaves. They also pointed out that all the other *Citrus* species occurred as a result of hybridization between these three species (Barrett and Rhodes, 1976; Scora, 1975).

* Corresponding author.
E-mail addresses: chsh1111@aliyun.com (C. Shi), shuowang@qust.edu.cn (S. Wang).

https://doi.org/10.1016/j.scienta.2023.111909
Received 14 December 2022; Received in revised form 16 January 2023; Accepted 2 February 2023
Available online 8 February 2023

DNA barcodes combined with traditional morphology-based taxonomy can be used to discriminate species and determine their relationships based on standardized DNA sequences (Hebert et al., 2003). In recent years, the *matK, rps14*, and *rbcL* genes have been used as DNA barcodes to study the taxonomic and phylogenetic relationships of *Citrus* species and their relatives, thus facilitating the taxonomy of *Citrus* species (Penjor et al., 2013, 2010; Wali et al., 2013). However, the gene sequence-based phylogenetic trees have nodes with low bootstrap values. As a result, it is unclear whether they are suitable for identifying and classifying all *Citrus* species, thus cannot resolve the current controversy regarding the discrimination of the *Citrus* subgenera (Penjor et al., 2013, 2010; Wali et al., 2013).

The chloroplast (cp) genome has been proposed as a DNA super barcode by some previous studies (Jiao et al., 2019; Liu et al., 2021). The chloroplast, which performs photosynthesis in plants, contains its own genetic material (Leister, 2003). Due to its highly conserved nature, the cp genome has been widely used in phylogenetic, taxonomic and evolutionary studies (Chen et al., 2022; Jheng et al., 2012). Despite the fact that the plant cp genome is highly conserved in terms of gene sequence and gene content (Parks et al., 2009), gene loss, mutation, and pseudogenization still occur (Henriquez et al., 2020). These variants can be applied to the comparative analysis. Comparative studies of the cp genomes of various species can enhance the understanding of plant taxonomy and the evolutionary relationships among plants (Song et al., 2022c, 2022a; Zhang et al., 2022). Particularly, the cp genome is uniparental, has a simple structure, and low molecular weight (Huang et al., 2020; Palmer et al., 1988). Therefore, the cp genome is suitable for analyses of matrilineal inheritance relationships of species, such as the *Citrus* species with significant hybridization, polyembryony, and asexual reproduction (Wu et al., 2018).

Hitherto, the complete cp genomes of *Citrus* species are not available abundantly in the NCBI database. *Citrus* species still need further development of their cp genomes. Herein, 16 new cp genomes of *Citrus* species (Table S1) were sequenced and *de novo* assembled to explore the phylogenetic relationships of *Citrus* species and the discrimination of *Citrus* subgenera. To better detect the divergences of *Citrus* species, we selected 12 representative species (*C. aurantium, C. australis, C. junos, C. madurensis, C. mangshanensis, C. maxima, C. medica, C. micrantha, C. reticulata, C. sinensis, C. tachibana*, and *C. trifoliata*) among the newly assembled and previously published *Citrus* species in the NCBI to conduct the comparative analysis. These species were selected based on Swingle, Tanaka, and Zeng's taxonomic perspectives on *Citrus* species. *C. medica* (Citron), *C. maxima* (Pummelo), and *C. reticulata* (Mandarin) are the key *Citrus* species. *C. mangshanensis* (Mangshan mandarin) is the most ancient representative of wild mandarin and tangerine (Yong et al., 2006). *C. mirantha* (Micrantha) is a wild species in the subgenus *Papeda* that may have also contributed to the evolution of cultivated *Citrus* species (Nicolosi et al., 2000). *C. madurensis* (Calamondin) is a hybrid of *Fortunella* species and *C. reticulata* (Cheng et al., 2005). *C. australis* (Australian round lime) is an Australian *Citrus* species that spread from Southeast Asia to Oceania 4 million years ago (Wu et al., 2018). *C. tachibana* (Tachibana mandarin) diverged from mainland Asian mandarins about 2 million years ago and is now mainly found in Taiwan and Japan (Fang et al., 1998; Wu et al., 2018). *C. junos* (Yuzu) is a hybrid of *C. reticulata* and *C. cavaleriei* (Ichang papeda) (Ozaki et al., 1991). *C. trifoliata* (Trifoliate orange, aka *Poncirus trifoliata*) is highly similar to *Citrus* species and may belong to the genus *Citrus* (Jarrell et al., 1992; Novelli et al., 2000; Torres et al., 1985). *Citrus aurantium* (Sour orange) and *Citrus sinensis* (Sweet orange) are hybrids of *Citrus maxima* and *Citrus reticulata* (Aseel et al., 2014; Barrett and Rhodes, 1976; Mabberley, 1997). This study aimed to: (a) Characterize the cp genomes of diverse *Citrus* species; (b) Explore divergences in the cp genomes of representative *Citrus* species, and (c) Establish the phylogenetic relationships among *Citrus* species and discriminate between their various subgenera based on the complete cp genome.

2. Materials and methods

2.1. Plant samples, DNA extraction, and sequencing

To improve the cp genomic resources of *Citrus* species, 16 *Citrus* species (*C. australis, C. hindsii, C. indica, C. jambhiri, C. junos, C. keraji, C. limonia, C. madurensis, C. mangshanensis, C. micrantha, C. nobilis, C. oto, C. paradisi, C. tachibana, C. tangerina, C. tarogayo*) that had not previously been available in the NCBI database were chosen. These species comprise geographically distinct species (*C. australis, C. indica, C. keraji, C. oto, C. tachibana*, and *C. tarogayo*) (Mabberley 2004; Wu et al., 2018, 2021), wild species (*C. hindsii, C. mangshanensis* and *C. micrantha*) (Nicolosi et al., 2000; Yong et al., 2006; Wang et al., 2022), hybrid species (*C. jambhiri, C. junos, C. limonia*, and *C. paradisi*) (Ozaki et al., 1991; Federici et al., 2000; Machado et al., 2002), and cultivated species (*C. nobilis* and *C. tangerina*) (Ji et al., 2011; Cuenca et al., 2020) to make sure the phylogenetic analysis is as extensive as possible. The botanical resources of the *Citrus* species were acquired from the Plant Germplasm and Genomics Center, Kunming Institute of Botany, and the Chinese Academy of Sciences. The permission was granted to choose three representative individuals of each species. All samples were identified by the authors and collected according to local laws. The voucher samples of various *Citrus* species were deposited at the Laboratory of Evolutionary Ecology, Qingdao University of Science and Technology (QUST). About 30 g of leaves from each *Citrus* species was used as a sample. A modified high salt approach was used to isolate cp DNA, as earlier reported (Shi et al., 2012). For evaluating the quality of the extracted DNA from *Citrus* species, 1% agarose gel electrophoresis (Green and Sambrook, 1972) was performed. Illumina Hiseq 4000 platform (Illumina, San Diego, CA, USA) was used to sequence genomic DNA of various *Citrus* species at Novogene (Beijing, China) following the manufacturer's protocol. Finally, 2 × 150 bp pair-end raw reads were obtained.

2.2. De novo assembly and annotation of cp genomes

The raw sequencing data should undergo quality clipping because it may contain some low-quality data, which will improve the accuracy of the following assembly. Almost 3.6 Gb of high-quality clean reads were acquired for diverse sample after the removal of low-quality reads and adaptors using Trimmomatic v0.40 software (Bolger et al., 2014). FastQC v0.11.9 (Brown et al., 2017) was used to evaluate the newly generated high-quality clean short reads. Then, the complete cp genomes of diverse *Citrus* species were *de novo* assembled via NOVOPlasty v4.3.1 (Dierckxsens et al., 2017). Multiple k-mer parameters were adjusted for each cp genome sequence in order to achieve the best assembly outcomes. GeSeq v1.42 (Tillich et al., 2017) was utilized with default settings to annotate the complete cp genomes. CPGAVAS v2 (Shi et al., 2019) was used to correct the previous annotation results for comparison. Finally, Sequin v16 (Lehwark and Greiner, 2018) was used to manually correct the codons and gene boundaries in the cp genomes with the *Citrus reticulata* cp genome (GenBank accession NC_034290) as a reference. The 16 *de novo* assembled *Citrus* cp genomes were submitted to GenBank under the accession numbers ON065546–ON065554 and ON872190–ON872196. Chloroplot v0.2.4 was used to draw the circular visual maps (Zheng et al., 2020). The built-in coloring theme of "volcano" was chosen.

2.3. Statistical analysis of characteristics in cp genomes

Geneious v9.0.2 (Kearse et al., 2012) was used to compare and determine the basic characteristics of the cp genomes of distinct *Citrus* species. These characteristics included the length of each sequence in each region, the percentage of various cp sequences, and the GC content of various regions. The annotated Genbank files of *Citrus* species (Table S1) were used to count the type and number of genes.

2.4. Genome comparison and structural analysis

The expansion and contraction of the IR regions were compared and visualized using the SVG module in Perl based on Genbank files of *Citrus* species. The protein-coding genes were collected from *Citrus* cp genomes to assess the non-synonymous (Ka) and synonymous (Ks) substitution rates and Ka/Ks ratios. All protein-coding genes were extracted from each cp genomic sequence of *Citrus* species via Geneious v9.0.2. Other species were compared using *C. mangshanensis*. The protein-coding genes between species were aligned using MAFFT v7.25 (Katoh and Standley, 2013). The Ka and Ks substitution rates and Ka/Ks values were determined using the Simple Ka/Ks Calculator in TBtools v1.1047 (Chen et al., 2020). Based on the protein-coding genes of different *Citrus* species, we counted the frequency of codon usage and estimated relative synonymous codon usage (RSCU) using MEGA v11 (Tamura et al., 2021). TBtools v1.1047 was employed to create the heat map for the RSCU analysis. The *Citrus* species' cp genome divergences were analyzed and visualized using the mVISTA tool (Frazer et al., 2004) with *C. mangshanensis* as a reference. The mVISTA alignment program selected LAGAN, and everything else was set to default. The nucleotide variability within the cp genome sequences of representative species was evaluated using DnaSP v6.12 (Rozas et al., 2017) (window length; 800 bp, step size; 200 bp).

2.5. Phylogenetic analysis

Phylogenetic reconstruction was performed using 35 *Citrus* species, and two *Zanthoxylum* species, *Z. armatum* (NC_050250) and *Z. simulans* (NC_037482), were chosen as outgroups. Phylogenetic trees were created using complete cp genomes via phylogenetic analysis. Sequence alignments were created using MAFFT v7.25 for the construction of phylogenetic trees. Maximum likelihood (ML) and Bayesian inference (BI) methods were used to conduct phylogenetic analysis to ensure reliability. The generalized-time reversible (GTR) with invariants (I) and discrete Gamma (G) (GTR+I + G) model was shown to be the best-fitted replacement model based on Jmodeltest v2.1.10 (Darriba et al., 2012) and the Bayesian information criterion (BIC). The ML tree was built using IQ-TREE v2.1.3 (Trifinopoulos et al., 2016), with 1000 bootstrap replicates at each branch node. Other parameters were set to their default values. For the Bl analysis, MrBayes (Huelsenbeck and Ronquist, 2001) was used. The Markov Chain Monte Carlo (MCMC) method was used for 2000,000 generations. Tree samples were taken every 5000 generations. After the first 25% of the trees were deemed "burn-in" and discarded, the consensus tree was constructed from the remaining trees.

