Li *et al. Horticulture Advances* (2024) 2:13
https://doi.org/10.1007/s44281-024-00032-9

Horticulture Advances

RESEARCH ARTICLE Open Access

Insights into chloroplast genome evolution in Rutaceae through population genomics

Chao-Chao Li[1], Yi Bao[1], Ting Hou[2], Jia-Cui Li[2], Zhi-Yao Ma[2], Nan Wang[2], Xiao-Meng Wu[1], Kai-Dong Xie[1], Yong-Feng Zhou[2] and Wen-Wu Guo[1*]

Abstract

Chloroplast genomes, pivotal for understanding plant evolution, remain unexplored in Rutaceae, a family with key perennial crops like citrus. Leveraging next-generation sequencing data from 509 Rutaceae accessions across 15 species, we conducted a de novo assembly of 343 chloroplast genomes, unveiling a chloroplast variation map highlighting the heterogeneous evolution rates across genome regions. Notably, differences in chloroplast genome size primarily originate from large single-copy and small single-copy regions. Structural variants predominantly occurred in the single-copy region, with two insertions located at the single-copy and inverted repeat region boundary. Phylogenetic analysis, principal component analysis, and population genetic statistics confirmed the cohesive clustering of different *Citrus* species, reflecting evolutionary dynamics in *Citrus* diversification. Furthermore, a close chloroplast genetic affinity was revealed among *Atalantia* (previously regarded as primitive citrus), *Clausena*, and *Murraya*. *Zanthoxylum* formed a distinct group with heightened genetic diversity. Through expanding our analysis to include 34 published chloroplast genomes, we explored chloroplast gene selection, revealing divergent evolutionary trends in photosynthetic pathways. While Photosystem I and Photosystem II exhibited robust negative selection, indicating stability, the Nicotinamide adenine dinucleotide (NADH) dehydrogenase pathway demonstrated rapid evolution, which was indicative of environmental adaptation. Finally, we discussed the effects of gene length and GC content on chloroplast gene evolution. In conclusion, our study reveals the genetic characterization of chloroplast genomes during Rutaceae diversification, providing insights into the evolutionary history of this family.

Keywords Chloroplast genome, Rutaceae, Negative selection, Diversification

Introduction

Chloroplasts, intracellular organelles commonly found in eukaryotic autotrophs, house their genome. In addition to facilitating photosynthesis, they play a crucial role in synthesizing amino acids, nucleotides, fatty acids, phytohormones, vitamins, and various metabolites (Daniell et al. 2016). The chloroplast genome in plants exhibits its unique genetic and evolutionary patterns, including maternal inheritance, a low mutation rate, and a circular, double-stranded DNA structure, which aids in revealing plant evolution and hybridization relationships.

Maternal inheritance, where all offspring genes originate from the female parent, is characteristic of most plant chloroplast genomes (Hipkins et al. 1994; Hagemann 2004). In chloroplast phylogenetic trees, hybrids typically cluster with their female parental type, allowing for quick identification of maternal origin, even in cases where the male parent is unknown.

*Correspondence:
Wen-Wu Guo
guoww@mail.hzau.edu.cn
[1] National Key Laboratory for Germplasm Innovation & Utilization of Horticultural Crops, College of Horticulture and Forestry Sciences, Huazhong Agricultural University, Wuhan 430070, People's Republic of China
[2] Shenzhen Branch, Guangdong Laboratory of Lingnan Modern Agriculture, Key Laboratory of Synthetic Biology, Ministry of Agriculture and Rural Affairs, Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen 518000, People's Republic of China

The low mutation rate of the chloroplast genome is noteworthy (Green 2011), with chloroplast gene synonymous rates in angiosperms being three times that of mitochondria. However, structural variation in the chloroplast genome is generally lower than in nuclear and mitochondrial genomes (Green 2011). Similar chloroplast genome sequences can be easily obtained through polymerase chain reaction (PCR) (Palmer 1987; Kelchner 2000; Heinze 2007). However, numerous samples cluster within the same evolutionary clade, suggesting that the chloroplast genomes of these samples exhibit limited genetic variation, which may hinder phylogenetic resolution. Therefore, chloroplast genomes are not well suited for studying the evolutionary relationships of closely related species with direct parental lineage.

The chloroplast genome structure typically comprises one large single-copy (LSC) region, one small single-copy (SSC) region, and two inverted repeat (IR) regions. The distinct structure of the IR region contributes to its lower mutation rate than the single-copy (SC) region. In early studies, partial chloroplast sequences were commonly used for phylogenetic analysis using sequences of the IR region, resulting in abnormal results. Advances in sequencing and analytical methods have enabled the widespread use of whole chloroplast genomes in phycological classification (Grivet et al. 2001; Chung and Staub 2003). For instance, a family-level phylogenetic tree for angiosperms was constructed based on chloroplast genes from 4792 plastomes representing 4660 species across 2024 genera (Li et al. 2021). Chloroplast genomes have been used to construct phylogenetic trees for many plant families (Zhang et al. 2017; Saarela et al. 2018; Ogoma et al. 2022; Yan et al. 2022).

The variability of chloroplast genomes is significantly influenced by the expansion and contraction of the IR region, a phenomenon observed in various species where the IR region can be nearly lost (Yi et al. 2013). This variability affects the frequency of gene variations, with synonymous substitution rates in the IR region being approximately 3.7-fold slower than those located in the SC region. Lower substitution rates in the IR region are linked to copy-dependent repair activity (Zhu et al. 2015). In *Plantago*, chloroplast genomes undergo sequence expansions and inversions, affecting IR region size and gene order. Genes near inversion breakpoints exhibit accelerated nucleotide substitution rates and local hypermutations associated with rearrangements (Mower et al. 2021). Essential genes related to photosynthesis are present in the chloroplast genome, and many have undergone natural selection to adapt to environmental changes during radiation evolution. For example, a genome-wide scan in rice identified 14 chloroplast orthologous genes associated with the photosynthetic system and adapted to shade tolerance or sun-loving characteristics (Gao et al. 2019). During the evolution of *Paphiopedilum*, seven chloroplast genes underwent positive selection, while in *Osmanthus*, all chloroplast genes experienced purifying selection (Li et al. 2022; Liu et al. 2022).

Research on chloroplast pan-genomes, facilitated by large-scale sequencing data, enhances our understanding of chloroplast DNA diversity and evolution across different plant species. Extensive comparative analyses of chloroplast genomes from various species or within the same species shed light on key genes or genetic variations associated with environmental adaptability (Magdy et al. 2019). Concurrently, studies on chloroplast pan-genomes offer crucial insights into the phylogenetic relationships of plants (Wang et al. 2022a, b; Zhou et al. 2023).

Rutaceae, primarily found in the tropics and subtropics, contains 154 genera and approximately 2100 species (Kubitzki 2011). Notably, subfamilies Aurantioideae and Rutoideae are of economic importance. Rutaceae species, including fruit trees (*Citrus*, *Clausena*), rootstocks (*Poncirus*), seasoning (*Zanthoxylum*), Chinese medicine (*Phellodendron*), and ornamental plants (*Murraya*), are widely utilized. Botanists historically classified Rutaceae based on phenotype, with Engler dividing it into seven subfamilies according to flower, fruit, and gland characteristics (Engler 1896; Engler 1931). Rutaceae exhibits four androecium character types: halostemony, diplostemony, obdiplostemony, and polyandry. Research on androecium characters in subfamilies Rutoideae and Aurantioideae suggests that the genus in Rutoideae with obdiplostemony serves as the ancestor genus of Aurantioideae (Wei et al. 2015).

Advancements in sequencing technology have allowed DNA-based studies on the taxonomy and evolution of Rutaceae. Chloroplast genome studies have classified several Rutaceae genera, revealing that *Cneorum*, *Ptaeroxylon*, *Spathelia*, and *Dictyoloma* form a clade sister to other Rutaceae species. Ovary and fruit characters are insufficient to accurately delineate families, as subfamilies with more than one genus (except Aurantioideae) are not monophyletic (Groppo et al. 2008). Phylogenetic analysis using nine chloroplast sequences shows that the three traditionally recognized subclades in the Aurantieae tribes are not monophyletic, with *Triphasiinae* and *Balsamocitrinae* found to be polyphyletic. Nuclear genomic studies help infer the timing of differentiation in several Rutaceae species (Bayer et al. 2009). Phylogenetic analysis of 34 chloroplast genomes from *Citrus* and related genera reveals that the genus *Citrus* underwent three radiation events. *Citrus* speciation occurred between 7.5 and 6.3 million years ago (Ma), leading to ancestors like citron (*Citrus medica*, 5.0–3.7 Ma) from Australian limes, micrantha from pummelos, and

Ichangensis from mandarins. Further radiations of *Fortunella*, sour orange, sweet orange, lemon, and mandarin occurred later (1.5–0.2 Ma) (Carbonell-Caballero et al. 2015). A 4.23 gigabases (Gb) reference genome of *Zanthoxylum*, assembled by Feng, ten times larger than that of sweet orange, shows a divergence between *Zanthoxylum* and sweet orange approximately 35 Ma ago (Feng et al. 2021). Wu analyzed 60 representative citrus species worldwide (Wu et al. 2018). The diversification of citrus species occurred in the late Miocene (6–8 Ma), spreading to Southeast Asia, likely related to the weakening of monsoon. Australian citrus diversification began later, crossing the Wallace Line (approximately 4 Ma). The hybridization history of several citrus species is systematically described. By comparing several wild and cultivated mandarin groups, two independent domestication events for cultivated mandarins in China were revealed, roughly distributed in the south and the north of the Lingnan Mountains (Wang et al. 2018).

