Plant Systematics and Evolution (2023) 309:31
https://doi.org/10.1007/s00606-023-01868-w

RESEARCH

Complete chloroplast genomes of four *Atalantia* (Rutaceae) species: insights into comparative analysis, phylogenetic relationships, and divergence time estimation

Wenbo Shi[1] · Weicai Song[1] · Yuqi Zhao[1] · Chao Shi[1,2] · Shuo Wang[1]

Received: 28 January 2023 / Accepted: 17 July 2023 / Published online: 16 August 2023

Abstract

Atalantia plants are small trees in the family Rutaceae that have great potential for use in agriculture, medicine, and carving. However, the phylogenetic relationships of *Atalantia* species have not been systematically studied due to the lack of available molecular resources and highly variable molecular markers. In this study, the complete chloroplast genomes of four *Atalantia* species were newly sequenced and characterized. The length of the chloroplast genomes in *Atalantia* species ranged from 159,557 to 160,341 bp, which was more similar to *Fortunella* and *Citrus* species in the Rutaceae. The *rpl*22 genes were pseudogenized at the SSC/IRa boundary. Certain mechanisms have been evolved to compensate for the pseudogenization of the *acc*D and *inf*A genes in *Atalantia* chloroplast genomes. Some highly divergent intergenic regions (*trn*H-GUG–*psb*A, *rps*16–*trn*Q-UUG, *trn*S-GCU–*trn*G-UCC, *atp*F–*atp*I, *pet*N–*psb*M, *psb*M–*trn*D-GUC, *ndh*C–*trn*V-UAC, *atp*B–*rbc*L, *acc*D–*psa*I, and *ycf*4–*cem*A) and genes (*psb*M, *trn*D-GUC, *rpl*32, *ndh*F, and *ycf*1) were detected, which serve as potential phylogenetic markers. The phylogenetic relationships and divergence times of *Atalantia* species were explored based on the complete chloroplast genome. Based on the phylogenetic trees, *Atalantia* species were grouped into two sub-branches. *Atalantia monophylla* and *A. buxifolia* were found to be more closely related, whereas *A. ceylanica* and *A. roxburghiana* exhibited a closer relationship. The divergence time estimation indicated that the common ancestor of the *Atalantia* species diverged approximately 14.26 million years ago. These findings will help with the identification, categorization, and utilization of *Atalantia* species.

Keywords *Atalantia* · Chloroplast genome · Comparative analysis · Divergence times · Molecular markers · Phylogeny

Introduction

The genus *Atalantia* (Rutaceae, Aurantioideae) contains approximately 17 species, which are distributed across India, Southeast Asia, and China (Bayer et al. 2009; Esser 2021). *Atalantia* plants are small trees or shrubs. They produce fragrant white flowers and small spherical fruits that resemble small greenish-yellow oranges or tangerines (Grosser et al. 1996; Ranade et al. 2009). In classical morphology, *Atalantia* species are classified as near citrus fruit trees (Swingle and Reece 1967). The *Atalantia* species have a broad range of applications. Its roots and leaves are used in herbal remedies, in combination with other herbs, to treat coughs, colds, and fevers (Das and Swamy 2016; Li and Xing 2016; Pang et al. 2020). The yellowish-white, dense, and strong *Atalantia* wood is ideal for intricate carving (Scora et al. 1969). Moreover, *Atalantia* species contain alkaloids and flavonoid glycosides that exhibit anti-insect and antibacterial

Handling Editor: Julien Boutte.

✉ Chao Shi
chsh1111@aliyun.com

✉ Shuo Wang
shuowang@qust.edu.cn

[1] College of Marine Science and Biological Engineering, Qingdao University of Science and Technology, Qingdao 266042, China

[2] 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

properties, making them potent and eco-friendly insecticides (Baskar et al. 2018). *Atalantia* species also make excellent graft unions with citrus fruit trees (Herrero et al. 1996). Using *Atalantia* species to develop citrus rootstocks can significantly reduce virus and weather damage (Alves et al. 2021). Therefore, there is a need to study *Atalantia* species to establish a foundation for a more comprehensive and efficient utilization of *Atalantia* resources.