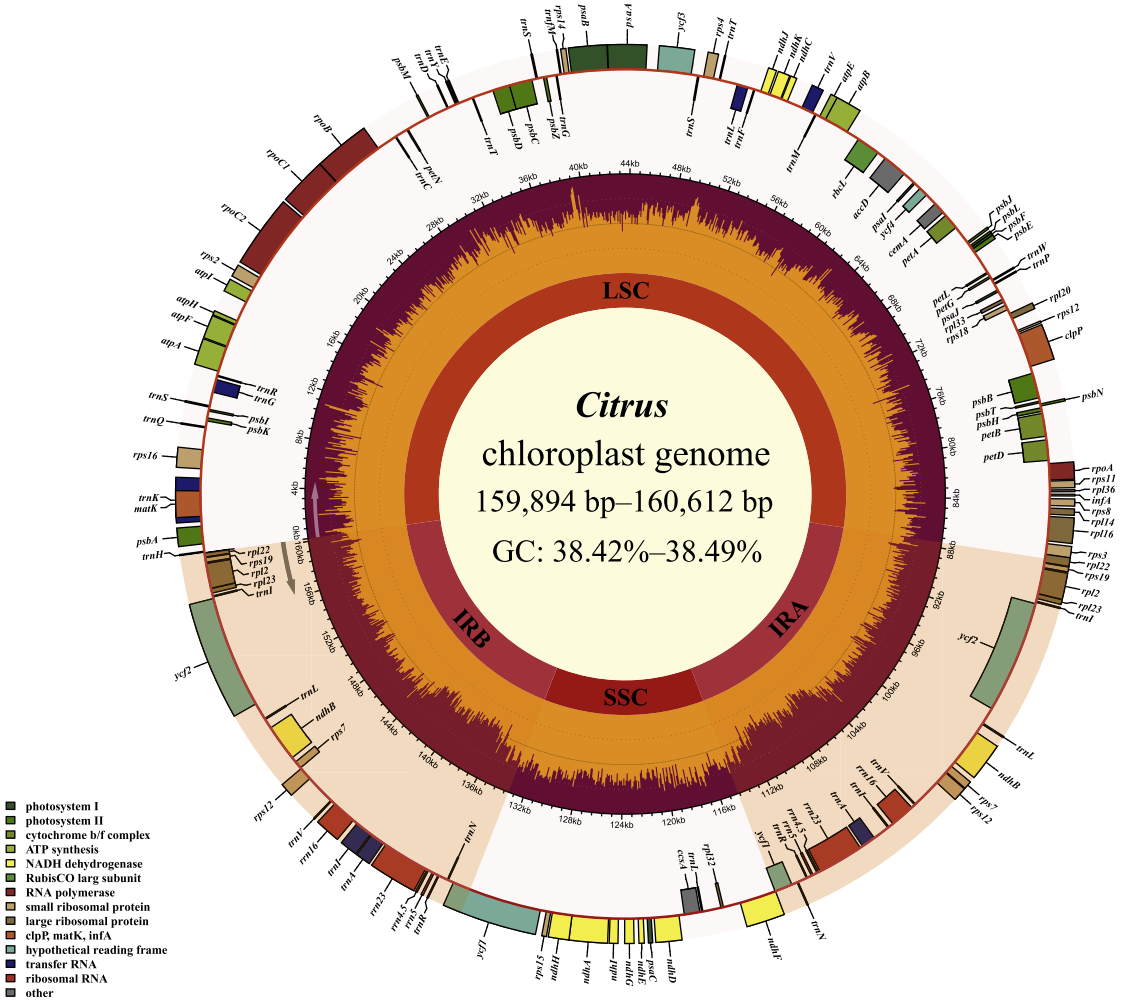

Fig. 1. The circular map of the *Citrus* chloroplast genomes. The genes are shown along the inner and outer sides of the circle. Genes on the outside of the circle are transcribed counterclockwise, and those on the inside are transcribed clockwise. Genes with different functions are shown in the bottom left corner using different colors. The inner circles are divided into dark-orange and light-orange, indicating GC and AT contents, respectively. The large single copy (LSC), the small single copy (SSC), and the inverted repeat (IRA and IRB) regions are also represented.

3. Results

3.1. Characterization of the citrus cp genomes

The cp genomes of *Citrus* species have a typical quadripartite structure found in most angiosperms (Fig. 1), with two inverted repeats (IR) regions separating the large single copy (LSC) region and small single copy (SSC) region (Palmer and Stein, 1986). The characterizations of the 12 cp genomes are shown in Table 1. *C. micrantha* and *C. junos* have the shortest and longest chloroplast genome (159,894 bp and 160,612 bp, respectively). The lengths of the IR regions for *C. australis* and *C. trifoliata* are 26,946 bp and 27,029 bp, respectively. The lengths of the LSC region for *C. micrantha* and *C. mangshanensis* are 87,149 bp and 87,852 bp, respectively. *C. aurantium* and *C. tachibana* have the shortest and longest SSC regions (18,385 bp and 18,835 bp, respectively). The *Citrus* species contained similar GC content (38.42% to 38.49%). Additionally, GC content was higher in the IR (42.91–42.91%) regions than in the LSC (36.75–36.75%) and SSC (33.09–33.38%) regions. Furthermore, 133 genes, including 88 protein-coding genes, 17 tRNA genes, and 8 rRNA genes, have been identified in the cp genomes of most *Citrus* species (Table 1). However, five species (*C. australis, C. maxima, C. medica, C. sinensis*, and *C. trifoliata*) did not contain the *infA* (transition initiation factor 1) gene. In addition, *C. tachibana* and *C. medica* do not contain one *rpl22* gene. *C. medica* and *C. aurantium* also lack one *ycf1* gene and one *rps19* gene, respectively (Table 1 and 2). A single intron was present in 15 genes (*ndhA, ndhB, petB, petD, atpF, rpl16, rpl2, rps16, rpoC1, trnA*-UGC, *trnG*-UCC, *trnI*-GAU, *trnK*-UUU, *trnL*-UAA, and *trnV*-UAC), while two introns were present in only three genes (*rps12, clpP*, and *ycf3*).

3.2. Contraction and expansion of IR regions

The SVG module in Perl was used to analyze the expansion and contraction of the IR regions of the cp genomes in 12 *Citrus* species. Gene arrangement in the cp genomes is substantially conserved in *Citrus* species. However, these chloroplast genomes are divergent at the IR/SC region junctions. The *rps3, rps22, rpl19, ycf1, ndhF, trnN, trnH*, and *psbA* genes are present at the LSC/IRb (JLB), IRb/SSC (JSB), SSC/IRa (JSA), and IRa/LSC (JLA) junctions (Fig. 2). A complete gene, *rpl22*, crosses the JLB junction in *C. australis, C. junos, C. madurensis, C. mangshanensis, C. medica, C. micrantha, C. reticulata*, and *C. tachibana*. The pseudogene *rpl22*$^{\psi}$ is located at the JLA junction in most *Citrus* species. Notably, the JLA junction in *C. medica* and *C. tachibana* lacks a pseudogene *rpl22*$^{\psi}$. However, *rpl22* is located at the JSB junction in *C. aurantium, C. maxima, C. sinensis*, and *C. trifoliata* and is constricted in the IRb region by 7–23 bp. The *ycf1* copy gene located at the JSB junction in the other *Citrus* species is absent in *C. medica*. The *trnH* gene is located at the JLA junction and is constricted in the LSC region by 0 to 63 bp. *C. sinensis* has the gene *trnH* containing 2 bp in the IRa region.

3.3. Codon usage bias

Citrus species shared similar preferences for codons, with average codons ranging from 26,474 (*C. australis*) to 27,001 (*C. trifoliata*) (Table S2). Based on the RSCU values, the 64 codons were divided into three groups (Fig. 3): two codons (UUG and AGA) with RSCU values higher than 1.81 were placed in one group, 22 codons with RSCU values higher than 1.10 were placed in another group, and the remaining codons were placed in the third group. When amino acid codons were examined, those that ended in U or A frequently had RSCU values higher than 1, while those that ended in G or C usually had RSCU values lower than 1. Additionally, the RUSC values for tryptophan (UGG) and methionine (AUG) were both 1. Based on RSCU values, the 12 representative *Citrus* species were divided into two groups: *C. medica, C. maxima, C. reticulata, C. trifoliata, C. sinensis*, and *C. aurantium* in one group; *C. mangshanensis, C. junos, C. tachibana, C. madurensis, C. micrantha*, and *C. australis* in the other (Fig. 3).

3.4. Estimation of evolutionary rates for CP genomes

The Ka, Ks, and Ka/Ks values of protein-coding genes from *Citrus* species were estimated using *C. mangshanensis* as a reference, to examine the molecular evolution of *Citrus* protein-coding genes (Fig. 4). The Ka values in *Citrus* cp genomes ranged from 0.0000 to 0.0189. The Ka values of all genes were less than 0.010 except for *ccsA, psbJ, rbcL, rpl20*, and *rpl22* genes (Fig. 4A and Table S3). The Ks values in *Citrus* cp genomes ranged from 0.0000 to 0.0496. The Ks values of all *Citrus* species were above 0.040 except for *petN* and *psbL* genes. However, only

Table 1
Summary of complete chloroplast genomes of 12 *Citrus* species.

Genome features	Accession number	Length (bp)				Gene number				GC cotent (%)			
		Full	LSC	SSC	IR	Full [unique]	PCG [unique]	tRNA [unique]	rRNA [unique]	Full	LSC	SSC	IR
C. aurantium	NC_052719	160,140	87,755	18,385	27,000	132 (113)	87 (79)	37 (30)	8 (4)	38.48%	36.80%	33.38%	42.94%
C. australis	ON872190	160,413	87,764	18,757	26,946	132 (112)	87 (78)	37 (30)	8 (4)	38.43%	36.76%	33.21%	42.97%
C. junos	ON065547	160,612	87,800	18,760	27,026	133 (113)	88 (79)	37 (30)	8 (4)	38.44%	36.81%	33.16%	42.93%
C. madurensis	ON065549	160,229	87,560	18,723	26,973	133 (113)	88 (79)	37 (30)	8 (4)	38.42%	36.75%	33.20%	42.95%
C. mangshanensis	ON065550	160,262	87,852	18,396	27,007	133 (113)	88 (79)	37 (30)	8 (4)	38.44%	36.78%	33.24%	42.93%
C. maxima	NC_034290	160,133	87,739	18,395	27,000	134 (113)	89 (79)	37 (30)	8 (4)	38.48%	36.81%	33.34%	42.95%
C. medica	NC_050939	160,031	87,476	18,573	26,991	134 (114)	89 (80)	37 (30)	8 (4)	38.44%	36.79%	33.28%	42.91%
C. micrantha	ON872194	159,894	87,149	18,763	26,991	133 (113)	88 (79)	37 (30)	8 (4)	38.44%	36.79%	33.15%	42.94%
C. reticulata	NC_034671	160,100	87,750	18,428	26,961	134 (113)	89 (79)	37 (30)	8 (4)	38.49%	36.85%	33.22%	42.96%
C. sinensis	NC_008334	160,129	87,736	18,393	27,000	132 (112)	87 (78)	37 (30)	8 (4)	38.48%	36.81%	33.34%	42.95%
C. tachibana	ON065552	160,600	87,769	18,835	26,998	132 (113)	87 (79)	37 (30)	8 (4)	38.42%	36.79%	33.09%	42.93%
C. trifoliata	NC_057088	160,260	87,442	18,760	27,029	138 (115)	91 (80)	37 (30)	8 (4)	38.44%	36.77%	33.25%	42.92%

Table 2
Genes in the chloroplast genome of *Citrus* species.

Category	Gene group	Gene name
Photosynthesis	Subunits of photosystem I	*psaA, psaB, psaC, psaI, psaJ*
	Subunits of photosystem II	*psbA, psbB, psbC, psbD, psbE, psbF, psbH, psbI, psbJ, psbK, psbL, psbM, psbN, psbT, psbZ*
	Subunits of NADH dehydrogenase	*ndhA**, *ndhB**(2), *ndhC, ndhD, ndhE, ndhF, ndhG, ndhH, ndhI, ndhJ, ndhK*
	Subunits of cytochrome b/f complex	*petA, petB**, *petD**, *petG, petL, petN*
	Subunits of ATP synthase	*atpA, atpB, atpE, atpF**, *atpH, atpI*
	Large subunit of rubisco	*rbcL*
Self-replication	Proteins of large ribosomal subunit	*rpl14, rpl16**, *rpl2**(2), *rpl20, rpl22*[a]*(2), rpl23*(2), *rpl32, rpl33, rpl36*
	Proteins of small ribosomal subunit	*rps11, rps12***(2), *rps14, rps15, rps16**, *rps18, rps19*[a](2), *rps2, rps3, rps4, rps7*(2), *rps8*
	Subunits of RNA polymerase	*rpoA, rpoB, rpoC1**, *rpoC2*
	Ribosomal RNAs	*rrn16*(2), *rrn23*(2), *rrn4.5*(2), *rrn5*(2)
	Transfer RNAs	*trnA*-UGC*(2), *trnC*-GCA, *trnD*-GUC, *trnE*-UUC, *trnF*-GAA, *trnG*-GCC, *trnG*-UCC*, *trnH*-GUG, *trnI*-CAU(2), *trnI*-GAU, *trnI*-GAU*, *trnK*-UUU*, *trnL*-CAA(2), *trnL*-UAA*, *trnL*-UAG, *trnM*-CAU, *trnN*-GUU(2), *trnP*-UGG, *trnQ*-UUG, *trnR*-ACG(2), *trnR*-UCU, *trnS*-GCU, *trnS*-GGA, *trnS*-UGA, *trnT*-GGU, *trnT*-UGU, *trnV*-GAC(2), *trnV*-UAC*, *trnW*-CCA, *trnY*-GUA, *trnfM*-CAU
Other genes	Maturase	*matK*
	Protease	*clpP***
	Envelope membrane protein	*cemA*
	Acetyl-CoA carboxylase	*accD*[a]
	c-type cytochrome synthesis gene	*ccsA*
	Translation initiation factor	*infA*
Genes of unknown function	Conserved hypothetical chloroplast ORF	*orf56*[a](2), *ycf1*[a](2), *ycf15*[a](2), *ycf2*(2), *ycf3***, *ycf4*, *ycf68*[a](2)

Notes:Gene.
* :Gene with one introns;
** :Gene with two introns; (2):Number of copies of multi-copy genes;
[a] : Gene that was missing from the chloroplast genomes of several *Citrus* species.