In this study, 378 Rutaceae chloroplast genomes were assembled and annotated using next-generation sequencing (NGS) data from 509 accessions in public databases and 34 published Rutaceae chloroplast genomes. Our analysis focused on exploring phylogenetic relationship and population structure within Rutaceae by leveraging information on chloroplast DNA variations. Comparative and evolutionary analyses of gene sequences were conducted to examine structural variations and gene selection in Rutaceae chloroplast genomes. The findings contribute valuable data for population genetic analysis and evolutionary studies of related taxa.

Materials and methods

Chloroplast genome assembly and annotation

We processed 509 NGS datasets through FastQC (v0.11.9) (https://github.com/s-andrews/FastQC), followed by batch assembly using GetOrganelle (v1.7.5) (Jin et al. 2020) with parameters '-fast -k 65,105,127 -w 0.68 -t 10 -f embplant_pt'. This yielded 344 complete chloroplast genome sequences. Plastid Genome Annotator (PGA) (Qu et al. 2019) was used to annotate 378 chloroplast genomes (including 34 published chloroplast Rutaceae genomes) with GeneBank files as the database. Subsequently, OGDRAW (v1.3.1) (Greiner et al. 2019) was used to map the chloroplast genome loops of the Hong Kong kumquat. The genome lengths and GC contents were measured using SeqKit (v2.3.0) (Shen et al. 2016) and visualized using Origin2019. Protein-coding genes were extracted based on the annotated GB format chloroplast genome files. The pan-genome of chloroplast protein-encoding genes was then constructed and annotated using Bacterial Pan Genome Analysis (BPGA) (Chaudhari et al. 2016) with default parameters.

Chloroplast genome size variation analysis

A phylogenetic tree was constructed using the LSC region sequences from 378 chloroplast genomes. Subsequently, the samples were categorized into three groups based on this tree: group A (*Ruta*, *Casimiroa*, *Tetradium*, *Phellodendron*, *Zanthoxylum*, *Murraya*, *Clausena*, *Micromelum*, *Glycosmis*, *Atalantia*, and lime), group B (*Citron*, *Fortunella*, *Ichangensis*, mandarin, and *Poncirus*), and group C (sweet orange, sour orange, grapefruit, and lemon). The trends in genome lengths within the LSC, SSC, and IR regions for these three sample groups were assessed using R packages ggpmisc (Aphalo 2021) and ggpubr (Kassambara 2023).

Mapping and variant calling

Clean data from 509 NGS runs were mapped to the Hong Kong kumquat chloroplast genome using BWA (Li and Durbin 2009). SAM (sequence alignment map) files were converted to sorted BAM (binary version of SAM) files using SAMtools (Li et al. 2009). Duplicate removal was performed using Picard (http://broadinstitute.github.io/picard/). Variant call format (VCF) files were generated using DeepVariant (rc1.0.0) (Poplin et al. 2018). GVCF files from 509 accessions were merged into one VCF file using GLnexus (v1.2.7) (Yun et al. 2020). VCF files were annotated with SnpEff (v5.1) (Cingolani et al. 2012). SNP and indel densities were calculated using BEDTools (v2.30.0) (Quinlan and Hall 2010), and variation information was visualized using Circos (v0.69–8) (Krzywinski et al. 2009). Structural variants (SV) were predicted using BCFtools (v1.8) (Narasimhan et al. 2016) and DELLY (v0.9.1) (Rausch et al. 2012). VCF files were filtered *via* VCFtools (v0.1.16) using the following criteria: variant quality > 2.0, quality score > 40.0, mapping quality > 30.0, genotype calls with a depth > 2 or < 100, and < 20% missing genotypes across all samples.

Phylogeny and population genomics analyses

To explore the evolutionary relationships of Rutaceae chloroplast genomes, a maximum likelihood-based method was used to construct phylogenetic trees. SNP variation data was extracted from VCF files using VCFtools (v0.1.16) (Danecek et al. 2011). SNPs were used to construct the phylogenetic tree using IQ-TREE (v2.1.4) (Minh et al. 2020), with 1000 ultrafast bootstrap replicates yielding support values for each node with the "GTR + I + G" model. The resulting tree file was plotted into a phylogenetic tree using Table 2itol (https://github.com/mgoeker/table2itol) and itol (https://itol.embl.de/).

For population structure and diversity analysis of Rutaceae chloroplast genomes (509 samples), principal component analysis (PCA) was conducted using PLINK

(v1.90b4) (Purcell et al. 2007), and results were visualized with ggplot2 (Wickham 2011). Considering the predominant maternal lineage of pummelo in sweet orange, sour orange, grapefruit, and lemon, we classified their chloroplast genomes into a group of pummelo (PU) and the remaining clustered into ten groups: *Atalantia* (at), *Citron* (ci), *Clausena* (cla), *Fortunella* (fo), *Ichangensis* (ich), Australian lime (lime), mandarin (ma), *Murraya* (mu), *Poncirus* (po), and *Zanthoxylum* (za).

To ensure calculation accuracy, samples with inconsistent nuclear and chloroplast classifications were excluded. Divergence (Dxy), fixation index (FST), and nucleotide diversity (Pi) statistics were calculated based on SNP variations using genomics_general (https://github.com/simonhmartin/genomics_general). The Pi and Dxy values of populations were visualized using ggplot2 (Wickham 2011).

Haplotype network construction

To study genealogical relationships among Rutaceae populations, we constructed a haplotype network based on SNP variation data. Plink (v1.90b4) (Purcell et al. 2007) was employed to filter low-frequency variant loci in VCF files. Subsequently, VCF files were converted into NEXUS (NEX) format files using vcf2phylip (Ortiz 2019), BioEdit (v5) (Hall et al. 2011), and DnaSP (Librado and Rozas 2009). Grouping information was then incorporated into NEX files. Finally, we used POPART (Leigh and Bryant 2015) to visualize Templeton, Crandall, and Sing (TCS) haplotype networks.

Nucleotide substitution rate analysis

After eliminating stop codons and non-coding genes from chloroplast genes, we estimated the nucleotide substitution rates of 81 chloroplast genes across 378 samples. The 81 protein-coding genes (PCG) from the 378 samples were aligned separately using MAFFT (v7.490) (Katoh and Standley 2013). Utilizing the aligned gene sequences, a constraint tree was constructed for each gene using IQTREE (Minh et al. 2020). The non-synonymous (*dN*), synonymous (*dS*), and non-synonymous to synonymous ratios (*dN*/*dS*) were calculated using the codeml program in the PAML package (Yang 2007), employing the F3×4 model for codon frequency estimation. Gap regions were excluded using the "clean data = 1" option, and the "model = 0" option was applied while keeping other parameters in the codeml control file at default settings. The *dN*, *dS*, and *dN*/*dS* values for the 81 genes were visualized using Origin2019.

For comparisons across different functional groups of PCGs, we selected seven groups with more than five members: photosystem I (PSA), photosystem II (PSB), cytochrome B6f complex (PET), ATP synthase (ATP), Nicotinamide adenine dinucleotide (NADH), ribosomal proteins large subunit (RPL), and ribosomal proteins small subunit (RPS). The *dN*, *dS*, and *dN/dS* values for these seven functional groups were visualized using ggplot2 (Wickham 2011).

Gene diversity and correlation analysis

DnaSP (v5) (Librado and Rozas 2009) was used to calculate gene length, GC content, Pi, per site from total mutations (Eta), and the number of variable sites (S) based on the 81 PCG alignments. The correlation between nucleotide substitution rates (including *dN*, *dS*, and *dN/dS* values for each gene) with Pi and Eta were calculated using ggscatterstats functions from the ggstatsplot package (Patil 2021). Additionally, correlation analysis for gene length, GC content, Pi, Eta, S, and nucleotide substitution rate was conducted using ggscatterstats.

Results

Rutaceae chloroplast genome variation analysis

A chloroplast variation map was constructed using short paired-end sequencing reads from 509 accessions (Table S1), mapping the reads to the *Fortunella hindsii* (Hong Kong kumquat) chloroplast genome. The analysis revealed 11,580 SNPs, 1,401 insertions, and 1,080 deletions. Notably, a significant difference in variation prevalence existed between the LSC and SSC regions, with 78.55% of the variants identified in the LSC region and 20.56% in the SSC region. Conversely, the IR region exhibited remarkably low variation (0.89%) (Fig. 1), consistent with observations in other higher plants (Daniell et al. 2016).

To reduce the influence of region length on chloroplast genome, we calculated variant density separately for the SSC, LSC, and IR regions. A similar pattern was observed in the SSC (131.2 SNPs and InDels per kb) and LSC regions (107.26 SNPs and InDels per kb). In contrast, the IR region exhibited significantly lower density, approximately 1.97 SNPs and InDels per kb. Furthermore, we annotated the chloroplast variation map to investigate the effects of variation on chloroplast genes. Notably, 83.89% of the variants were concentrated in the upstream and downstream regions of genes, while only 4.32% occurred in the exon region. Synonymous and non-synonymous mutations constituted 1% and 2.5%, respectively (Table S2).