To date, most studies on the genus *Atalantia* have primarily focused on its chemical characteristics (Chukaew et al. 2008; Fernando and Soysa 2014; Yang et al. 2015). Nevertheless, the phylogenetic system of the *Atalantia* species, which could provide a foundation for identification and molecular breeding, requires further research. Bayer et al. (2009) analyzed the phylogenetic relationships among Aurantioideae species using nine chloroplast sequences to aid in the taxonomy of *Atalantia* species. Additionally, studies using chloroplast *rbc*L and *mat*K genes have further uncovered the phylogenetic relationships of some *Atalantia* species (Penjor et al. 2010, 2013). However, these gene fragments may not represent the complete genome due to a lack of polymorphic loci (Guo et al. 2022). Moreover, the RAD-Seq study on Aurantioideae species, which involves extensive DNA sequences, contradicts previous studies that employed chloroplast gene fragments (Nagano et al. 2018). Thus, additional analysis is required to determine the phylogenetic relationships among *Atalantia* species. Herein, we aim to use the complete chloroplast genome to investigate the phylogenetic relationships of *Atalantia* species and compare them to earlier studies.

The chloroplast (cp), believed to have arisen from endosymbiosis between a photosynthetic bacterium and a non-photosynthetic host (Yang et al. 2010), has its own genetic system (Shimada and Sugiura 1991). It contains all the structures necessary for photosynthesis and plays a critical role in other essential plant activities (Leister 2003). Cp genomes are widely studied in plant molecular biology (Dobrogojski et al. 2020), and their gene content, organization, and structure are highly conserved (Song et al. 2022c). Most cp genomes have a typical quadripartite structure, including two equal inverted repeat (IR) regions separated by a small single copy (SSC) region and a large single copy (LSC) region (Song et al. 2022b). Despite being relatively conserved, the entire cp genome contains sufficiently informative loci and highly variable regions. These loci and regions can aid in determining the origin, evolution, and phylogenetic relationships of species (Song et al. 2022a; Zhang et al. 2022). Moreover, a comparative cp genome analysis of various species can aid in understanding their evolution and classification (Liu et al. 2021; Xia et al. 2022). Therefore, a comprehensive study of the cp genome in the genus *Atalantia* is particularly relevant.

With the development of high-throughput sequencing technologies, many cp genome sequences are available in the National Center for Biotechnology Information (NCBI) databases (Sayers et al. 2021). However, only two complete cp genomes of *Atalantia* species (*A. buxifolia* and *A. kwangtungensis*) are accessible in the NCBI database (Zhu et al. 2018). Here, we present the complete de novo assembly of cp genomes from four *Atalantia* species (*A. ceylanica*, *A. monophyla*, *A. racemose*, and *A. roxburghiana*). This study provides the first comparative and phylogenetic analysis of species within the genus *Atalantia*. The main objectives were to: (a) characterize the cp genomes of *Atalantia* species; (b) explore informative loci and highly variable regions in the cp genome of *Atalantia* species; and (c) determine the phylogenetic relationships and divergence times of *Atalantia* species based on the complete cp genome.

Materials and methods

Plant materials, DNA extraction, and sequencing

The plant materials of four *Atalantia* species (*A. ceylanica*, *A. monophyla*, *A. racemose*, and *A. roxburghiana*) were collected from the Plant Germplasm and Genomics Center, Kunming Institute of Botany, and the Chinese Academy of Sciences. For each species, four representative individuals were selected and collected in compliance with local laws. The authors identified all of the samples, and voucher specimens were deposited in the Evolutionary Biology Laboratory of Qingdao University of Science and Technology (QUST). Approximately 30 g of leaves from each *Atalantia* species was collected as a sample. The DNA was extracted from silica gel-dried leaf tissues using the modified high salt approach (Shi et al. 2012). The DNA quality was tested using 1% agarose gel electrophoresis (Green and Sambrook 1972). The DNA was randomly fragmented to create paired-end libraries with an average insert size of about 300 bp in length, according to the guidelines provided by the manufacturer (Illumina, San Diego, CA, USA). Genomic DNA was sequenced at Novogene (Beijing, China) on the Illumina Hiseq 4000 platform (Illumina, San Diego, CA, USA), following the manufacturer's instructions. Finally, two pair-end raw reads, each measuring 151 bp, were obtained.

De novo assembly and annotation of cp genomes

Clean, high-quality reads were generated for each sample after removing adaptors and low-quality reads using Trimmomatic v0.40 (Bolger et al. 2014). The high quality of newly produced clean short reads was assessed using FastQC v0.11.9 (Brown et al. 2017). NOVOPlasty v4.3.1 (Dierckxsens et al. 2017) was used to de novo assemble

the cp genomes of four *Atalantia* species. The entire cp genomes were validated and calibrated using GetOrganelle v1.7.4 (Jin et al. 2020). To assess the assembly results, Minimap v2.17 (Li 2018) was used to map raw reads to the assembled chloroplast genome sequences. Samtools v1.9 (Li et al. 2009) was used to detect and annotate variants in mapping results. The complete cp genome of *Atalantia* species obtained from the assembly was annotated using GeSeq v1.42 (Tillich et al. 2017). The Plastid Genome Annotator (PGA) (Qu et al. 2019) was used to compare and supplement the results of cp genome annotation. Codons and gene boundaries in the cp genomes were manually adjusted using Sequin v16 (Lehwark and Greiner 2018), with *A. buxifolia* (NC_063578) and *Citrus madurensis* (NC_064137) cp genomes used as references. Additionally, using the aforementioned method, the complete cp genome sequences of *A. buxifolia* (NC_063578) and *A. kwangtungensis* (MH329190) were re-annotated. The four newly sequenced *Atalantia* cp genomes were submitted to GenBank under the accession numbers NC_065396–NC_065399.