C. madurensis had Ks values above 0.040 for the *infA* gene (Fig. 4B and Table S3). Furthermore, 14 genes (*petG, psaI, psbE, psbF, psbI, psbM, psbN, psbT, psbZ, rpl2, rpl33, rps18, rps7, ycf1*) of the *Citrus* species had zero values for Ka and Ks (Fig. 3C and Table S3). The Ka/Ks values of 79 unique protein-coding genes ranged between 0.0 and 2.3 (Fig. 4D and Table S3). The Ka/Ks value for the *rpoC1* gene was higher than 1 in most *Citrus* species except for five (*C. madurensis, C. medica, C. reticulata, C. tachibana*, and *C. trifoliata*). The Ka/Ks values of the *matK* gene in *C. australis, C. madurensis*, and *C. micrantha* were higher than 1.0. Moreover, the Ka/Ks values of five genes (*ccsA, matK, rbcL, rpoC1*, and *ycf4*) in *C. australis* were greater than 1.0.

3.5. Comparative analysis of cp genomes

The *Citrus* cp genomes were compared using mVISTA with *C. mangshanensis* as a reference to validate the potential for divergence in the cp genomes. The cp genomes of the 12 *Citrus* species were similar (Fig. 5). Unlike the SSC and LSC regions, the IR regions of these species had less divergence. Besides, there was more divergence in non-coding regions than in coding regions. The *ndhA, ycf1, rpl22, trnI*-GAU, and *trnH*-GUG genes were more divergent than the other genes. The intergenic regions (*trnK*-UUU–*trnQ*-UUG, *atpF–atpH, trnG*-GCC–*trnfM*-CAU, *accD–psaI, petD–rpoA*, and *rpl32–trnL*-UAG) were the most divergent regions in the cp genomes.

3.6. Nucleotide diversity in citrus cp genomes

The highly variable regions were examined to measure the degree of divergence among the cp genomes of *Citrus* species at the sequence level. Nine highly variable regions (intergenic regions: *trnK*-UUU–*rps16, petN–psbM, trnT*-UGU–*trnL*-UAA, *psbL–psbE, petG–psaJ, rps3–rpl23*, and *rpl32–ndhD*) and two genes (*ndhA* and *ycf1*) were found in the cp genomes of the *Citrus* species (Fig. 6). The values of nucleotide diversity (Pi) ranged from 0 to 0.018 (average; 0.003). The Pi value in the IR regions was very low (0.001). Furthermore, the Pi value was higher in the SSC region (0.006) than in the LSC region (0.004).

3.7. Phylogenetic analysis

The ML and BI trees had similar topological structures based on the complete cp genome (Fig. 7). The nodes of the ML and BI trees demonstrated high dependability. All ML bootstrap and Bayesian posterior probability (PP) values were higher than 50 and 0.5, respectively. Results showed that *Citrus* species were monophyletic. The 35 *Citrus* species were divided into four branches: four *Citron* species (*C. medica, C. indica, C. australasica*, and *C. australis*) formed the first clade and were located at the most basal position; *Poncitrus* species (*C. trifoliata* and *C. polytrifolia*) and *Fortunella* species (*C. hindsii, C. japonica*, and *C. madurensis*) formed the second branch; Two *Papeda* species (*C. cavaleriei* and *C. junos*) and nine *Sinocitrus* species (*C. reticulata, C. sunki, C. clementina, C. tangerine, C. nobilis, C. limonia, C. jambhiri, C. tachibana*, and *C. mangshanensis*) formed the third branch; *Papeda* species (*C. micrantha, C. aurantiifolia*, and *C. hongheensis*), *Aurantium* species (*C. aurantium, C. depressa*, and *C. sinensis*), *Cephalocitrus* species (*C. maxima* and *C. paradisi*), and *Sinocitrus* species (*C. unshiu, C. platymamma, C. limon, C. erythrose, C. tarogayo, C. oto*, and *C. keraji*) formed the fourth branch.

4. Discussion

4.1. Characterization of chloroplast genomes

In this study, a comparative study was conducted using 12 representative *Citrus* species, and the essential features of the cp genomes of these *Citrus* species were also analyzed. The 12 complete cp genomes of the *Citrus* species have similar characteristics to other angiosperm cp genomes (a quadripartite structure with LSC, SSC, and two IR regions) (Wicke et al., 2011). The size of the cp genomes of the *Citrus* species ranges from 159,894 to 160,612 bp. Several reports have shown that gene loss, variation in IR regions, and variation in intergenic spacer regions significantly influence the size of the plant cp genome (Liu et al., 2019; Yang et al., 2021, 2016). A total of 132–138 cp genes were identified in the 12 *Citrus* species. Previous studies have shown that the *ycf15* and *ycf68* genes may have been pseudogenized because they do not encode proteins (Do et al., 2013; Tangphatsornruang et al., 2010). In this study, most *Citrus* species had 133 cp genes, excluding the pseudogenes *ycf15* and *ycf68* (Fig.1 and Table 1). Six *Citrus* species (*C. australis, C. maxima, C. medica, C. reticulata, C. sinensis*, and *C. trifoliata*) do not have *infA*, a gene coding a translation initiation factor. Several studies have shown that the loss of the *infA* gene from the cp genome to the nucleus occurs during angiosperm evolution (Millen et al., 2001). Herein, the *infA* gene in the genomes of six *Citrus* species

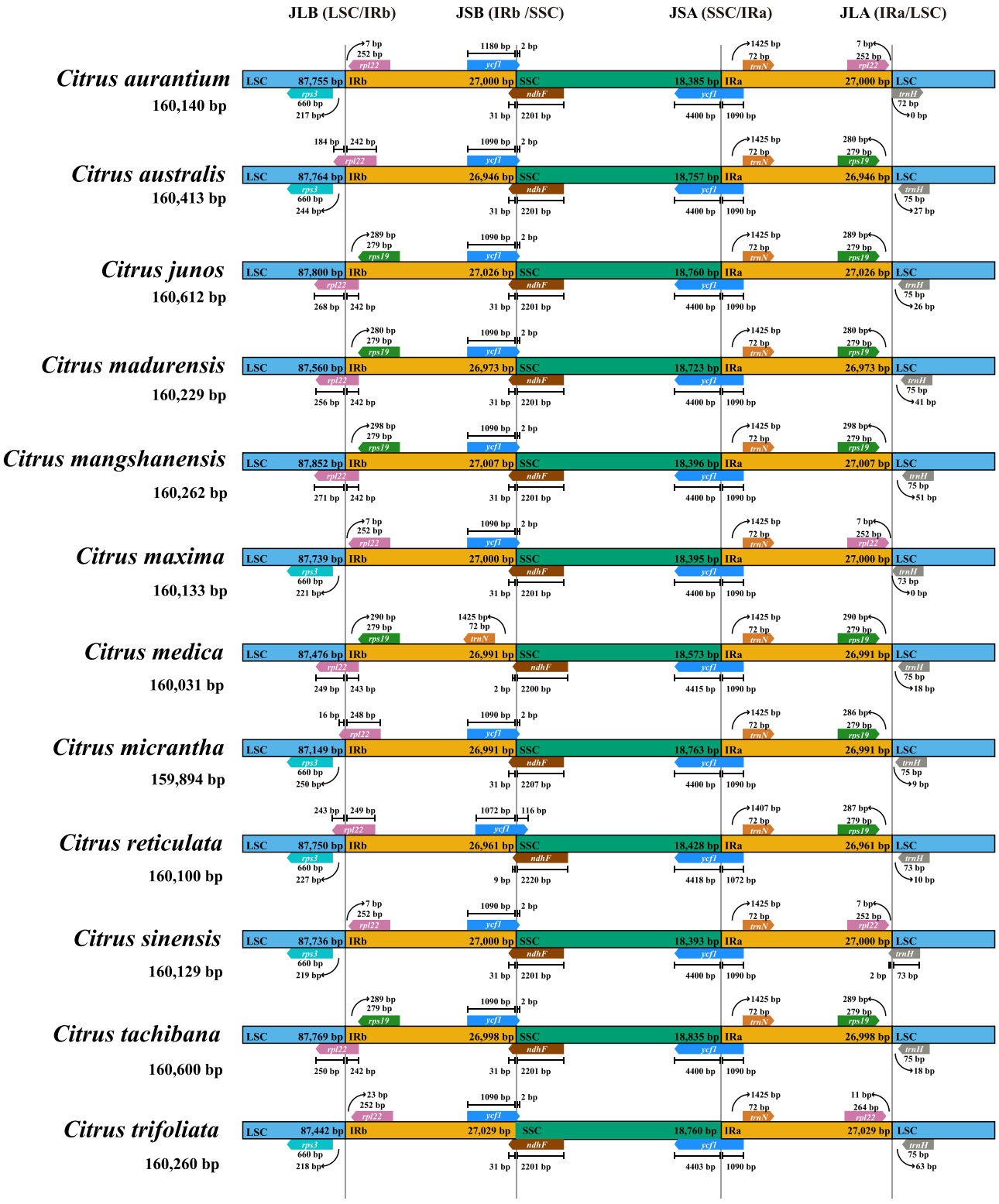

Fig. 2. The contraction and expansion of the inverted repeat/single copy (IR/SC) junctions. The junctions of the IR, short single copy (SSC), and large single copy (LSC) regions in the chloroplast genomes of *Citrus* species were compared. Loci JLB, JSB, JSA, and JLA represent the LSC/IRb, IRb/SSC, SSC/IRa, and IRa/LSC junctions, respectively.

(*C. aurantium, C. junos, C. madurensis, C. mangshanensis, C. micrantha*, and *C. tachibana*) had an internal stop codon (TAG), suggesting that the *infA* gene may be a potential pseudogene in *Citrus* species. The cp genome of *C. trifoliata* has an open reading frame (*orf56*). Furthermore, *orf56* has been reported in the cp genomes of early-diverging angiosperms, such as *Nymphaea* (Gruenstaeudl et al., 2017). Previous research found that *orf56* and the ACRS gene in the mitochondrial genome of *Citrus* have a significant degree of sequence similarity (Ohtani et al., 2002). The *orf56* appears to have been independently transferred from cp to mitochondrion in some species (Su et al., 2014).

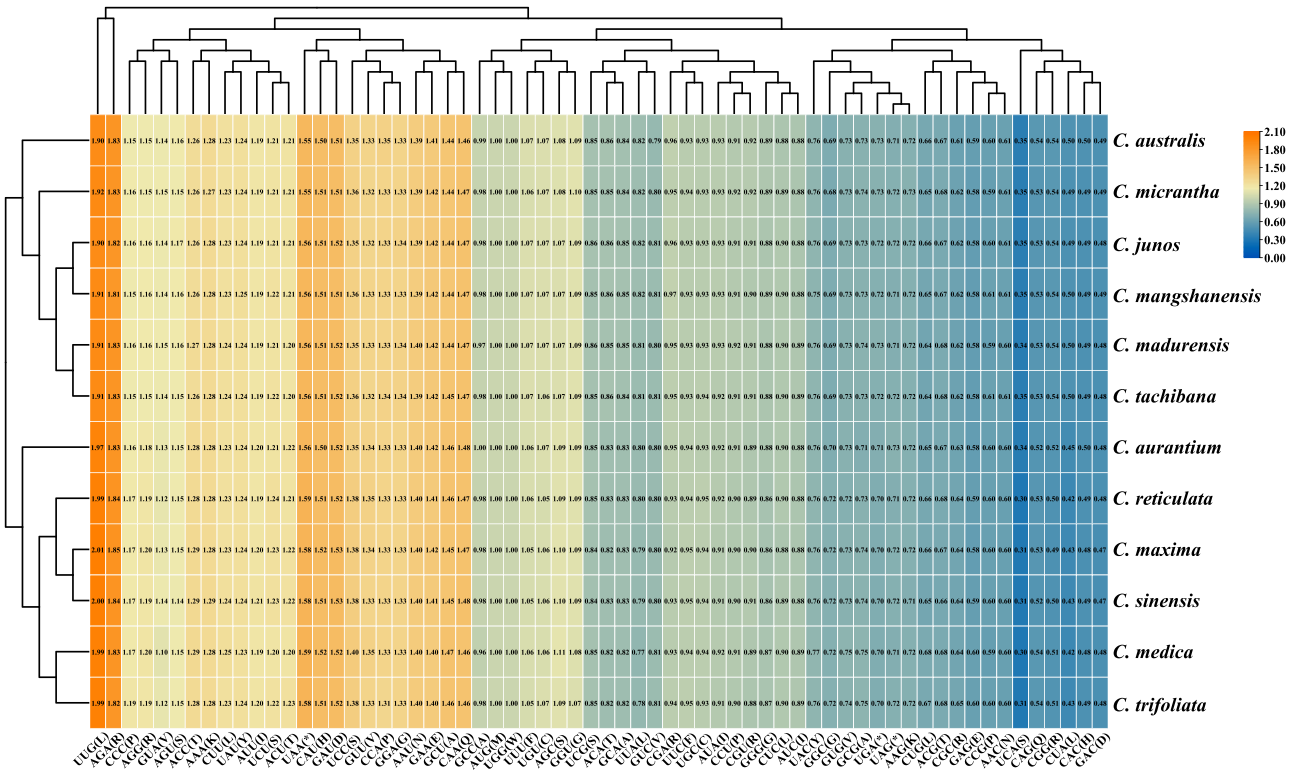

Fig. 3. The heat map for the RSCU analysis of *Citrus* species. Tree clustering was based on the RSCU values of 64 codons. The RSCU value rises as the orange color gets darker. The RSCU value falls as the blue color darkens. The various codons are represented by each column. Each different species of *Citrus* is represented by each row.