Delly was employed to identify structural variations in the Rutaceae chloroplast genome. After filtering low-frequency variant loci and large fragment variants, 125 SV loci were obtained, including 78 deletions, 30 insertions, 14 inversions, and three duplications. All SV loci were located in the SC region, except for two insertions

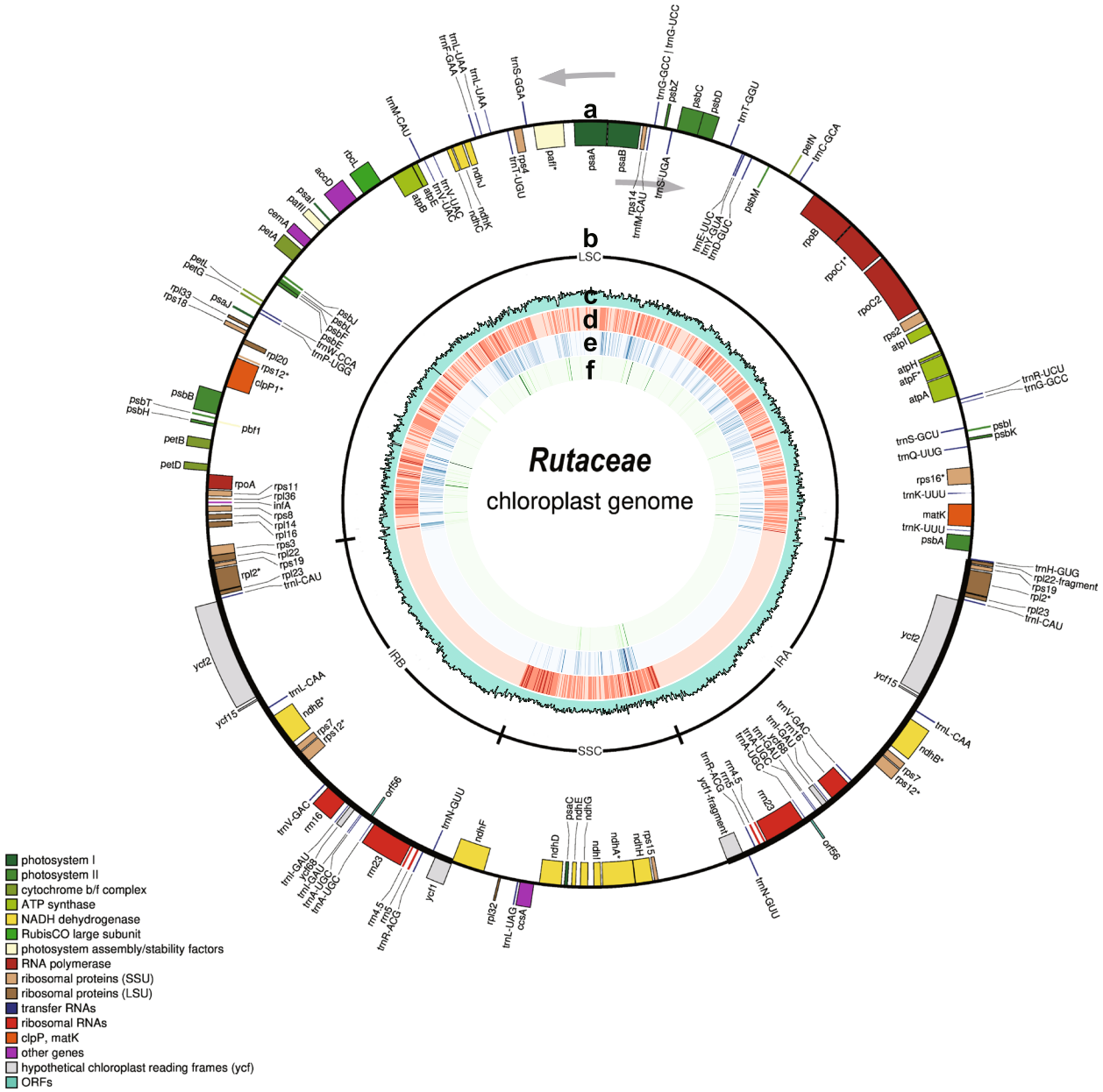

Fig. 1 Exploration of chloroplast variation in Rutaceae using the Hong Kong kumquat chloroplast as a reference genome. **a** Gene map of the Hong Kong kumquat chloroplast, with arrows indicating gene transcriptional direction. **b** Locations of the four regions within the chloroplast genome. **c** Bar graph illustrating the GC content of the Hong Kong kumquat chloroplast genome. **d**–**f** Heatmaps depicting SNP, Indel, and structural variants (SV) densities across 509 Rutaceae samples. LSC, large single-copy; IRA, inverted repeat A; SSC, small single-copy; IRB, inverted repeat B

located at the SC and IR region boundary (Fig. 1f, Table S3).

The GC contents of the 378 Rutaceae chloroplast genomes ranged from 38.4% to 38.5%. The IR regions exhibited the highest GC content at 42.9%–43%, ensuring genomic stability. Conversely, the LSC region demonstrated a GC content of 36.7%–36.9%, while the SSC region exhibited the lowest GC content at 33.1%–33.4% (Fig. 3b, Table S4).

The assessment of simple sequence repeats (SSRs) in 378 genomes revealed that 70% of the samples harbored 75–80 SSRs. Samples exhibited variability, with the maximum number of SSRs observed in *Glycosmis pentaphylla* (NC_032687, 112 SSRs), while the minimum was

recorded in *Zanthoxylum madagascariense* (NC_046744, 58 SSRs) (Table S4).

Chloroplast population genomic analysis in Rutaceae

To investigate the genetic relationships among different genera and species, including interspecific hybrids, we employed the maximum likelihood (ML) method to construct a phylogenetic tree based on the chloroplast variation map. The phylogeny highlighted two significant genetic features: widespread hybridization in modern citrus cultivars and the cohesive clustering of wild and cultivated citrus species. The chloroplasts of sweet orange, sour orange, grapefruit, and lemon, primarily inherited from pummelo (Wang et al. 2022a, b) (Fig. S1a), led to the clustering of some hybrids with other groups, reflecting interspecific hybridization through maternal inheritance. For instance, chloroplasts of five sweet oranges (sweet_orange_A20, sw_HML, sw_BZH, sw_NMH, sw_NF) clustered with mandarin. Despite *Atalantia* being considered primitive citrus that diverged from ichangensis (*Citrus ichangensis*) around 20 Ma (Wang et al. 2017), the length of inner branches suggested a close grouping of *Atalantia, Zanthoxylum*, *Murraya*, and *Clausena* (Fig. 2a, Fig. S2). Our chloroplast phylogeny thus revealed both species diversification and extensive hybridization, offering insights into the evolutionary dynamics by breeding improvements and natural hybridization.

The genetic landscape of Rutaceae was examined using PCA. *Zanthoxylum* emerged with the longest genetic distance, showcasing distinctive features in its nuclear genome (2n = 46, genome size = 4.6 G) compared to other species (2n = 18). Combining these findings with the phylogenetic analysis, three distinct clusters were identified: *Zanthoxylum* (cluster1), a cluster comprising *Atalantia*, *Clausena*, and *Murraya* (cluster2), and a cluster with wild and cultivated *Citrus* species (cluster3) (Fig. 2b). Examining Pi and Dxy in the chloroplast genome of Rutaceae populations (Fig. 2c, d) revealed significantly higher diversity in cluster1 and cluster2 than in cluster3. The FST in the chloroplast genome across the three clusters indicated that in most of the SC regions, genetic differentiation was most pronounced between cluster1 and cluster3, followed by cluster1 and cluster2, while the differentiation of cluster2 and cluster3 was the least. Remarkably, FST in the IR region was significantly lower than in the SC region, with the lowest differentiation observed between cluster1 and cluster3, indicating distinct differentiation dynamics in different structural regions of the chloroplast (Fig. 2e). Citron and lime exhibited the highest genetic diversity in the chloroplast genome, potentially linked to increased effective population size (Ne), as reported previously (Wang et al. 2022a, b). Notably, *Clausena* displayed the lowest genetic diversity in clutser1, possibly attributed to the small sample size limiting diversity. Moreover, the Dxy statistic aligned with the phylogenetic tree and PCA results.

To trace the genetic variants contributing to chloroplast genome divergence among species, we constructed a chloroplast haplotype network based on whole-genome variations. The network exhibited four independent main branches. Within mandarins, both cultivated and wild varieties formed separate branches, indicating single domestication events in mandarin evolution. Similarly, the *Fortunella* genus showcased a split between cultivated kumquat and wild kumquat (Hong Kong kumquat) populations. Notably, the long-term clonal propagation of sweet orange resulted in lower chloroplast genome variation, with a dominant haplotype representing 63.27% of the variance. Despite variations in chromosome numbers and nuclear genomic architectures between *Zanthoxylum* and cluster2 species (*Atalantia, Murraya,* and *Clausena*), the chloroplast genome haplotypes unexpectedly exhibited proximity (Fig. S2). The intriguing question of whether genetic factors influence nuclear and chloroplast genome differentiation during species diversification remains open.

Comparative genomics of Rutaceae chloroplast genomes

To investigate the evolution of chloroplast genome size in Rutaceae, we conducted de novo assemblies of 509 chloroplast genomes based on deep NGS data, with 67.5% (344/509) assembled into a circular genome. Combining these with 34 published chloroplast genomes, we obtained a dataset of 378 genomes, representing 17 species. The average sequencing depth for each assembly was 4565.39 (Fig. 3a, Fig. S5a and Table S5). The chloroplast genome size ranged from 157,339 to 161,204 bp (average 159,760 bp). Mandarin exhibited the largest genome (average 160,694 bp), while *Ruta* had the smallest (average 157,339 bp). LSC region length variation was a major contributor to chloroplast genome size changes ($R = 0.92$, $P = 3.97E{-}159$) (Fig. S3a). Positive correlations were observed between the size of the SC region and the whole genome, with the LSC region showing a higher correlation than the SSC region (Fig. S3b). Conversely, the IR region size was negatively correlated with the chloroplast genome size (Fig. S3c).