Besides, some *Citrus* species (*C. medica, C. tachibana, C. aurantium*) do not have *rpl22, rps19*, and *ycf1* genes (Table 2). These genes are found at the junctions of the IR regions. Contraction and expansion of the IR regions are frequently related to variations in the cp genome structure. Additionally, contraction and expansion can enhance the divergence of genes in the cp genome, resulting in gene loss and pseudogene formation (Li et al., 2019). Herein, visualization results (Fig. 2) show that the species had some differences, particularly in the size and location of some genes, such as *rpl22, rps19*, and *ycf1*. The variation in gene size and location changes genome size. The 12 *Citrus* species were separated into two groups based on the location of the *rpl22* gene: species where the *rpl22* gene crossed the IR/SC junctions and those where the *rpl22* gene was found in the IR regions. Herein, *C. medica* and *C. tachibana*, species whose *rpl22* gene crosses the IR/SC junctions, lost the *rpl22* copy gene. *C. aurantium* also lost the *rps19* copy gene. The gene losses occurred at the IRa/LSC junction, indicating that most *Citrus* species diverge at his region. It has been demonstrated that mutation and selection are the causes of the relative frequency of RSCU values in the cp genomes (Song et al., 2022b). The similar codon preferences indicated that during the course of evolution, *Citrus* species had established their own codon usage systems. *Citrus* species showed a preference for arginine (UUG) with the RSCU values up to 2. This suggested the *trnQ*-UUG gene may have played a significant role in the evolution of *Citrus* species. Ka, Ks, and their ratio (Ka/Ks) are widely used to evaluate gene divergence rates (Yang and Nielsen, 2000). Herein, more genes had Ka= 0 than Ks=0 (Fig. 3), indicating that synonymous genes of most protein-coding genes undergo more frequent nucleotide alterations than non-synonymous genes (Tian et al., 2021). Most genes, including *rpl22, rps19*, and *ycf1*, had Ka/Ks values less than 1, indicating that they were undergoing purifying selection. Notably, Ka/Ks value of the *rpoC1* gene was higher than 1 in most *Citrus* species. Moreover, the Ka/Ks values of genes (*ccsA, matK, rbcL, rpoC1*, and *ycf4* genes) of some *Citrus* species were higher than 1, indicating that these genes were under positive selection and thus may enhance the adaption of *Citrus* species to their environment.

4.2. Identification of highly variable regions

The mVISTA results (Fig. 4) showed that there was less variation in the IR regions than in the LSC and SSC regions, and there were more variation sites in the non-coding regions than in the coding regions. The GC content was higher in the IR regions than in the LSC and SSC regions. A previous study showed that sequence stability is positively related to GC content (Zhang et al., 2011). Herein, variations occurred in *ndhA, ycf1, rpl22, trnI*-GAU, and *trnH*-GUG genes, and the intergenic regions (*trnK*-UUU–*trnQ*-UUG, *atpF–atpH, trnG*-GCC–*trnfM*-CAU, *accD–psaI, petD–rpoA*, and *rpl32–trnL*-UAG). Nucleotide diversity of chloroplast genomes of the *Citrus* species was about 0.003 (Fig. 5). The *Citrus* species had a slightly higher average Pi value than other reported species (Wang et al., 2021). Besides, at least nine highly variable regions that could serve as powerful DNA barcodes for *Citrus* species were identified in the *Citrus* cp genomes. The highly variable regions included the intergenic regions (*trnK*-UUU–*rps16, petN–psbM, trnT*-UGU–*trnL*-UAA, *psbL–psbE, petG–psaJ, rps3–rpl23*, and *rpl32–ndhD*) and two genes (*ndhA* and *ycf1*) with Pi values above 0.01. Among these highly variable regions, *trnG*-GCC–*trnfM*-CAU, *rpl32–trnL*-UAG, and *rps3–rpl23* have been identified in previous studies (Su et al., 2014). Furthermore, *ycf1* (Dong et al., 2015), *ndhA, psbL–psbE* (Huang et al., 2019), and *rpl32-trnL*-UAG (Xiang et al., 2022) have been used as suitable plant barcodes. However, further research should be performed to further assess whether the variable regions are suitable for *Citrus* species delineation and identification.

4.3. Phylogenetic analysis

Previous studies could not clearly clarify the phylogenetic relationship of *Citrus* species because of the inadequate number of polymorphic

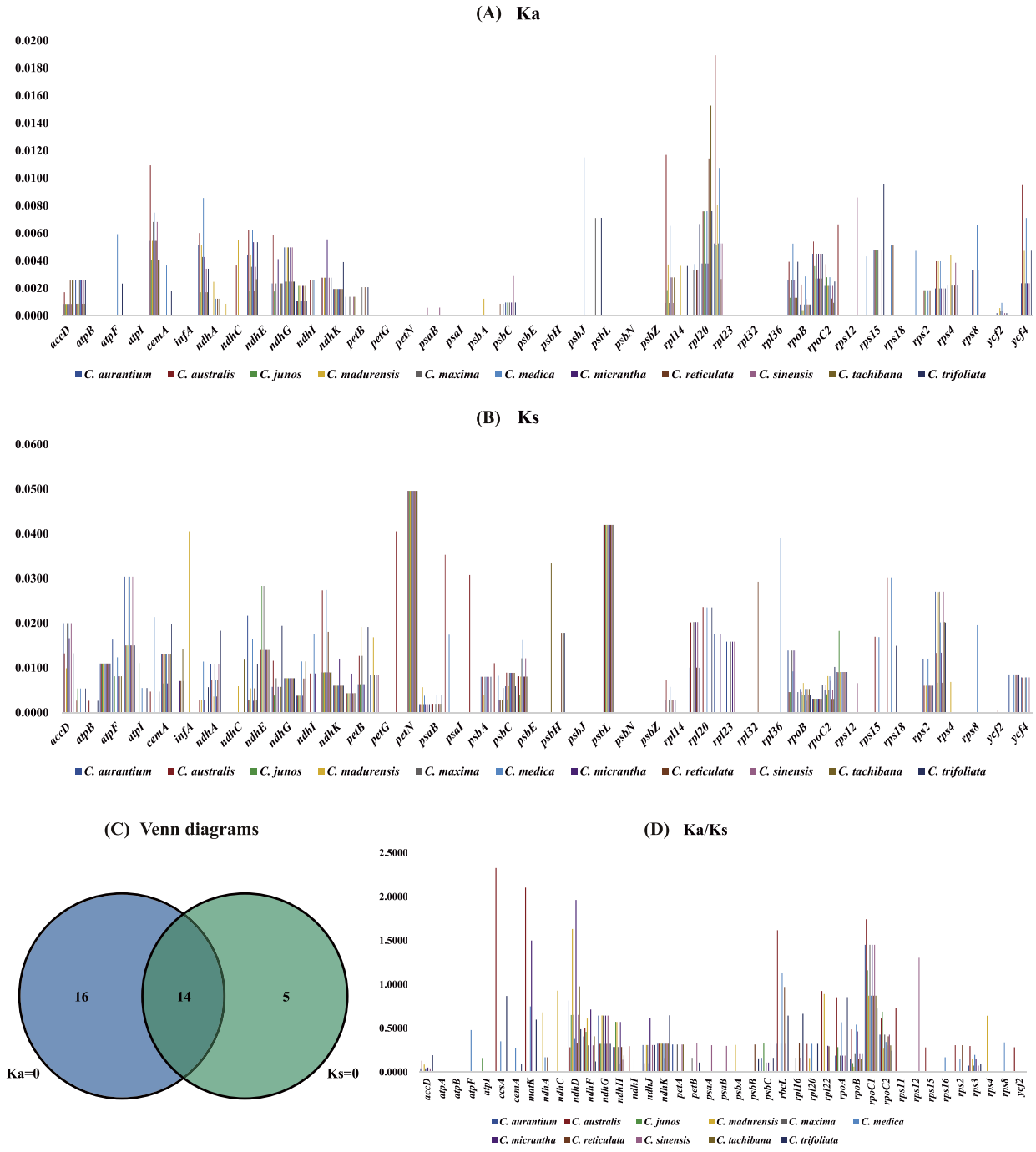

Fig. 4. Estimation of evolutionary rates for cp genomes. (A) Ka value statistics. (B) Ks value statistics. (C) Venn diagrams with Ka = 0 and Ks = 0. (D) Part Ka/Ks value statistics.

sites (Penjor et al., 2013, 2010; Wali et al., 2013). The complete cp genome sequence contains enough informative loci that can be used to assess phylogenetic relationships (Jansen et al., 2007, 2006; Song et al., 2022b) and thus is suitable for delimitation analysis of *Citrus* subgenera. Herein, the nodes of ML and BI trees showed high dependability (Fig. 6), suggesting that the complete cp genome can be used as a super barcode to identify *Citrus* species. The phylogenetic trees exhibited similar topological structures and showed that the *Citrus* species were monophyletic, similar to previous research (Lu et al., 2011; Sun et al., 2015). The phylogenetic trees were divided into four branches and subsequently separated into eight sub-branches.

The first major branch of phylogenetic trees had to sub-branches: *C. medica* (Citron) and *C. indica* (Indian wild orange), native to southwestern China, northeastern India, and northern Myanmar formed the first sub-branch (Hynniewta et al., 2014; Yang et al., 2015). *C. medica*, was at the root of the phylogenetic trees. Earlier studies have shown that *C. medica* can act as the male parent during the hybridization of numerous *Citrus* species (Curk et al., 2016; Nicolosi et al., 2000). *C. indica* is a hybrid between a citron and a mandarin (Mabberley, 2004). However, *C. indica* cannot be categorized as a mandarin based on RFLP

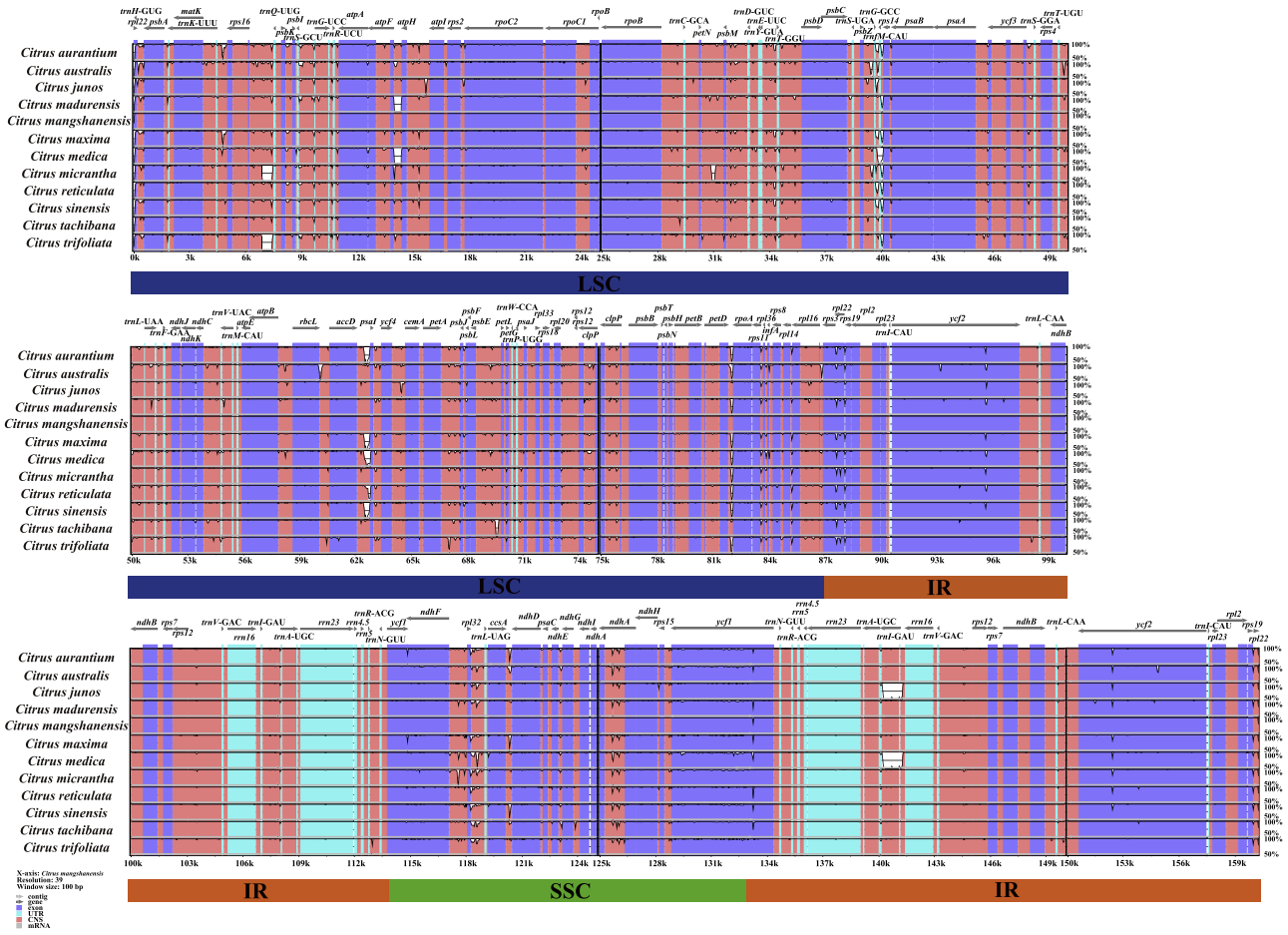

Fig. 5. Analysis of the whole chloroplast genomes of *Citrus* species. The sequence analysis of the *Citrus* species was performed with mVISTA using *Citrusmangshanensis* as the reference sequence. The dark gray arrows indicate the orientation of individual genes; pink bars indicate non-coding sequences (CNS); purple bars indicate exons; blue bars indicate RNA; gray bars indicate mRNA: and the *y*-axis indicates percentage identity (50%−100%).

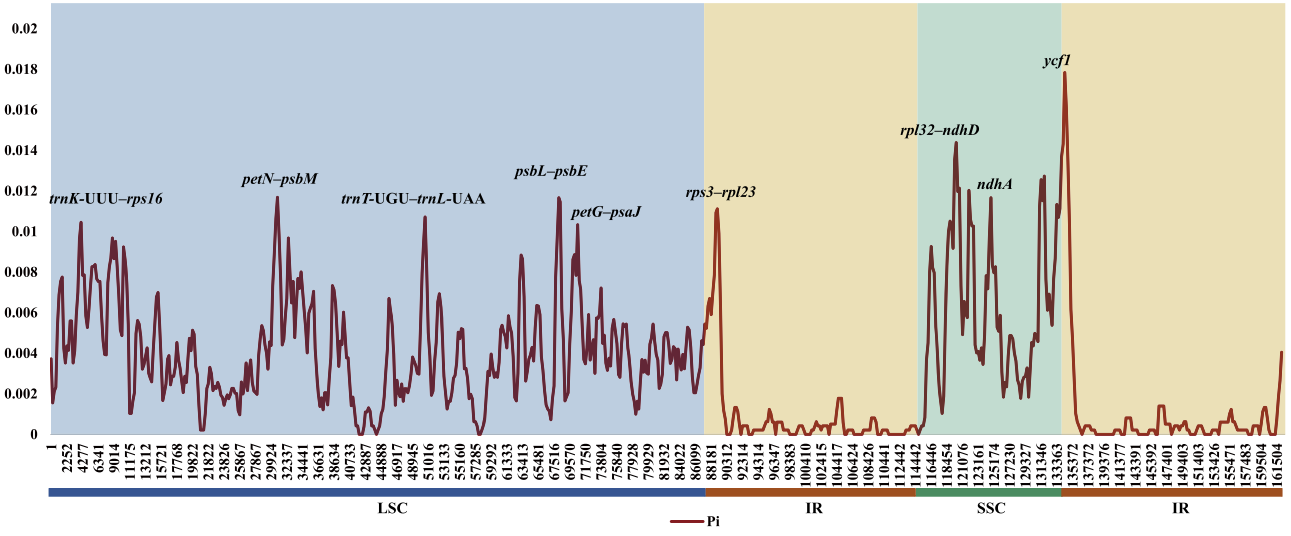

Fig. 6. The sliding window analysis of nucleotide diversity (Pi) among the whole chloroplast genomes of *Citrus* species. The *x*-axis and *y*-axis represent the base position along the sequences and Pi value, respectively.

marker analysis (Federici et al., 1998). A phylogenetic study using the complete cp genome showed that *C. indica* and *C. medica* are closely related, indicating that *C. medica* may be the female parent during the hybridization of *C. indica*. Australian citrus species (*C. australasica* (Australian finger lime) and *C. australis* (Australian round lime)) formed the second sub-branch. These results demonstrate that citron and Australian citrus species are closely related. A previous study also showed that Australian citrus species and citron can undergo nesting

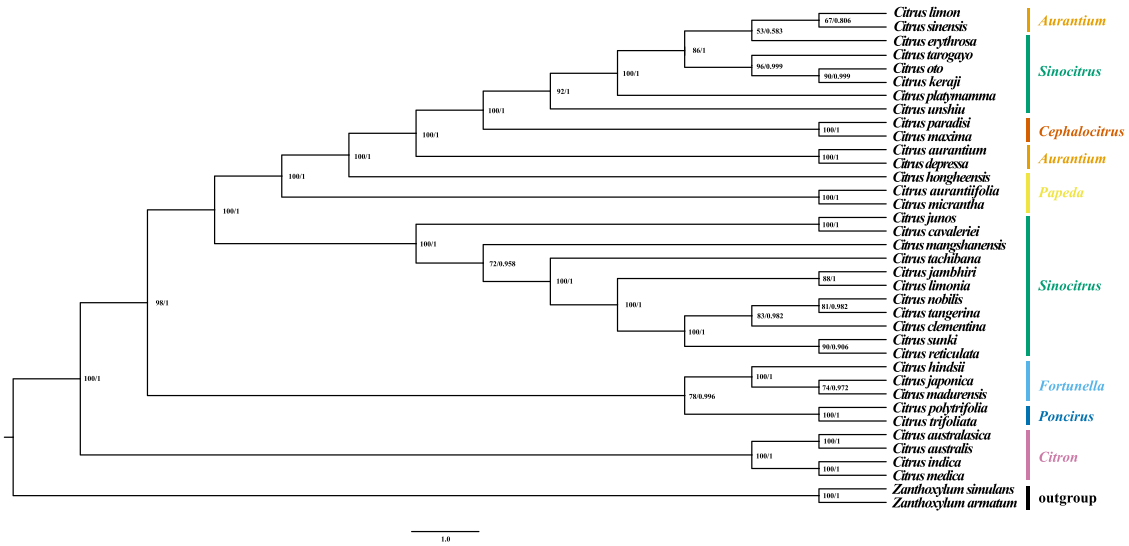

Fig. 7. Phylogenetic trees constructed using the maximum likelihood (ML) and Bayesian inference (BI) methods based on the complete chloroplast genomes of *Citrus* species. *Zanthoxylum armatum* and *Zanthoxylum simulans* were used as outgroups. The numbers above the nodes indicate support values.

(Bayer et al., 2009). Taken together, these findings indicate that the Australian citrus species belong to the *Citron* subgenus.

The second major branch of phylogenetic trees also had two sub-branches: *Poncirus* species (*C. trifoliata* (Trifoliate orange) and *C. polytrifolia* (Fuminzhi)) and *Fortunella* species (*C. japonica* (Marumi kumquat), *C. hindsii* (Hongkong kumquat), and *C. madurensis* (Calamondin)) formed the first and second sub-branch, respectively. However, research using *rbcL* molecular fragments showed that *Poncirus* and *Fortunella* should not be classified separately from the genus *Citrus* (Penjor et al., 2010). Furthermore, studies using *matK* molecular fragments classified some *Poncirus* species in the genus *Citrus* (Penjor et al., 2013). Additionally, a study based on morphological, palynological, and molecular data recommended that *Fortunella* belongs to the genus *Citrus* (Wang et al., 2022). In this study, phylogenetic analysis of the complete cp genomes showed that *Poncirus* and *Fortunella* are closely related to the genus *Citrus* and thus may belong to the subgenera of the genus *Citrus*.

The third major branch of phylogenetic trees included numerous mandarin *Citrus* species and was classified as the *Sinocitrus* subgenus. Notably, this branch included *C. cavaleriei* (Ichang papeda) and *C. junos* (Yuzu). Some studies have shown that the two *Citrus* species belong to the *Papeda* subgenus (Abkenar et al., 2008; Demarcq et al., 2021). However, the earlier isoenzyme analysis (Hirai et al., 1986), RFLP and RAPD analysis (Federici et al., 1998), and nSSR analysis (Nicolosi et al., 2000) showed that *C. cavaleriei* does not belong to the *Papeda* subgenus. In this study, phylogenetic analysis showed that *C. cavaleriei* and *C. junos* formed a sub-branch, which formed the same major branch together with the mandarin citrus subbranch. These results suggest that *C. cavaleriei* and *C. junos* are closely related to mandarin citrus and thus belong to the *Sinocitrus* subgenus. *C. reticulata* (Mandarin), one of the three basic *Citrus* species, was closely related to *C. sunki (Sunki mandarin)*. A previous study made a similar conclusion (Yoo et al., 2020). *C. sunki* may have evolved from the wild *C. reticulata* (Cornélio et al., 2003). *C. nobilis* (King mandarin), *C. tangerina* (Dancy mandarin), and *C. clementina* (Clementine mandarin) are clustered together and are closest to *C. reticulata*. The three *Citrus* species are closely related and have been used as parents during hybridization (Cuenca et al., 2020). There are both wild and cultivated *C. nobilis* varieties (Ji et al., 2011; Liu et al., 2013). *C. nobilis* may be an ancestor of *C. tachibana* (Tachibana mandarin) and the intermediate type formed during hybridization between *C. cavaleriei* and *C. reticulata* (Liu et al., 2013). Phylogenetic results based on complete cp genomes showed similar results. *C. tachibana* was located between *C. cavaleriei* and *C. reticulata* in the phylogenetic trees. *C. tachibana* diverged from mainland Asian mandarins (Wu et al., 2018). Earlier studies showed that *C. mangshanensis* (Mangshan mandarin) and common mandarin are mainland Asian mandarin (Wu et al., 2021), consistent with the constructed phylogenetic trees. Although *C. mangshanensis* and *C. reticulata* belonged to the same major branch, they were separated by a great distance. Additionally, *C. limonia* (Rangpur lime) and *C. jambhiri* (Rough lemon) were clustered together. The two *Citrus* species may be varieties of *Citrus limon* (Lemon) (Federici et al., 2000). However, some studies have shown that the two *Citrus* species are hybrids of *C. medica* and *C. reticulata* (Federici et al., 2000; Gmitter and Hu, 1990). The phylogenetic tree also showed that *C. limonia* and *C. jambhiri* were in the same major branch with *C. reticulata*, indicating that *C. reticulata* may be their female parent.

The fourth major branch had a complex composition, including orange, lemon, lime, pomelo, etc. This branch was divided into two sub-branches. The first sub-branch was formed at the root of the branch by *C. micrantha* (Micrantha), *C. aurantiifolia* (Lime), and *C. hongheensis* (Honghe Papeda). *C. micrantha*, a wild species of the subgenus *Papeda*, may be the ancestor of cultivated citrus (Dugrand-Judek et al., 2015), similar to phylogenetic analysis. A previous study suggested that *C. aurantiifolia* is a hybrid of *C. medica* and *C. micrantha* (Rouiss et al., 2018). Herein, phylogenetic analysis also showed that *C. micrantha* might be the female parent of *C. aurantiifolia*. A previous study showed that *C. hongheensis*, which is endemic to China (Zhang et al., 2020), may be the original type of the subgenus *Papeda* (Liang et al., 2007). Herein, the three *Citrus* species (*C. micrantha, C. aurantiifolia*, and *C. hongheensis*) belonged to the *Papeda* subgenera. The fruit characteristics of *C. depressa* (Shiikuwasha) are similar to those of *C. sunki* (Yamamoto et al., 2017). However, the phylogenetic tree showed that *C. depressa* is distantly related to *C. sunki* while it is closely related to *C. aurantium* (Sour orange). *C. depressa* experiences a drop in flavonoid content during maturity, similar to *Aurantium* subgenus, which is represented by *C. aurantium* (Chien et al., 2022). Therefore, *C. depressa* was classified under the *Aurantium* subgenus, while *C. maxima* (Pummelo) and *C. paradisi* (Grapefruit) were clustered into the *Cephalocitrus* subgenus. *C. paradisi* may be a natural hybrid of *C. maxima* and *C. sinensis* (Sweet orange) (Machado et al., 2002). In this study, phylogenetic analysis showed that *C. maxima* is the female parent of *C. paradisi*. Bendiguangju (*C. reticulata*), in China, may be the origin of the *C. unshiu* (Satsuma mandarin), which has been cultivated in Japan for more than a century (Fujii et al., 2016). However, phylogenetic trees showed that *C. unshiu* and *C. reticulata* were distantly related, possibly due to cross-breeding

and geographic isolation. The *Sinocitrus* subgenus was divided into two sub-regions based on the botanical characteristics and geographic distribution of *Citrus* species: *Macoacrumen* and *Microacrumen*. The *Macoacrumen* subregion may be a hybrid origin (Zeng, 1962). The phylogenetic analysis showed that the grouping was reasonable. *C. platymamma* (Byungkyool) is an endemic Korean *Citrus* species with unknown origin (Jung et al., 2005). In this study, phylogenetic analysis of revealed that *C. platymamma* was the closest relative to *C. unshiu. C. unshiu* and *C. platymamma* were categorized as *Macoacrumen* subregions. The three *Citrus* species (*C. tarogayo* (Tarogayo), *C. oto* (Oto), and *C. keraji* (Keraji)) in the Ryukyu Islands were clustered together. The three *Citrus* species may have been derived from *C. nobilis* (Yamamoto et al., 2013). A recent study showed that *C. depressa* and *C. nobilis* are the parents of the three *Citrus* species (Yamamoto et al., 2021). Phylogenetic analysis showed that *C. depressa* could be the female parent of the three *Citrus* species. *C. erythrosa* (Dongjeongkyool) is a commonly cultivated *Citrus* species in Korea (Shin et al., 2022). In this study, phylogenetic trees clustered *C. erythrosa* with *C. sinensis* and *C. limon* (with a low bootstrap values). *C. limon* is a hybrid of *C. medica,* the male parent, and *C. aurantium*, the female parent. However, the phylogenetic trees showed that *C. limon* was more closely related to *C. sinensis* than any other species.