The chloroplast genome has undergone dynamic expansion and contraction, particularly through the progressive enlargement caused by the integration of the IR region into the SC region—a trend observed broadly across land plants (Raubeson and Jansen 2005). Sorting chloroplast genomes for Rutaceae based on the phylogenetic tree (Fig. S4) revealed evident patterns of expansion and contraction. Grouping the samples into A, B, and C, we analyzed the trends in overall

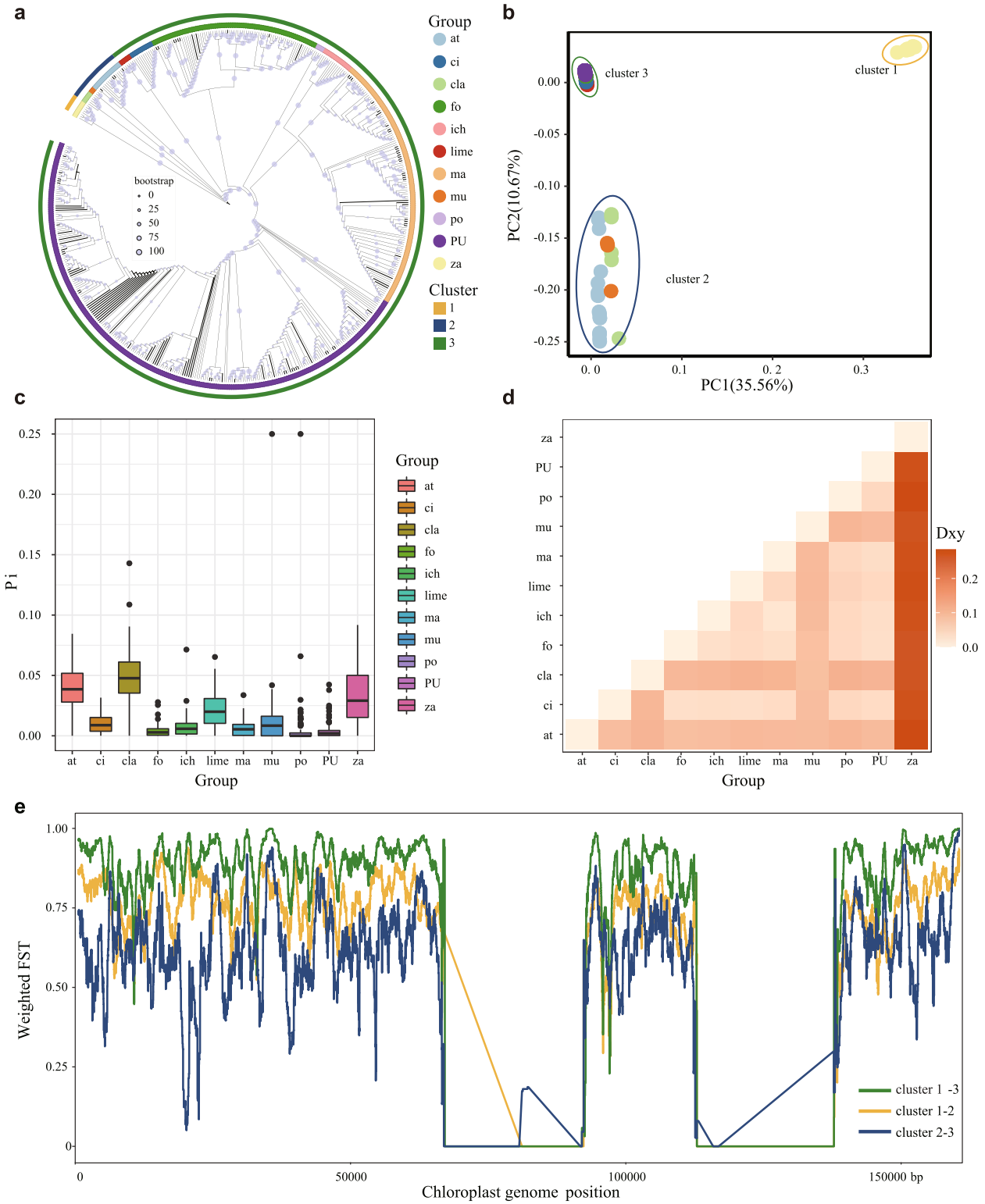

Fig. 2 Population structure of Rutaceae chloroplast genomes. **a** Phylogenetic tree constructed based on chloroplast variation data, with *Zanthoxylum* set as an outgroup and bootstrap values represented by circle size. **b** Principal component analysis (PCA) of chloroplast genomes from 11 Rutaceae populations, with PC1 and PC2 accounting for 35.56% and 10.67% variance, respectively. **c** Analysis of nucleotide diversity (Pi) in chloroplast genomes across 11 Rutaceae populations. **d** Divergence (Dxy) analysis of chloroplast genomes in 11 Rutaceae populations. The degree of divergence is represented by a heatmap. **e** Fixation index (FST) depicting differentiation between three clusters in the chloroplast genome

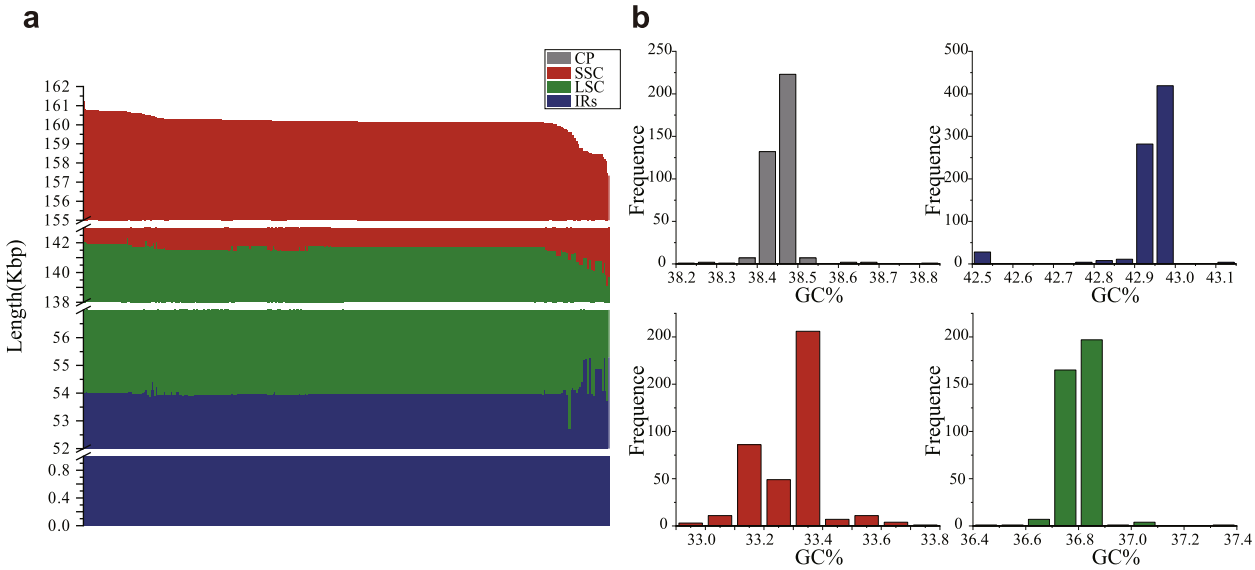

Fig. 3 Rutaceae chloroplast genome length and GC content. A total of 378 chloroplast whole genomes were assembled and displayed in gray, with red representing the small single-copy (SSC) region, green the large single-copy (LSC) region, and blue the two inverted repeat (IR) regions. **a** Length statistic of 378 chloroplast whole genomes, with samples sorted by size on the x-axis and genome length on the y-axis. **b** Distribution of GC content in 378 chloroplast genomes, with the x-axis representing GC content and the y-axis indicating the number of samples. CP, chloroplast genome length

genome length and three partitions: LSC, SSC, and the IR region. In groups A and B, the chloroplast genome length significantly increased with correlation coefficients of 0.72 ($P = 1.5E{-}8$) and 0.76 ($P < 2.2E{-}16$), respectively. Conversely, group C exhibited a decrease with a correlation coefficient of −0.17 ($P = 0.015$) (Fig. 4a), likely influenced by phylogenetic branch samples, whose chloroplast genome length is about 70 bp longer than other types in the chloroplast phylogenetic tree. In the LSC region, there was a significant increase in size for groups A ($R = 0.72$, $P = 1.1E{-}8$) and B ($R = 0.75$, $P < 2.2E{-}16$), whereas group C showed non-significant growth ($R = 0.04$, $P = 0.58$) (Fig. 4b). The SSC region exhibited size increases in groups A ($R = 0.33$, $P = 0.024$) and B ($R = 0.47$, $P = 9.8E{-}9$) and a decrease in group C ($R = -0.2$, $P = 0.0056$) (Fig. 4c). In the IR region, groups A ($R = -0.28$, $P = 0.058$) and C ($R = -0.11$, $P = 0.14$) exhibited significant decreases, while group B displayed a slight increase ($R = 0.18$, $P = 0.037$). Notably, the IR region displayed minimal changes in size for groups B and C, whereas the significant decrease in the length of group A was attributed to the increase in the IR region in *Zanthoxylum* (Fig. 4d).

In summary, Rutaceae chloroplast genome size increases during diversification, primarily driven by amplifications in the LSC and SSC regions. *Zanthoxylum* exhibited an extended IR region in group A compared to other Rutaceae species, with groups B and C showing relatively stable IR region sizes.

Chloroplast genome selection patterns in Rutaceae

To elucidate the selection dynamics of the chloroplast genome, we comprehensively analyzed the chloroplast pan-genome (Fig. S5). The annotation encompassed 150 conserved genes, including 82 PCGs, 36 tRNA genes, 4 rRNA genes, and one open reading frame (ORF). Notably, 4 rRNA and 7 tRNA genes exhibited two duplications, with one tRNA gene presenting four duplicates. Additionally, 10 PCGs displayed multi-copy variations. The pan-genome revealed individual-level distinction in gene numbers, with Hong Kong kumquat variety 21B1 exhibiting the lowest number of conserved genes (144 genes). Focusing on PCG genetic diversity, we used two different statistics, Pi and Eta (Table S6). The *hypothetical chloroplast open reading frame 1* (*ycf1*) gene had the most variants, while *ycf15* demonstrated greater conservation and fewer variations (Fig. S6a, b). The extensive length of the *ycf1* gene, spanning the IR and SSC regions, and the dynamics of IR region expansion and contraction contributed to its elevated variant count and genetic diversity.

Although the gene count in Rutaceae chloroplasts remains remarkably conserved, a broader family-level examination reveals extensive loss of chloroplast translation *initiation factor 1* (*infA*) genes, with 344 out of 378 chloroplast genomes annotated with *infA* genes. Within Rutaceae, 20 different *infA* gene sequences were identified, with 307 samples exhibiting consistent *infA* protein sequences. Additionally, *ribosomal protein l2* (*rpl2*) genes

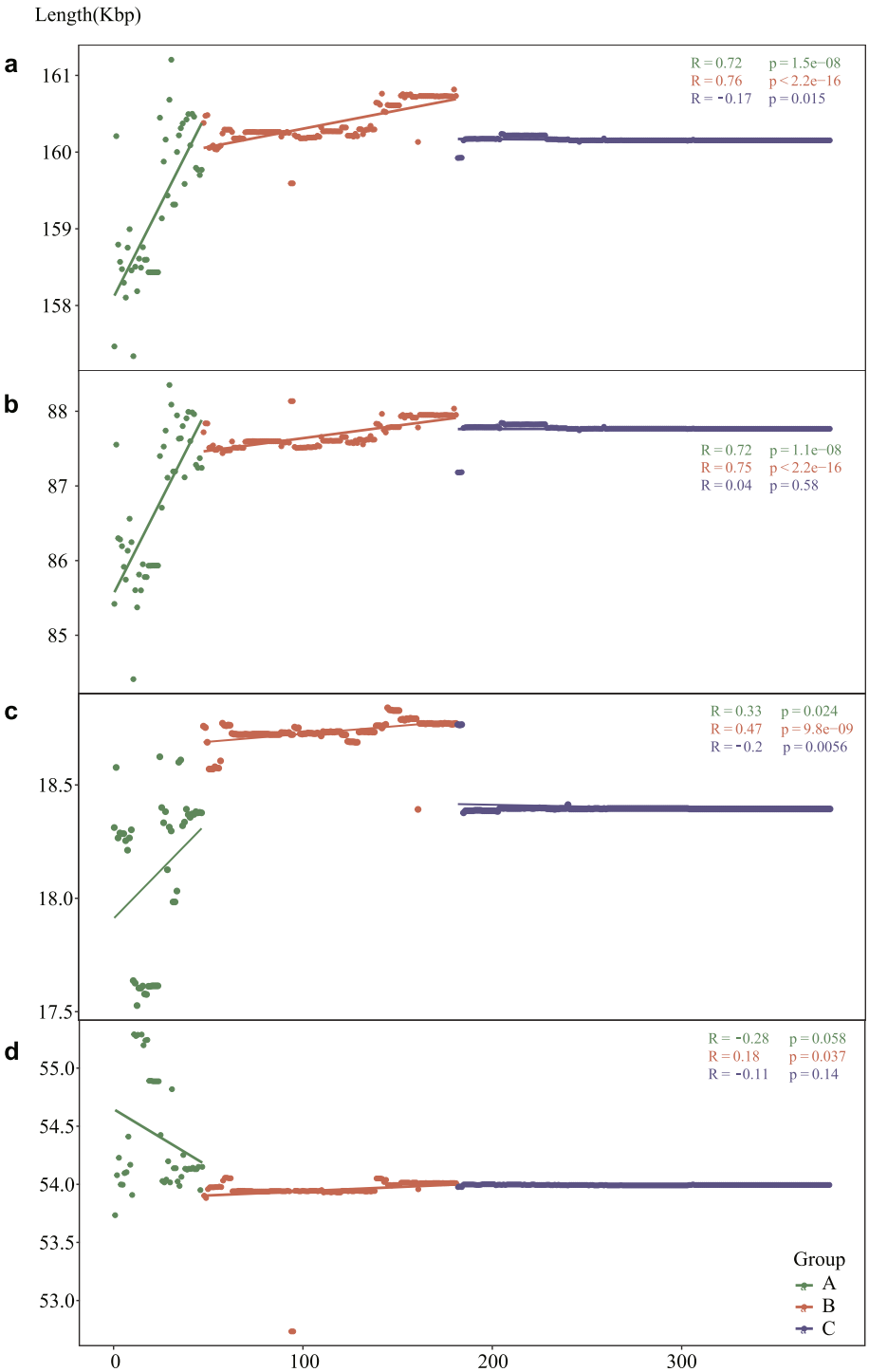

Fig. 4 Trends in chloroplast genome size variation. **a** The 378 chloroplast whole genomes were sorted according to the genome phylogenetic tree. Trends in chloroplast complete genome, **b** large single-copy (LSC) region, **c** small single-copy (SSC) region, and **d** inverted repeat (IR) region size changes were measured for three groups of samples

were lost in two samples, and *ribosomal protein s4* (*rps4*) genes were lost in one sample. A pan-genome of chloroplast PCGs in the Rutaceae subfamily was constructed. Despite an increase in the total number of chloroplast genes, core and pan genes exhibited minimal variation at the family level, underscoring the overall genomic stability (Fig. S5).

To identify potential functions under selection in the chloroplast genome, we conducted a focused analysis on 15 pathways, each comprising more than five genes. Notably, genes within the PSA and PSB pathways displayed higher conservation, while those within the NADH pathway exhibited increased variability (Fig. S6c, d). A substantial 91% of NADH functional genes were located in the SC region, potentially contributing to the increased variability. However, despite being located in the same SC region, both PSA and PSB pathway genes remained conserved. This suggests that the interplay between region location and gene function may collectively influence the genetic diversity observed in chloroplast genes.

To study the evolution of chloroplast genes in Rutaceae, the ratio of substitution rates at dN and dS sites was calculated for 81 PCGs, excluding the trans-splice gene *rps12* (Table S5). The dN/dS values of 80 genes were all < 1, with only *rps16* exhibiting a dN/dS > 1 (Fig. 5a). Notably, the dN values of PSA genes were the lowest, whereas those of *RPS* genes were the highest (Fig. 5b). Conversely, the dS values of *Nicotinamide Adenine Dinucleotide* (*NADH*) genes were the highest, while those of PSB genes were the lowest (Fig. 5c). To explore correlations between evolutionary rates and gene characteristics, we conducted a correlation analysis between dN, dS, and dN/dS of 81 PCGs and Pi and Eta, respectively. All correlations were statistically significant ($P < 0.05$), exhibiting a positive trend (Fig. 6).

Discussion

Chloroplast phylogeny and diversity in Rutaceae

The exploration of the cytoplasmic and nuclear genomes of citrus revealed the complete evolutionary and hybridization history of the *Citrus* genus. Citrus originated in the Himalayas around 8 Ma, giving rise to several foundational species during migration. The chloroplast phylogenetic tree illustrates that the chloroplast genomes of lime,