Herein, phylogenetic analysis based on the complete cp genome showed that the genus *Citrus* was divided into seven subgenera: *Poncirus, Fortunella, Papeda, Citron, Cephalocitrus, Aurantium*, and *Sinocitrus*. We are pleased to find that this result is similar to that obtained by Bayer et al. (2009). In their study, the phylogenetic analysis used more than 10,000 bp and nine cpDNA sequences. The clades' support confidence was also high. Bayer et al. divided the genus *Citrus* into two clades, the first of which included *C. medica, C. indica*, and Australian citrus species. In this study, this clade was classified into the *Citron* subgenus. In the study of Bayer et al., the most economically significant *Citrus* species and cultivated species were found in the second clade of the genus *Citrus*, which is highly similar to the current study. Bayer et al. classified the second clade into four groups: the kumquat group, the mandarin group, the lime group, and the pommelo group. However, this clade was divided into the *Poncirus, Fortunella, Papeda, Cephalocitrus, Aurantium*, and *Sinocitrus* subgenera in this study.

5. Conclusions

In this study, the complete cp genomes of 16 *Citrus* species were *de novo* assembled to explore the phylogenetic relationships of *Citrus* species and discriminate *Citrus* subgenera. The comparative analysis of representative wild, domesticated, and hybrid *Citrus* species revealed their relatively conserved chloroplast genomes. The characteristics of the cp genomes of *Citrus* species were also explored. Furthermore, *infA* and *rpl22* gene losses were common in the *Citrus* species. The IRa/LSC junction was the most diverged region of the cp genomes. The *trnQ*-UUG gene may have played a significant role in the evolution of *Citrus* species. The *rpoC1* gene in cp genomes of *Citrus* species had undergone positive selection. Several highly variable loci (*trnK*-UUU–*trnQ*-UUG, *atpF–atpH, trnG*-GCC–*trnfM*-CAU, *accD–psaI, petD–rpoA*, and *rpl32–trnL*-UAG) that could serve as potential markers for phylogenetic studies were found in the cp genomes of *Citrus* species. Taken together, these findings show that *Poncirus* and *Fortunella* can be categorized as two subgenera of the genus *Citrus* based on the ML tree and BI tree constructed using the complete cp genome. Finally, the genus *Citrus* was divided into seven subgenera: *Poncirus, Fortunella, Papeda, Citron, Cephalocitrus, Aurantium*, and *Sinocitrus*.

CRediT authorship contribution statement

Wenbo Shi: Conceptualization, Data curation, Investigation, Methodology, Project administration, Software, Visualization, Writing – review & editing, Writing – original draft. **Weicai Song:** Conceptualization, Data curation, Formal analysis, Investigation, Project administration. **Jin Liu:** Data curation, Validation. **Chao Shi:** Conceptualization, Data curation, Funding acquisition, Project administration, Supervision, Writing – original draft. **Shuo Wang:** Data curation, Funding acquisition, Project administration.

Declaration of Competing Interest

The authors declare that they have no conflict of interest associated with the work described in this manuscript.

Data availability

Data will be made available on request.

The data supporting the findings of this study are openly available in the GenBank database at https://www.ncbi.nlm.nih.gov/, under accession numbers ON065546–ON065554 and ON872190–ON872196.

Acknowledgements

This work was supported by the State Key Laboratory of Palaeobiology and Stratigraphy (No. 223123 and No. 213119), National Natural Science Foundation of China (NO. 31801022), and Shandong Province Natural Science Foundation of China (NO. ZR2019BC094).

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.scienta.2023.111909.

References

Abkenar, A.A., Isshiki, S., Matsumoto, R., Tashiro, Y., 2008. Comparative analysis of organelle DNAs in acid citrus grown in Japan using PCR-RFLP method. Genet. Resour. Crop Evol. 55, 487–492. https://doi.org/10.1007/s10722-007-9253-2.

Agouillal, F., Taher, M., Moghrani, Z., Nasrallah, H., El Enshasy, N., 2017. A review of genetic taxonomy, biomolecules chemistry and bioactivities of citrus hystrix DC. Biosci. Biotechnol. Res. Asia 14, 285–305. https://doi.org/10.13005/bbra/2446.

Aseel, K.A.A., Nantawan, K., Talib, A.A.K., Athia, A.M., Sahapat, B., Pimwadee, P., Theerakulpisut, P., 2014. Genetic diversity of citrus (Rutaceae) in Iraq based on random amplified polymorphic DNA (RAPD) markers. African J. Agric. Res. 9, 1012–1019. https://doi.org/10.5897/ajar2013.8306.

Barrett, H.C., Rhodes, A.M., 1976. A Numerical Taxonomic Study of Affinity Relationships in Cultivated *Citrus* and Its Close Relatives. Syst. Bot. 1, 105–136. https://doi.org/10.2307/2418763.

Bayer, R.J., Mabberley, D.J., Morton, C., Miller, C.H., Sharma, I.K., Pfeil, B.E., Rich, S., Hitchcock, R., Sykes, S., 2009. A molecular phylogeny of the orange subfamily (Rutaceae: aurantioideae) using nine cpDNA sequences. Am. J. Bot. 96, 668–685. https://doi.org/10.3732/ajb.0800341.

Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114–2120. https://doi.org/10.1093/bioinformatics/btu170.

Brown, J., Pirrung, M., Mccue, L.A., 2017. FQC Dashboard: integrates FastQC results into a web-based, interactive, and extensible FASTQ quality control tool. Bioinformatics 33, 3137–3139. https://doi.org/10.1093/bioinformatics/btx373.

Chen, C.J., Chen, H., Zhang, Y., Thomas, H.R., Frank, M.H., He, Y., Xia, R., 2020. TBtools: an integrative toolkit developed for interactive analyses of big biological data. Mol. Plant 13, 1194–1202. https://doi.org/10.1016/j.molp.2020.06.009.

Chen, Q., Hu, H., Zhang, D., 2022. DNA barcoding and phylogenomic analysis of the genus *Fritillaria* in China based on complete chloroplast genomes. Front. Plant Sci. 13, 1–16. https://doi.org/10.3389/fpls.2022.764255.

Cheng, Y.J., De Vicente, M.C., Meng, H.J., Guo, W.W., Tao, N.G., Deng, X.X., 2005. A set of primers for analyzing chloroplast DNA diversity in *Citrus* and related genera. Tree Physiol 25, 661–672. https://doi.org/10.1093/treephys/25.6.661.

Chien, W.J., Saputri, D.S., Lin, H.Y., 2022. Valorization of Taiwan's *Citrus depressa* Hayata peels as a source of nobiletin and tangeretin using simple ultrasonic-assisted extraction. Curr. Res. Food Sci. 5, 278–287. https://doi.org/10.1016/j.crfs.2022.01.013.

Cornélio, M.T.M.N., Figueirôa, A.R.S., Santos, K.G.B., Carvalho, R., Soares Filho, W.S., Guerra, M., 2003. Chromosomal relationships among cultivars of *Citrus* reticulata Blanco, its hybrids and related species. Plant Syst. Evol. 240, 149–161. https://doi.org/10.1007/s00606-003-0012-3.

Cuenca, J., García-Lor, A., Juárez, J., Pina, J.A., Navarro, L., Aleza, P., 2020. Alborea: a new mid-late mandarin triploid hybrid [(*Citrus clementina* × *C. tangerina*) ×