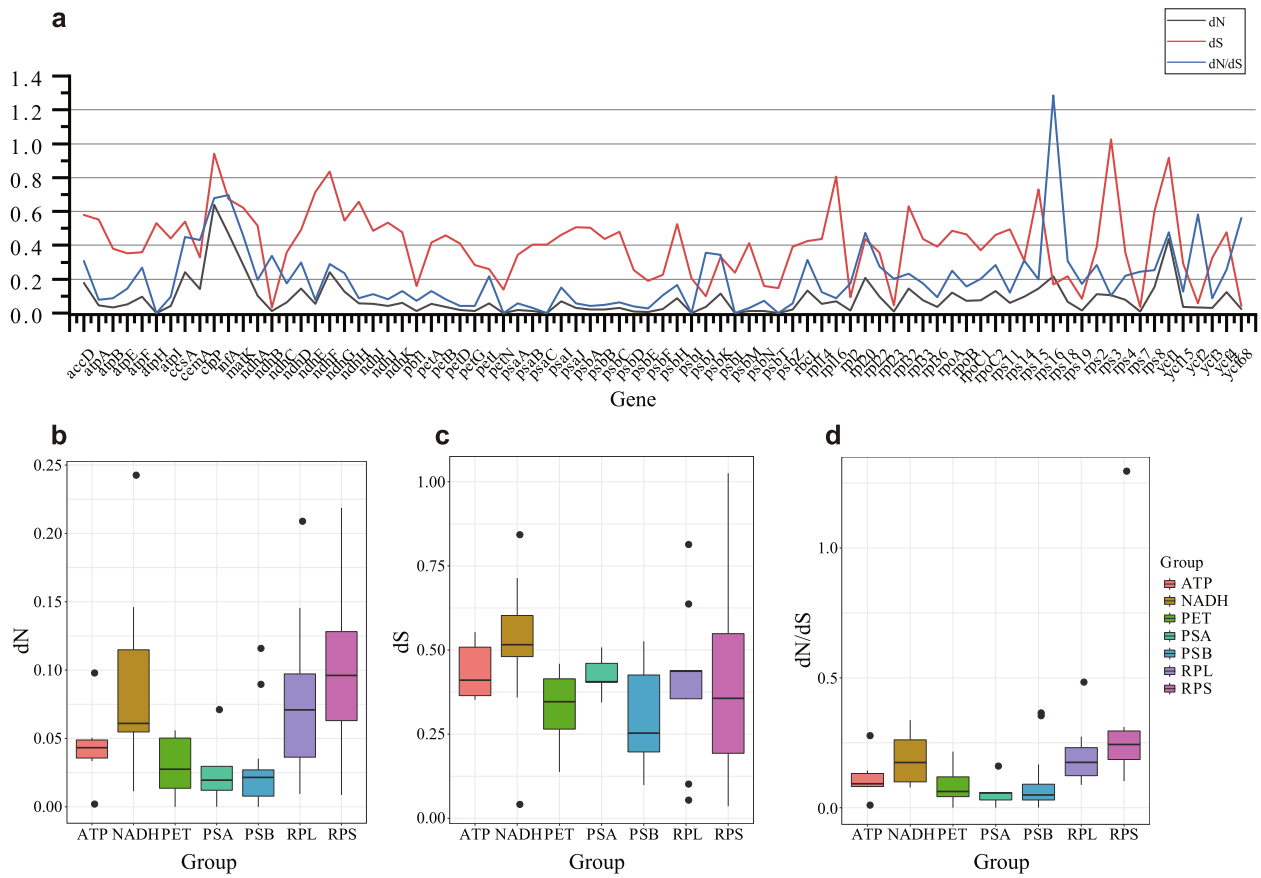

Fig. 5 Diversity of chloroplast protein-encoding genes in Rutaceae. Nucleotide diversity (Pi) and per site from total mutations (Eta) values of 81 chloroplast genes were measured. **a** Non-synonymous (dN), synonymous (dS) substitution rates, and dN/dS of 81 protein-coding genes. Classification of 81 protein-coding genes into functional pathways, with dN **b**, dS **c**, and dN/dS **d** counted for seven functional pathways (number of genes > 5). ATP, ATP synthase; NADH, Nicotinamide adenine dinucleotide; PET, cytochrome B6f complex; PSA, photosystem I; PSB, photosystem II; RPL, ribosomal proteins large subunit; RPS, ribosomal proteins small subunit

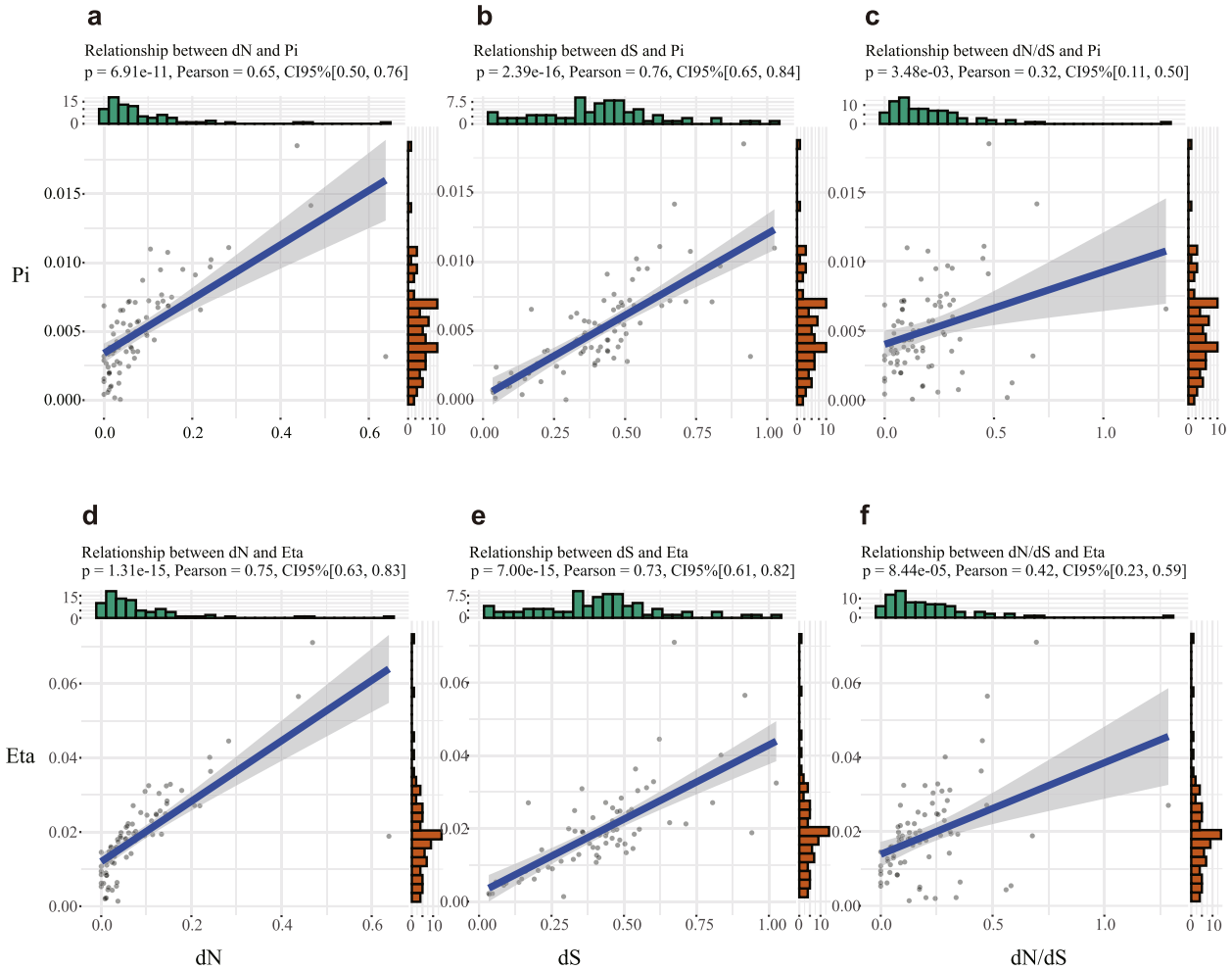

Fig. 6 Correlation between selective pressure and diversity of chloroplast genes. **a**–**c** Correlation of non-synonymous (dN), synonymous (dS) substitution rates, and dN/dS with nucleotide diversity. **d**–**f** Correlation of dN, dS substitution rates, and dN/dS with per site from total mutations (Eta)

sweet orange, lemon, and grapefruit are derived from pummelo. Additionally, citron and Australian lime (lime) form a distinct group, while *Ichangensis* and mandarin cluster together. *Poncirus* was an independent group situated between the above two groups (Wu et al. 2018). In this study, the *Citrus* topology in the chloroplast phylogenetic tree closely resembled previous reports. However, a notable distinction was the positioning of *Poncirus*, which was robustly supported between *Fortunella* and *Ichangensis*, as indicated by a high bootstrap value of 99.4% (Fig. 2a). Plant domestication, involving the continuous human selection of agricultural traits in wild plants over time, often leads to a reduction in population diversity. Among Rutaceae, citrus has undergone extensive domestication events (Rao et al. 2021). The Pi values of chloroplast genomes in the three domesticated citrus species (*Fortunella*, mandarin, and pummelo) are lower than those in the undomesticated species (citron and *Ichangensis*) (Fig. 2c), indicating the impact of domestication on genetic diversity.

Differences in the relationship between chloroplast and nuclear gene populations in Rutaceae

The observed disparity between chloroplast and nuclear phylogenies in Rutaceae prompts a deeper investigation into the underlying factors intrinsic to the evolutionary dynamics of these genomes. Primarily, the maternal inheritance of chloroplast DNA in most plants, including Rutaceae, contrasts with the biparental inheritance of nuclear DNA. This difference in inheritance patterns leads to distinct evolutionary trajectories, with chloroplast genomes reflecting maternal lineage history and nuclear genomes representing a mix of both parental lineages. Additionally, the lower mutation rate in chloroplast genomes, compared to nuclear genomes, results in the retention of ancient evolutionary signals for more

extended periods. This divergence in mutation rates may explain why chloroplast phylogeny tends to reflect more ancient divergences within Rutaceae, while nuclear phylogeny captures more recent evolutionary events.

To analyze the relationship of chloroplast DNA sequences among and within Rutaceae populations, we constructed a haplotype network map. This map revealed a closer relationship between the chloroplast genomes of *Atalantia* and *Murraya* from the Aurantioideae subfamily with those of the Rutoideae subfamily than with other Aurantioideae species. Various haplotype branches were observed in the chloroplasts of mandarin, pummelo, *Fortunella*, and grapefruit within the Aurantioideae species (Fig. S2). The Dxy showed differences in chloroplast genomes among Rutaceae groups, with the *Ichangensis*–mandarin, ichangensis–*Fortunella*, and *Fortunella*–citron showing the least chloroplast differences (Fig. 2d). However, the phylogenetic tree which was constructed based on nuclear genome sequencing differed from our chloroplast-based results due to distinct genetic patterns. In Restriction-site Associated DNA Sequence (RAD-seq) results, *Atalantia* and *Murraya* remained classified under Aurantioideae (Nagano et al. 2018). Nevertheless, the three groups mentioned above, closely related at the chloroplast genome level, exhibited greater nuclear genome distance, indicating the inaccuracies in classification relying solely on the chloroplast genome.