(*C. nobilis* × *C. deliciosa*)]. HortScience 55, 1387–1392. https://doi.org/10.21273/HORTSCI15109-20.
Curk, F., Ollitrault, F., Garcia-Lor, A., Luro, F., Navarro, L., Ollitrault, P., 2016. Phylogenetic origin of limes and lemons revealed by cytoplasmic and nuclear markers. Ann. Bot. 117, 565–583. https://doi.org/10.1093/aob/mcw005.
Darriba, D., Taboada, G.L., Doallo, R., Posada, D., 2012. JModelTest 2: more models, new heuristics and parallel computing. Nat. Methods 9, 772. https://doi.org/10.1038/nmeth.2109.
Demarcq, B., Cavailles, M., Lambert, L., Schippa, C., Ollitrault, P., Luro, F., 2021. Characterization of Odor-Active Compounds of Ichang Lemon (*Citrus wilsonii* Tan.) and Identification of Its Genetic Interspecific Origin by DNA Genotyping. J. Agric. Food Chem. 69, 3175–3188. https://doi.org/10.1021/acs.jafc.0c07894.
Dierckxsens, N., Mardulyn, P., Smits, G., 2017. NOVOPlasty: de novo assembly of organelle genomes from whole genome data. Nucleic Acids Res. 45, 1–9. https://doi.org/10.1093/nar/gkw955.
Ding, Y., Zhang, J., Lu, Y., Lin, E., Lou, L., Tong, Z., 2015. Development of EST-SSR markers and analysis of genetic diversity in natural populations of endemic and endangered plant *Phoebe chekiangensis*. Biochem. Syst. Ecol. 63, 183–189. https://doi.org/10.1016/j.bse.2015.10.008.
Do, H.D.K., Kim, J.S., Kim, J.H., 2013. Comparative genomics of four Liliales families inferred from the complete chloroplast genome sequence of *Veratrum patulum* O. Loes. (Melanthiaceae). Gene 530, 229–235. https://doi.org/10.1016/j.gene.2013.07.100.
Dong, W., Xu, C., Li, C., Sun, J., Zuo, Y., Shi, S., Cheng, T., Guo, J., Zhou, S., 2015. ycf1, the most promising plastid DNA barcode of land plants. Sci. Rep. 5, 1–5. https://doi.org/10.1038/srep08348.
Dugrand-Judek, A., Olry, A., Hehn, A., Costantino, G., Ollitrault, P., Froelicher, Y., Bourgaud, F., 2015. The distribution of coumarins and furanocoumarins in Citrus species closely matches Citrus phylogeny and reflects the organization of biosynthetic pathways. PLoS ONE 10, e0142757. https://doi.org/10.1371/journal.pone.0142757.
Fang, D., Krueger, R.R., Roose, M.L., 1998. Phylogenetic relationships among selected *Citrus* Germplasm accessions revealed by inter-simple sequence repeat (ISSR) markers. J. Am. Soc. Hortic. Sci. https://doi.org/10.21273/JASHS.123.4.612.
Federici, C.T., Fang, D.Q., Scora, R.W., Roose, M.L., 1998. Phylogenetic relationships within the genus *Citrus* (Rutaceae) and related genera as revealed by RFLP and RAPD analysis. Theor. Appl. Genet. 96, 812–822. https://doi.org/10.1007/s001220050807.
Federici, C.T., Roose, M.L., Scora, R.W., 2000. RFLP analysis of the origin of *Citrus bergamia, Citrus jambhiri*, and *Citrus limonia*. Acta Hortic 535, 55–62. https://doi.org/10.1007/s001220050807.
Frazer, K.A., Pachter, L., Poliakov, A., Rubin, E.M., Dubchak, I., 2004. VISTA: computational tools for comparative genomics. Nucleic Acids Res 32, 273–279. https://doi.org/10.1093/nar/gkh458.
Froelicher, Y., Mouhaya, W., Bassene, J.B., Costantino, G., Kamiri, M., Luro, F., Morillon, R., Ollitrault, P., 2011. New universal mitochondrial PCR markers reveal new information on maternal citrus phylogeny. Tree Genet. Genomes 7, 49–61. https://doi.org/10.1007/s11295-010-0314-x.
Fujii, H., Ohta, S., Nonaka, K., Katayose, Y., Matsumoto, T., Endo, T., Yoshioka, T., Omura, M., Shimada, T., 2016. Parental diagnosis of satsuma mandarin (*Citrus unshiu* Marc.) revealed by nuclear and cytoplasmic markers. Breed. Sci. 66, 683–691. https://doi.org/10.1270/jsbbs.16060.
Gmitter, F.G., Hu, X., 1990. The possible role of Yunnan, China, in the origin of contemporary *Citrus* species (rutaceae). Econ. Bot. 44, 267–277. https://doi.org/10.1007/BF02860491.
Green, M.R., Sambrook, J., 1972. Agarose gel electrophoresis. Cold Spring Harb. Protoc. 124, 7–19. https://doi.org/10.1101/pdb.prot100404.
Gruenstaeudl, M., Nauheimer, L., Borsch, T., 2017. Plastid genome structure and phylogenomics of Nymphaeales: conserved gene order and new insights into relationships. Plant Syst. Evol. 303, 1251–1270. https://doi.org/10.1007/s00606-017-1436-5.
Hebert, P.D.N., Cywinska, A., Ball, S.L., DeWaard, J.R., 2003. Biological identifications through DNA barcodes. Proc. R. Soc. B Biol. Sci. 270, 313–321. https://doi.org/10.1098/rspb.2002.2218.
Henriquez, C.L., Abdullah, Ahmed, I., Carlsen, M.M., Zuluaga, A., Croat, T.B., McKain, M. R., 2020. Evolutionary dynamics of chloroplast genomes in subfamily Aroideae (Araceae). Genomics 112, 2349–2360. https://doi.org/10.1016/j.ygeno.2020.01.006.
Herrero, R., Asíns, M.J., Carbonell, E.A., Navarro, L., 1996. Genetic diversity in the orange subfamily Aurantioideae. I. Intraspecies and intragenus genetic variability. Theor. Appl. Genet. 92, 599–609. https://doi.org/10.1007/BF00224564.
Hirai, M., Kozaki, I., Kajiura, I., 1986. Isozyme analysis and phylogenic relationship of citrus. Japanese J. Breed. 36, 377–389. https://doi.org/10.1270/jsbbs1951.36.377.
Huang, S., Ge, X., Cano, A., Salazar, B.G.M., Deng, Y., 2020. Comparative analysis of chloroplast genomes for five *Dicliptera* species (Acanthaceae): molecular structure, phylogenetic relationships, and adaptive evolution. PeerJ 2020, e8450. https://doi.org/10.7717/peerj.8450.
Huang, Y., Li, Z., Wang, C., Zou, C., Wen, W., Shao, J., Zhu, X., 2019. *psbE-psbL* and *ndhA* Intron, The Promising Plastid DNA barcode of Fagopyrum. Int. J. Mol. Sci. 20, 3455. https://doi.org/10.3390/ijms20143455.
Huelsenbeck, J.P., Ronquist, F., 2001. MRBAYES: bayesian inference of phylogenetic trees. Bioinformatics 17, 754–755. https://doi.org/10.1093/bioinformatics/17.8.754.
Hynniewta, M., Malik, S.K., Rao, S.R., 2014. Genetic diversity and phylogenetic analysis of *Citrus* (L) from north-east India as revealed by meiosis, and molecular analysis of internal transcribed spacer region of rDNA. Meta Gene 2, 237–251. https://doi.org/10.1016/j.mgene.2014.01.008.
Jansen, R.K., Cai, Z., Raubeson, L.A., Daniell, H., Depamphilis, C.W., Leebens-Mack, J., Müller, K.F., Guisinger-Bellian, M., Haberle, R.C., Hansen, A.K., Chumley, T.W., Lee, S.B., Peery, R., McNeal, J.R., Kuehl, J.V., Boore, J.L., 2007. Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns. Proc. Natl. Acad. Sci. U. S. A. 104, 19369–19374. https://doi.org/10.1073/pnas.0709121104.
Jansen, R.K., Kaittanis, C., Saski, C., Lee, S.B., Tomkins, J., Alverson, A.J., Daniell, H., 2006. Phylogenetic analyses of *Vitis* (Vitaceae) based on complete chloroplast genome sequences: effects of taxon sampling and phylogenetic methods on resolving relationships among rosids. BMC Evol. Biol. 6, 1–14. https://doi.org/10.1186/1471-2148-6-32.
Jarrell, D.C., Roose, M.L., Traugh, S.N., Kupper, R.S., 1992. A genetic map of citrus based on the segregation of isozymes and RFLPs in an intergeneric cross. Theor. Appl. Genet. 84, 49–56. https://doi.org/10.1007/BF00223980.
Jheng, C., Chen, Tien-chih, Lin, J., Chen, Ting-chieh, Wu, W., Chang, C., 2012. Plant Science The comparative chloroplast genomic analysis of photosynthetic orchids and developing DNA markers to distinguish Phalaenopsis orchids. Plant Sci 190, 62–73. https://doi.org/10.1016/j.plantsci.2012.04.001.
Ji, Q.H., Zeng, J.W., Guo, Y.J., 2011. Using optimized random amplified polymorphic DNA (RAPD) markers to identify the category status of *Citrus nobilis* Lour. Gonggan. African J. Biotechnol. 10, 13982–13990. https://doi.org/10.5897/ajb11.1501.
Jiao, L., Lu, Y., He, T., Li, J., Yin, Y., 2019. A strategy for developing high-resolution DNA barcodes for species discrimination of wood specimens using the complete chloroplast genome of three *Pterocarpus* species. Planta 250, 95–104. https://doi.org/10.1007/s00425-019-03150-1.
Jung, Y.H., Kwon, H.M., Kang, S.H., Kang, J.H., Kim, S.C., 2005. Investigation of the phylogenetic relationships within the genus *Citrus* (Rutaceae) and related species in Korea using plastid *trnL-trnF* sequences. Sci. Hortic. (Amsterdam). 104, 179–188. https://doi.org/10.1016/j.scienta.2004.08.008.
Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. 30, 772–780. https://doi.org/10.1093/molbev/mst010.
Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P., Drummond, A., 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28, 1647–1649. https://doi.org/10.1093/bioinformatics/bts199.
Kumar, S., Malik, S.K., Uchoi, A., Chaudhury, R., Bhat, S.R., 2014. A new wild type of citron (*Citrus medica* L., Rutaceae) identified through morphology and *psbM-trnD* spacer region of chloroplast DNA. Trees 28, 1115–1124. https://doi.org/10.1007/s00468-014-1022-1.
Lehwark, P., Greiner, S., 2018. GB2sequin - a file converter preparing custom GenBank files for database submission. Genomics 111, 759–761. https://doi.org/10.1016/j.ygeno.2018.05.003.
Leister, D., 2003. Chloroplast research in the genomic age. Trends Genet 19, 47–56. https://doi.org/10.1016/S0168-9525(02)00003-3.
Li, W., Zhang, C., Guo, X., Liu, Q., Wang, K., 2019. Complete chloroplast genome of *Camellia japonica* genome structures, comparative and phylogenetic analysis. PLoS ONE 14, e0216645. https://doi.org/10.1371/journal.pone.0216645.
Liang, G., Xiong, G., Guo, Q., He, Q., Li, X., 2007. AFLP analysis and the taxonomy of *Citrus*. Acta Hortic 760, 137–142. https://doi.org/10.17660/ActaHortic.2007.760.17.
Liu, C., Jiang, D., Cheng, Y., Deng, X., Chen, F., Fang, L., Ma, Z., Xu, J., 2013. Chemotaxonomic study of citrus, poncirus and fortunella genotypes based on peel oil volatile compounds - deciphering the genetic origin of Mangshanyegan (*Citrus nobilis* Lauriro). PLoS ONE 8, e58411. https://doi.org/10.1371/journal.pone.0058411.
Liu, Y., Tahir Ul Qamar, M., Feng, J.W., Ding, Y., Wang, S., Wu, G., Ke, L., Xu, Q., Chen, L.L., 2019. Comparative analysis of miniature inverted-repeat transposable elements (MITEs) and long terminal repeat (LTR) retrotransposons in six *Citrus* species. BMC Plant Biol 19, 1–16. https://doi.org/10.1186/s12870-019-1757-3.
Liu, Z.F., Ma, H., Ci, X.Q., Li, L., Song, Y., Liu, B., Li, H.W., Wang, S.L., Qu, X.J., Hu, J.L., Zhang, X.Y., Conran, J.G., Twyford, A.D., Yang, J.B., Hollingsworth, P.M., Li, J., 2021. Can plastid genome sequencing be used for species identification in Lauraceae? Bot. J. Linn. Soc. 197, 1–14. https://doi.org/10.1093/botlinnean/boab018.
Lu, Z.hua, Zhou, Z.qin, Xie, R.jin, 2011. Molecular phylogeny of the "true citrus fruit trees" group (aurantioideae, rutaceae) as inferred from chloroplast DNA sequence. Agric. Sci. China 10, 49–57. https://doi.org/10.1016/S1671-2927(11)60306-4.
Mabberley, D.J., 2004. *Citrus* (Rutaceae): a review of recent advances in etymology, systematics and medical applications. Blumea-Biodiversity, Evol. Biogeogr. Plants 49, 481–498. https://doi.org/10.3767/000651904x484432.
Mabberley, D.J., 1997. A classification for edible *Citrus* (Rutaceae). Telopea 7, 167–172. https://doi.org/10.7751/telopea19971007.
Machado, M.A., Nunes, W.M.C., Cristofani, M., Targon, M.L.P.N., 2002. Assessment of genetic variability in grapefruits (*Citrus paradisi* Macf.) and pummelos (*C. maxima* (Burm.) Merr.) using RAPD and SSR markers. Euphytica 126, 169–176. https://doi.org/10.1023/A:1016332030738.
Millen, R.S., Olmstead, R.G., Adams, K.L., Palmer, J.D., Lao, N.T., Heggie, L., Kavanagh, T.A., Hibberd, J.M., Gray, J.C., Morden, C.W., Calie, P.J., Jermiin, L.S., Wolfe, K.H., 2001. Many parallel losses of infa from chloroplast DNA during angiosperm evolution with multiple independent transfers to the nucleus. Plant Cell 13, 645–658. https://doi.org/10.1105/tpc.13.3.645.

Moore, G.A., 2001. Oranges and lemons: clues to the taxonomy of *Citrus* from molecular markers. Trends Genet 17, 536–540. https://doi.org/10.1016/S0168-9525(01)02442-8.

Nicolosi, E., Deng, Z.N., Gentile, A., Malfa, S.La, Continella, G., Tribulato, E., 2000. *Citrus* phylogeny and genetic origin of important species as investigated by molecular markers. Theor. Appl. Genet. 100, 1155–1166. https://doi.org/10.1007/s001220051419.

Novelli, V.M., Machado, M.A., Lopes, C.R., 2000. Isoenzymatic polymorphism in *Citrus* spp. and *Poncirus trifoliata* (L.) Raf. (Rutaceae). Genet. Mol. Biol. 23, 163–168. https://doi.org/10.1590/S1415-47572000000100030.

Ohtani, K., Yamamoto, H., Akimitsu, K., 2002. Sensitivity to Alternaria alternata toxin in *Citrus* because of altered mitochondrial RNA processing. Proc. Natl. Acad. Sci. 99, 2439–2444. https://doi.org/10.1073/pnas.042448499.

Ozaki, Y., Miyake, M., Maeda, H., Ifuku, Y., Bennett, R.D., Herman, Z., Fong, C.H., Hasegawa, S., 1991. Ichangensin glucoside in *Citrus junos, Citrus sudachi* and *Citrus sphaerocarpa*. Phytochemistry 30, 2659–2661. https://doi.org/10.1016/0031-9422(91)85118-J.

Palmer, J.D., Jansen, R.K., Michaels, H.J., Chase, M.W., James, R., Palmer, D., Manhart, J.R., 1988. Chloroplast DNA variation and plant phylogeny. Ann. Missouri Bot. Gard. 75, 1180–1206.

Palmer, J.D., Stein, D.B., 1986. Conservation of chloroplast genome structure among vascular plants. Curr. Genet. 10, 823–833. https://doi.org/10.1007/BF00418529.

Parks, M., Cronn, R., Liston, A., 2009. Increasing phylogenetic resolution at low taxonomic levels using massively parallel sequencing of chloroplast genomes. BMC Biol 7, 1–17. https://doi.org/10.1186/1741-7007-7-84.

Penjor, T., Anai, T., Nagano, Y., Matsumoto, R., Yamamoto, M., 2010. Phylogenetic relationships of *Citrus* and its relatives based on *rbcL* gene sequences. Tree Genet. Genomes 6, 931–939. https://doi.org/10.1007/s11295-010-0302-1.

Penjor, T., Yamamoto, M., Uehara, M., Ide, M., Matsumoto, N., Matsumoto, R., Nagano, Y., 2013. Phylogenetic Relationships of *Citrus* and Its Relatives Based on *matK* Gene Sequences. PLoS ONE 8, 1–13. https://doi.org/10.1371/journal.pone.0062574.

Rouiss, H., Bakry, F., Froelicher, Y., Navarro, L., Aleza, P., Ollitrault, P., 2018. Origin of C. latifolia and C. aurantiifolia triploid limes: the preferential disomic inheritance of doubled-diploid "Mexican" lime is consistent with an interploid hybridization hypothesis. Ann. Bot. 121, 571–585. https://doi.org/10.1093/aob/mcx179.