Effect of chloroplast gene characteristics on genetic diversity in Rutaceae

Research on the evolutionary patterns of chloroplast genomes has predominantly explored the rates of different regions or functional categories (Birky 1995; Zhu et al. 2015), with limited attention given to the role of chloroplast gene characteristics such as length and GC content in genetic diversity and gene evolution. To address this gap, we examined Chloroplast genes, evaluating their Pi, Eta, S, gene length, and GC content (Table S6). The mean values for chloroplast genes Pi, Eta, and S were 0.0051, 0.0189, and 97.9, respectively. Notably, NADH, RPS, and RPL pathway genes exhibited increased variation over evolution, while PSA, PSB, PET, and ATP pathway genes were more conserved. Gene length and GC content affect the number of gene variants. Correlation analyses revealed that gene length exhibited no correlation with Pi and Eta but was significantly correlated with S. In contrast, GC content was significantly negatively correlated with Pi and Eta, but not correlated with S (Fig. S7). Although chloroplast genes with longer lengths accumulated more variants, their Pi and Eta values were comparable to other genes. The higher the GC content in the gene, the lower were Pi and Eta. Additionally, genes with higher GC content demonstrated greater stability, with no correlation between gene length and diversity. To investigate whether gene length and GC content influenced gene evolution, we assessed the dN, dS, and dN/dS of the 81 chloroplast genes in Rutaceae. The average dN, dS, and dN/dS were 0.0850, 0.4096, and 0.2047, respectively. Correlation analyses between gene length and GC content with dN, dS, and dN/dS revealed no significant correlations (Fig. S8). In summary, in Rutaceae chloroplasts, gene length appears to have no impact on genetic diversity and gene evolution. However, GC content is negatively correlated with genetic diversity, highlighting its role in maintaining stability.

Conclusions

In summary, this study utilized next-generation sequencing data from 509 samples across 15 species in the Rutaceae family to assemble 343 chloroplast genomes. The variation map revealed that 99.11% of the variation occurred within single-copy regions, with differences in chloroplast genome size correlating with the length of single-copy regions. The chloroplast genetic revealed the short chloroplast genetic distance among *Atalantia*, *Clausena* and *Murraya*. The analysis of gene selective pressure revealed that most chloroplast genes are under negative selection. Genes in the NADH, RPS, and RPL pathways showed increased variation in evolution, while genes in the PSA, PSB, PET, and ATP pathways were more conserved. Furthermore, the length of the gene had no impact on nucleotide diversity and gene evolution, while GC content was negatively correlated with nucleotide diversity.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1007/s44281-024-00032-9.

Additional file 1.

Additional file 2: Fig S1. (a) Phylogenetic tree (b) and principal component analysis (PCA) of 509 Rutaceae samples, classified by species information. **Fig S2.**Haplotype network map for 509 samples. Circle size and color represent sample number and category, respectively. **Fig S3.** (a) Correlation between chloroplast genome size and large single-copy (LSC), (b) small single-copy (SSC), (c) and inverted repeat (IR) region sizes. **Fig S4.** Phylogenetic tree from 378 chloroplast genomes, categorized into groups A, B, and C based on divergence relationships (grouping info in Table S4). **Fig S5.** Rutaceae chloroplast pan-genomes. The pan-genome was constructed using 378 protein-coding genes of the chloroplast genome. (a) Chloroplast genome assembly depth distribution frequency. (b) Trends in core and total gene changes. (c) Annotation of core and accessory genes using KEGG. **Fig S6.** Diversity of chloroplast protein-encoding genes in Rutaceae. Nucleotide diversity (Pi) and per site from total mutations (Eta) values for 81 chloroplast genes. (a), (b) Pi and Eta values of 81 protein-coding genes. (c), (d) Classification of 81 genes by functional pathways, with Pi and Eta values counted for seven functional pathways (number of genes > 5). ATP, ATP synthase; NADH, Nicotinamide adenine dinucleotide; PET, cytochrome B6f complex; PSA, photosystem I; PSB, photosystem II; RPL, ribosomal proteins large subunit; RPS, ribosomal proteins small subunit. **Fig S7.**Correlation analysis of length, GC content,

nucleotide diversity (Pi), per site from total mutations (Eta), and number of variable sites (S) of chloroplast genes. **Fig S8.** Correlation analysis of length, GC content, and the non-synonymous (dN), synonymous (dS), and dN/dS of chloroplast genes.

Acknowledgements
Not applicable.

Authors' contributions
W.W.G. and Y.F.Z. conceived and designed the project. C.C.L. assembled the genomes and performed population genomics analysis. N.W. collected samples. Y.B., T.H., J.C.L. and Z.Y.M. were involved in data analysis. C.C.L. wrote the manuscript with contributions from N.W., X.M.W., K.D.X., Y.F.Z and W.W.G.

Funding
This research was financially supported by grants from the National Natural Science Foundation of China (no. U23A20203), the HZAU-AGIS Cooperation Fund (no. SZYJY2022009), and the China Agriculture Research System (no. CARS-26).

Availability of data and materials
The BioProject ID of the 509 next-generation sequencing data, chloroplast genome assembly, annotation and variation files are available on https://github.com/Licc900/Chloroplast_evolution_Rutaceae.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Received: 27 November 2023 Revised: 9 January 2024 Accepted: 12 January 2024
Published online: 30 April 2024

References

Aphalo P. R package ggpmisc is an extension to ggplot2 and the Grammar of Graphics. Im: Ggpmisc. GitHub. 2021. https://github.com/aphalo/ggpmisc. Accessed 10 Oct 2023.
Bayer RJ, Mabberley DJ, Morton C, Miller CH, Sharma IK, Pfeil BE, et al. A molecular phylogeny of the orange subfamily (Rutaceae: Aurantioideae) using nine cpDNA sequences. Am J Bot. 2009;96:668–75. https://doi.org/10.3732/ajb.0800341.
Birky CW. Uniparental inheritance of mitochondrial and chloroplast genes: mechanisms and evolution. Proc Natl Acad Sci. 1995;92:11331–8. https://doi.org/10.1073/pnas.92.25.11331.
Carbonell-Caballero J, Alonso R, Ibanez V, Terol J, Talon M, Dopazo J. A phylogenetic analysis of 34 chloroplast genomes elucidates the relationships between wild and domestic species within the genus *Citrus*. Mol Biol Evol. 2015;32:2015–35. https://doi.org/10.1093/molbev/msv082.
Chaudhari NM, Gupta VK, Dutta C. BPGA- an ultra-fast pan-genome analysis pipeline. Sci Rep. 2016;6:24373. https://doi.org/10.1038/srep24373.
Chung SM, Staub JE. The development and evaluation of consensus chloroplast primer pairs that possess highly variable sequence regions in a diverse array of plant taxa. Theor Appl Genet. 2003;107:757–67. https://doi.org/10.1007/s00122-003-1311-3.
Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. 2012;6:80–92. https://doi.org/10.4161/fly.19695
Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and VCFtools. Bioinformatics. 2011;27:2156–8. https://doi.org/10.1093/bioinformatics/btr330.
Daniell H, Lin CS, Yu M, Chang WJ. Chloroplast genomes: diversity, evolution, and applications in genetic engineering. Genome Biol. 2016;17:134. https://doi.org/10.1186/s13059-016-1004-2.
Engler A. Rutaceae. In: Engler A, Prantl K, editors. Dia natürlichen pflanzenfamilien. 2nd ed. Leipzig, Germany; 1931:187–359.
Feng S, Liu Z, Cheng J, Li Z, Tian L, Liu M, et al. Zanthoxylum-specific whole genome duplication and recent activity of transposable elements in the highly repetitive paleotetraploid Z. bungeanum genome. Hortic Res. 2021;8:205. https://doi.org/10.1038/s41438-021-00665-1
Gao LZ, Liu YL, Zhang D, Li W, Gao J, Liu Y, et al. Evolution of Oryza chloroplast genomes promoted adaptation to diverse ecological habitats. Commun Biol. 2019;2:278. https://doi.org/10.1038/s42003-019-0531-2.
Green BR. Chloroplast genomes of photosynthetic eukaryotes. Plant J. 2011;66:34–44. https://doi.org/10.1111/j.1365-313X.2011.04541.x.
Greiner S, Lehwark P, Bock R. Organellar Genome DRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes. Nucleic Acids Res. 2019;47:W59–W64. https://doi.org/10.1093/nar/gkz238
Grivet D, Heinze B, Vendramin G, Petit R. Genome walking with consensus primers: application to the large single copy region of chloroplast DNA. Mol Ecol Notes. 2001;1:345–9. https://doi.org/10.1046/j.1471-8278.2001.00107.x.
Groppo M, Pirani JR, Salatino ML, Blanco SR, Kallunki JA. Phylogeny of Rutaceae based on two noncoding regions from cpDNA. Am J Bot. 2008;95:985–1005. https://doi.org/10.3732/ajb.2007313.
Hagemann R. The sexual inheritance of plant organelles. Molecular biology and biotechnology of plant organelles: Springer; 2004:93–113. https://doi.org/10.1007/978-1-4020-3166-3_4
Hall T, Biosciences I, Carlsbad C. BioEdit: an important software for molecular biology. GERF Bull Biosci. 2011;2:60–1. https://doi.org/10.55838/1980-3540.ge.2018.287
Heinze B. A database of PCR primers for the chloroplast genomes of higher plants. Plant Methods. 2007;3:4. https://doi.org/10.1186/1746-4811-3-4.
Hipkins VD, Krutovskii KV, Straws SH. Organelle genome in conifers: structure, evolution. For Genet. 1994;1:179–89.
Jin JJ, Yu WB, Yang JB, Song Y, Depamphilis CW, Yi TS, et al. GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. Genome Biol. 2020;21:241. https://doi.org/10.1101/256479.
Kassambara A. n.d. Ggpubr: 'ggplot2' Based Publication Ready Plots. ggpubr. https://rpkgs.datanovia.com/ggpubr/. Accessed 15 Oct 2023
Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013;30:772–80. https://doi.org/10.1093/molbev/mst010.
Kelchner SA. The evolution of non-coding chloroplast DNA and its application in plant systematics. Ann Mo Bot Gard. 2000;87:482. https://doi.org/10.2307/2666142.
Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, et al. Circos: an information aesthetic for comparative genomics. Genome Res. 2009;19:1639–45. https://doi.org/10.1101/gr.092759.109.
Kubitzki K. Flowering Plants. Eudicots: Sapindales, Cucurbitales, Myrtaceae. 1st ed. Berlin: Springer; 2011
Leigh JW, Bryant D. POPART: full-feature software for haplotype network construction. Methods Ecol Evol. 2015;6:1110–6. https://doi.org/10.1111/2041-210x.12410.
Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25:1754–60. https://doi.org/10.1093/bioinformatics/btp324.
Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and SAMtools. Bioinformatics. 2009;25:2078–9. https://doi.org/10.1093/bioinformatics/btp352.
Li HT, Luo Y, Gan L, Ma PF, Gao LM, Yang JB, et al. Plastid phylogenomic insights into relationships of all flowering plant families. BMC Biol. 2021;19:232. https://doi.org/10.1186/s12915-021-01166-2.
Li Y, Li X, Sylvester S, Duan Y. Plastid genomes reveal evolutionary shifts in elevational range and flowering time of Osmanthus (Oleaceae). Ecol Evol. 2022;12: e8777. https://doi.org/10.1002/ece3.8777.
Librado P, Rozas J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics. 2009;25:1451–2. https://doi.org/10.1093/bioinformatics/btp187.

Liu H, Ye H, Zhang N, Ma J, Wang J, Hu G, et al. Comparative analyses of chloroplast genomes provide comprehensive insights into the adaptive evolution of Paphiopedilum (Orchidaceae). Horticulturae. 2022;8:391. https://doi.org/10.3390/horticulturae8050391.
Magdy M, Ou L, Yu H, Chen R, Zhou Y, Heba Hassan, et al. Pan-plastome approach empowers the assessment of genetic variation in cultivated *Capsicum* species. Hort Res. 2019;6:108. https://doi.org/10.1038/s41438-019-0191-x.
Minh B, Schmidt H, Chernomor O, Schrempf D, Woodhams M, 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. https://doi.org/10.1093/molbev/msaa015.
Mower JP, Guo W, Partha R, Fan W, Levsen N, Wolff K, et al. Plastomes from tribe Plantagineae (Plantaginaceae) reveal infrageneric structural synapormorphies and localized hypermutation for *Plantago* and functional loss of *ndh* genes from *Littorella*. Mol Phylogenet Evol. 2021;162: 107217. https://doi.org/10.1016/j.ympev.2021.107217.
Nagano Y, Mimura T, Kotoda N, Matsumoto R, Nagano AJ, Honjo MN, et al. Phylogenetic relationships of Aurantioideae (Rutaceae) based on RAD-Seq. Tree Genet Genomes. 2018;14:6. https://doi.org/10.1007/s11295-017-1223-z.
Narasimhan V, Danecek P, Scally A, Xue Y, Tyler-Smith C, Durbin R. BCFtools/RoH: a hidden Markov model approach for detecting autozygosity from next-generation sequencing data. Bioinformatics. 2016;32:1749–51. https://doi.org/10.1093/bioinformatics/btw044.
Ogoma C, Liu J, Stull G, Wambulwa M, Oyebanji O, Milne R, et al. Deep insights into the plastome evolution and phylogenetic relationships of the Tribe Urticeae (Family Urticaceae). Front Plant Sci. 2022;13: 870949. https://doi.org/10.3389/fpls.2022.870949.
Ortiz E. vcf2phylip v2. 0: convert a VCF matrix into several matrix formats for phylogenetic analysis. 2019. https://doi.org/105281/zenodo, 2540861. Accessed 13 Nov 2023
Palmer JD. Chloroplast DNA evolution and biosystematic uses of chloroplast DNA variation. Am Nat. 1987;130:6–29. https://doi.org/10.1086/284689.
Patil, I. Visualizations with statistical details: The 'ggstatsplot' approach. J Open Source Softw. 2021;6:3167. https://doi.org/10.21105/joss.03167
Poplin R, Chang PC, Alexander D, Schwartz S, Colthurst T, Ku A, et al. A universal SNP and small-indel variant caller using deep neural networks. Nat Biotechnol. 2018;36:983–7. https://doi.org/10.1038/nbt.4235.
Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007;81:559–75. https://doi.org/10.1086/519795.
Qu XJ, Moore MJ, Li DZ, Yi TS. PGA: a software package for rapid, accurate, and flexible batch annotation of plastomes. Plant Methods. 2019;15:50. https://doi.org/10.1186/s13007-019-0435-7.
Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010;26:841–2. https://doi.org/10.1093/bioinformatics/btq033.
Rao MJ, Zuo H, Xu Q. Genomic insights into citrus domestication and its important agronomic traits. Plant Commun. 2021;2: 100138. https://doi.org/10.1016/j.xplc.2020.100138.
Raubeson LA, Jansen RK. Chloroplast genomes of plants. In: Henry RJ, editor. Plant diversity and evolution: genotypic and phenotypic variation in higher plants.Cambridge: CAB International; 2005:45–68.
Rausch T, Zichner T, Schlattl A, Stütz AM, Benes V, Korbel JO. DELLY: structural variant discovery by integrated paired-end and split-read analysis. Bioinformatics. 2012;28:i333–9. https://doi.org/10.1093/bioinformatics/bts378.
Saarela JM, Burke SV, Wysocki WP, Barrett MD, Clark LG, Craine JM, et al. A 250 plastome phylogeny of the grass family (Poaceae): topological support under different data partitions. PeerJ. 2018;6: e4299. https://doi.org/10.7717/peerj.4299.
Shen W, Le S, Li Y, Hu F. SeqKit: a cross-platform and ultrafast toolkit for FASTA/Q file manipulation. PLoS ONE. 2016;11: e0163962. https://doi.org/10.1371/journal.pone.0163962.
Wang X, Xu Y, Zhang S, Cao L, Huang Y, Cheng J, et al. Genomic analyses of primitive, wild and cultivated citrus provide insights into asexual reproduction. Nat Genet. 2017;49:765–72. https://doi.org/10.1038/ng.3839.
Wang L, He F, Yue H, He J, Yang S, Zeng J, et al. Genome of Wild Mandarin and Domestication History of Mandarin. Mol Plant. 2018;11:1024–37. https://doi.org/10.1016/j.molp.2018.06.001.
Wang N, Li C, Kuang L, Wu X, Xie K, Zhu A, et al. Pan-mitogenomics reveals the genetic basis of cytonuclear conflicts in citrus hybridization, domestication, and diversification. Proc Natl Acad Sci. 2022b;119: e2206076119. https://doi.org/10.1073/pnas.2206076119.
Wang J, Liao X, Gu C, Xiang K, Wang J, et al. The Asian lotus (Nelumbo nucifera) pan-plastome: diversity and divergence in a living fossil grown for seed, rhizome, and aesthetics. Ornamental Plant Research. 2022. https://doi.org/10.48130/OPR-2022-0002
Wei L, Xiang XG, Wang YZ, Li ZY. Phylogenetic relationships and evolution of the Androecia in Ruteae (Rutaceae). PLoS ONE. 2015;10: e0137190. https://doi.org/10.1371/journal.pone.0137190.
Wickham H. ggplot2. Wiley Interdiscip Rev Comput Stat. 2011;3:180–5. https://doi.org/10.1002/wics.147.
Wu GA, Terol J, Ibanez V, Lopez-Garcia A, Perez-Roman E, Borreda C, et al. Genomics of the origin and evolution of Citrus. Nature. 2018;554:311–6. https://doi.org/10.1186/s12864-015-1926-1.
Yan LJ, Zhu ZG, Wang P, Fu CN, Guan XJ, Kear P, et al. Comparative analysis of 343 plastid genomes of *Solanum* section *Petota*: Insights into potato diversity, phylogeny, and species discrimination. J Syst Evol. 2022;61:599–612. https://doi.org/10.1111/jse.12898.
Yang Z. PAML 4: Phylogenetic analysis by maximum likelihood. Mol Biol Evol. 2007;24:1586–91. https://doi.org/10.1093/molbev/msm088.
Yi X, Gao L, Wang B, Su YJ, Wang T. The complete chloroplast genome sequence of *Cephalotaxus oliveri* (Cephalotaxaceae): Evolutionary comparison of *Cephalotaxus* chloroplast DNAs and insights into the loss of inverted repeat copies in gymnosperms. Genome Biol Evol. 2013;5:688–98. https://doi.org/10.1093/gbe/evt042.
Yun T, Li H, Chang PC, Lin MF, Carroll A, McLean CY. Accurate, scalable cohort variant calls using DeepVariant and GLnexus. Bioinformatics. 2020;36:5582–9. https://doi.org/10.1093/bioinformatics/btaa1081.
Zhang SD, Jin JJ, Chen SY, Chase MW, Soltis DE, Li HT, et al. Diversification of Rosaceae since the late cretaceous based on plastid phylogenomics. New Phytol. 2017;214:1355–67. https://doi.org/10.1111/nph.14461.
Zhou J, He W, Wang J, Liao X, Xiang K, et al. The pan-plastome of tartary buckwheat (*fagopyrum tataricum*): Key insights into genetic diversity and the history of lineage divergence. BMC Plant Biol. 2023;23:212. https://doi.org/10.1186/s12870-023-04218-7.
Zhu A, Guo W, Gupta S, Fan W, Mower J. Evolutionary dynamics of the plastid inverted repeat: The effects of expansion, contraction, and loss on substitution rates. New Phytol. 2015;209:1747–56. https://doi.org/10.1111/nph.13743.

Publisher's Note