Rozas, J., Ferrer-Mata, A., Sanchez-DelBarrio, J.C., Guirao-Rico, S., Librado, P., Ramos-Onsins, S.E., Sanchez-Gracia, A., 2017. DnaSP 6: DNA sequence polymorphism analysis of large data sets. Mol. Biol. Evol. 34, 3299–3302. https://doi.org/10.1093/molbev/msx248.

Scora, R.W., 1975. On the History and Origin of *Citrus*. Bull. Torrey Bot. Club 102, 369–375. https://doi.org/10.2307/2484763.

Shi, C., Hu, N., Huang, H., Gao, J., Zhao, Y.J., Gao, L.Z., 2012. An improved chloroplast DNA extraction procedure for whole plastid genome sequencing. PLoS ONE 7, e31468. https://doi.org/10.1371/journal.pone.0031468.

Shi, L., Chen, H., Jiang, M., Wang, L., Wu, X., Huang, L., Liu, C., 2019. CPGAVAS2, an integrated plastome sequence annotator and analyzer. Nucleic Acids Res. 47, W65–W73. https://doi.org/10.1093/nar/gkz345.

Shin, S.C., Song, J.H., Yoo, Y.H., Lee, J.S., Kang, S.Il, Kim, H.J., Lee, H., Kim, H.B., 2022. The complete chloroplast genome sequence of a medicinal citrus landrace, *Citrus erythrosa* Hort. ex Tanaka in Jeju Island. Korea. Mitochondrial DNA Part B Resour. 7, 580–582. https://doi.org/10.1080/23802359.2022.2057243.

Song, W., Chen, Z., He, L., Feng, Q., Zhang, H., Du, G., Shi, C., 2022a. Comparative Chloroplast Genome Analysis of Wax Gourd (*Benincasa hispida*) with Three Benincaseae Species, Revealing. Genes (Basel) 13, 461. https://doi.org/10.3390/genes13030461.

Song, W., Chen, Z., Shi, W., Han, W., Feng, Q., Shi, C., Engel, M.S., Wang, S., 2022b. Comparative analysis of complete chloroplast genomes of nine species of *Litsea* (Lauraceae): hypervariable regions, positive selection, and phylogenetic relationships. Genes (Basel) 13, 1550. https://doi.org/10.3390/genes13091550.

Song, W., Ji, C., Chen, Z., Cai, H., Wu, X., Shi, C., Wang, S., 2022c. Comparative Analysis the complete chloroplast genomes of Nine *Musa* Species: genomic features, comparative analysis, and phylogenetic implications. Front. Plant Sci. 13, 1–15. https://doi.org/10.3389/fpls.2022.832884.

Su, H.J., Hogenhout, S.A., Al-Sadi, A.M., Kuo, C.H., 2014. Complete chloroplast genome sequence of omani lime (*Citrus aurantiifolia*) and comparative analysis within the rosids. PLoS ONE 9, e113049. https://doi.org/10.1371/journal.pone.0113049.

Sun, Y.L., Kang, H.M., Han, S.H., Park, Y.C., Hong, S.K., 2015. Taxonomy and phylogeny of the genus citrus based on the nuclear ribosomal dna its region sequence. Pakistan J. Bot. 47, 95–101. https://doi.org/10.1371/journal.pone.0113049.

Swingle, W.T., 1967. The botany of *Citrus* and its wild relatives. Citrus Industry 190–430.

Tanaka, T., 1977. Fundamental discussion of *Citrus* classification. Studia Citrologica 14, 1–6.

Tamura, K., Stecher, G., Kumar, S., 2021. MEGA11: molecular Evolutionary Genetics Analysis Version 11. Mol. Biol. Evol. 38, 3022–3027. https://doi.org/10.1093/molbev/msab120.

Tangphatsornruang, S., Sangsrakru, D., Chanprasert, J., Uthaipaisanwong, P., Yoocha, T., Jomchai, N., Tragoonrung, S., 2010. The chloroplast genome sequence of mungbean (*Vigna radiata*) determined by high-throughput pyrosequencing: structural organization and phylogenetic relationships. DNA Res. 17, 11–22. https://doi.org/10.1093/dnares/dsp025.

Tian, S., Lu, P., Zhang, Z., Wu, J.Q., Zhang, H., Shen, H., 2021. Chloroplast genome sequence of Chongming lima bean (Phaseolus lunatus L.) and comparative analyses with other legume chloroplast genomes. BMC Genomics 22, 1–14. https://doi.org/10.1186/s12864-021-07467-8.

Tillich, M., Lehwark, P., Pellizzer, T., Ulbricht-Jones, E.S., Fischer, A., Bock, R., Greiner, S., 2017. GeSeq - Versatile and accurate annotation of organelle genomes. Nucleic Acids Res. 45, W6–W11. https://doi.org/10.1093/nar/gkx391.

Torres, A.M., Mau-lastovicka, T., Williams, T.E., Soost, R.K., 1985. Segregation distortion and linkage of Citrus and Poncirus isozyme genes. J. Hered. 76, 289–294. https://doi.org/10.1093/oxfordjournals.jhered.a110094.

Trifinopoulos, J., Nguyen, L.T., von Haeseler, A., Minh, B.Q., 2016. W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. Nucleic Acids Res. 44, W232–W235. https://doi.org/10.1093/NAR/GKW256.

Wali, S., Munir, F., Mahmood, T., 2013. Phylogenetic studies of selected *Citrus* species based on chloroplast gene, *rps14*. Int. J. Agric. Biol. 15, 357–361.

Wang, L., He, F., Huang, Y., He, J., Yang, S., Zeng, J., Deng, C., Jiang, X., Fang, Y., Wen, S., Xu, R., Yu, H., Yang, X., Zhong, G., Chen, C., Yan, X., Zhou, C., Zhang, H., Xie, Z., Larkin, R.M., Deng, X., Xu, Q., 2018. Genome of Wild Mandarin and Domestication History of Mandarin. Mol. Plant 11, 1024–1037. https://doi.org/10.1016/j.molp.2018.06.001.

Wang, T., Chen, L.L., Shu, H.-J., You, F., Liang, X.L., Li, J., Ren, J., Wanga, V.O., Mutie, F.M., Cai, X.Z., Liu, K.M., Hu, G.W., 2022. *Fortunella venosa* (Champ. ex Benth.) C. C. Huang and *F. hindsii* (Champ. ex Benth.) swingle as independent species: evidence from morphology and molecular systematics and taxonomic revision of *Fortunella* (Rutaceae). Front. Plant Sci. 13, 867659 https://doi.org/10.3389/fpls.2022.867659.

Wang, T., Kuang, R.P., Wang, X.H., Liang, X.L., Wanga, V.O., Liu, K.M., Cai, X.Z., Hu, G. W., 2021. Complete chloroplast genome sequence of *Fortunella venosa* (Champ. ex benth.) C.C.Huang (Rutaceae): comparative analysis, phylogenetic relationships, and robust support for its status as an independent species. Forests 12, 996. https://doi.org/10.3390/f12080996.

Wicke, S., Schneeweiss, G.M., DePamphilis, C.W., Müller, K.F., Quandt, D., 2011. The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. Plant Mol. Biol. 76, 273–297. https://doi.org/10.1007/s11103-011-9762-4.

Wu, G.A., Sugimoto, C., Kinjo, H., Azama, C., Mitsube, F., Talon, M., Gmitter, F.G., Rokhsar, D.S., 2021. Diversification of mandarin citrus by hybrid speciation and apomixis. Nat. Commun. 12, 1–10. https://doi.org/10.1038/s41467-021-24653-0.

Wu, G.A., Terol, J., Ibanez, V., López-García, A., Pérez-Román, E., Borredá, C., Domingo, C., Tadeo, F.R., Carbonell-Caballero, J., Alonso, R., Curk, F., Du, D., Ollitrault, P., Roose, M.L., Dopazo, J., Gmitter, F.G., Rokhsar, D.S., Talon, M., 2018. Genomics of the origin and evolution of *Citrus*. Nature 554, 311–316. https://doi.org/10.1038/nature25447.

Xiang, K.L., Mao, W., Peng, H.W., Erst, A.S., Yang, Y.X., He, W.C., Wu, Z.-Q., 2022. Organization, Phylogenetic Marker Exploitation, and Gene Evolution in the Plastome of *Thalictrum* (Ranunculaceae). Front. Plant Sci 13, 897843. https://doi.org/10.3389/fpls.2022.897843.

Yamamoto, M., Takakura, A., Tanabe, A., Teramoto, S., Kita, M., 2017. Diversity of *Citrus depressa* Hayata (Shiikuwasha) revealed by DNA analysis. Genet. Resour. Crop Evol. 64, 805–814. https://doi.org/10.1007/s10722-016-0403-2.

Yamamoto, M., Tani, K., Kozai, N., 2021. The Morphological and Genetic Characteristics of Local *Citrus* Grown on the Ryukyu Islands, Japan Fig Kyushu Fig. 1. Collection sites (islands) of local citrus accessions on the Ryukyu Islands. Trop. Agric. Dev. 65, 206–215. https://doi.org/10.11248/jsta.65.206.

Yamamoto, M., Tsuchimochi, Y., Ninomiya, T., Koga, T., Kitajima, A., Yamasaki, A., Inafuku-Teramoto, S., Yang, Xuehu, Yang, Xiaoling, Zhong, G., Nasir, N., Kubo, T., Tominaga, S., 2013. Diversity of chloroplast DNA in various mandarins (*Citrus* spp.) and other citrus demonstrated by caps analysis. J. Japanese Soc. Hortic. Sci. 82, 106–113. https://doi.org/10.2503/jjshs1.82.106.

Yang, X., Li, H., Liang, M., Xu, Q., Chai, L., Deng, X., 2015. Genetic diversity and phylogenetic relationships of citron (*Citrus medica* L.) and its relatives in southwest China. Tree Genet. Genomes 11, 1–13. https://doi.org/10.1007/s11295-015-0955-x.

Yang, X., Zhou, T., Su, X., Wang, G., Zhang, X., Guo, Q., Cao, F., 2021. Structural characterization and comparative analysis of the chloroplast genome of *Ginkgo biloba* and other gymnosperms. J. For. Res. 32, 765–778. https://doi.org/10.1007/s11676-019-01088-4.

Yang, Y., Zhou, T., Duan, D., Yang, J., Feng, L., Zhao, G., 2016. Comparative analysis of the complete chloroplast genomes of five *Quercus* species. Front. Plant Sci. 7, 1–13. https://doi.org/10.3389/fpls.2016.00959.

Yang, Z., Nielsen, R., 2000. Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. Mol. Biol. Evol. 17, 32–43. https://doi.org/10.1093/oxfordjournals.molbev.a026236.

Yong, L., Zhong-Hai, S., Yong, L., de-Chun, L., bo, wu, 2006. Genetic diversity of pummelo (*Citrus grandis* Osbeck) and its relatives based on simple sequence repeat markers. Chinese J. Agric. Biotechnol. 3, 119–126. https://doi.org/10.1079/CJB2006102.

Yoo, Y.H., Oh, C.J., Shin, S.C., Seo, S., Kim, M., Yun, S.H., Song, K.J., Lee, H., Kim, H.B., 2020. Complete chloroplast genome sequence of a medicinal landrace citrus Jinkyool (*Citrus sunki* Hort. ex Tanaka) in Jeju Island. Korea. Mitochondrial DNA Part B Resour. 5, 3719–3720. https://doi.org/10.1080/23802359.2020.1833771.

Zeng, M., 1962. The understanding and opinion of the *Citrus* classification. China Fruits 2, 31–37. Chinese.

Zhang, J., Kuo, C.C.J., Chen, L., 2011. GC content around splice sites affects splicing through pre-mRNA secondary structures. BMC Genomics 12, 1–11. https://doi.org/10.1186/1471-2164-12-90.

Zhang, Y.M., Han, L.J., Yang, C.W., Yin, Z.L., Tian, X., Qian, Z.G., Li, G.D., 2022. Comparative chloroplast genome analysis of medicinally important *Veratrum*

(Melanthiaceae) in China: insights into genomic characterization and phylogenetic relationships. Plant Divers 44, 70–82. https://doi.org/10.1016/j.pld.2021.05.004.
Zhang, Z.hao, Long, C.rui, Jiang, Y., Bei, X.jun, Wang, S., 2020. Characterization of the complete chloroplast genome of *Citrus hongheensis*, a key protected wild plant in Yunnan province of China. Mitochondrial DNA Part B Resour. 5, 3532–3533. https://doi.org/10.1080/23802359.2020.1825137.
Zheng, S., Poczai, P., Hyvönen, J., Tang, J., Amiryousefi, A., 2020. Chloroplot: an Online Program for the Versatile Plotting of. Organelle Genomes. Front. Genet. 11, 576124 https://doi.org/10.3389/fgene.2020.576124.