Characterization of chloroplast genomes in *Atalantia* species

The fundamental features of the cp genomes of various *Atalantia* species were analyzed using Geneious v9.0.2 (Kearse et al. 2012). These features comprise the length of each sequence in different regions, including LSC, SSC, and IR regions; the proportion of different sequences; the gene composition, which includes protein-coding genes, tRNA genes, rRNA genes, introns, and exons; and the GC content of different regions. OGDRAW v1.3.1 (Greiner et al. 2019) was used to generate the circular genome maps of *Atalantia* species.

Structural analysis of the cp genomes

Geneious v9.0.2 was used to extract all protein-coding genes from each complete cp genome to perform codon preference analysis. Then, MEGA v11 (Tamura et al. 2021) was used to calculate the codon usage frequency and relative synonymous codon usage (RSCU) for various *Atalantia* species based on protein-coding genes. TBtools v1.098761 (Chen et al. 2020) was used to generate the heat map for the RSCU analysis. MACSE v2 (Ranwez et al. 2018) was applied to the protein-coding genes of various *Atalantia* species. The assessment of the ratio between non-synonymous (Ka) and synonymous (Ks) substitution rates was conducted using KaKs_Calculator v2 (Wang et al. 2010). By comparing with *A. roxburghiana*, the Ka/Ks values for each *Atelantia* species were determined. Using MISA v2.1 (Beier et al. 2017), a Perl program, simple sequence repeats (SSRs) in the cp genomes were detected. The settings were determined as follows: nine for mononucleotide repeats, four for dinucleotide repeats, and three for trinucleotide, tetranucleotide, pentanucleotide, and hexanucleotide repeats. Additionally, REPuter v2 (Kurtz et al. 2001) was used to examine the forward (F), palindromic (P), reverse (R), and complement (C) repeats. The hamming distance was set to three with a minimum repeat size of 30 bp and a maximum repeat size of 300 bp.

Genomes comparison

The cp genome sequences from various *Atalantia* species were aligned using mVISTA (Frazer et al. 2004) to investigate the divergence. The alignment was performed using the LAGAN model (Brudno et al. 2003), with *A. roxburghiana* serving as the reference species. MAFFT v7.308 (Katoh and Standley 2013) was used to align the cp genome sequences of all *Atalantia* species. Then, the DNA polymorphism module of DnaSP v6.12.03 (Rozas et al. 2017) was used for nucleotide variability (Pi) analysis with a selected step size of 200 sites and a window length of 800 sites.

Phylogenetic analysis

For the phylogenetic analysis, we selected 30 Aurantioideae species (including four *Atalantia* species newly uploaded) that have complete cp genomes in the NCBI database (Online Resource 1). *Toddalia asiatica* (NC_056094) and *Zanthoxylum madagascariense* (NC_046744) were used as the outgroups. MAFFT v7.308 was used to align all of the complete cp genome sequences with default settings. To determine the optimal evolutionary model for the aligned sequences, we employed jmodeltest v2.1.10 (Darriba et al. 2012) and selected the best model based on the Bayesian information criterion (BIC) selection results. The TPM1uf + I + G model was selected for further analysis. The maximum likelihood (ML) phylogenetic tree was constructed using IQ-Tree v2.1.2 (Trifinopoulos et al. 2016), and bootstrap replications were performed 1000 times to determine branch support values. Additionally, the Bayesian inference (BI) phylogenetic tree was constructed using MrBayes v3.2.7a (Huelsenbeck and Ronquist 2001). BI analysis involved the parallel running of four Markov Chain Monte Carlo (MCMC). A sampling tree was constructed after every 1000 generations in the MCMC method, which calculated 2,000,000 generations. The first 25% of generations were eliminated as "burn-in," with the remaining trees used to construct the consensus tree.

Divergence time estimation

Using the MCMCtree module (Puttick 2019) in PAML v4.9 (Yang 2007), the divergence times of Aurantioideae species

were estimated based on the complete cp genomes. To estimate divergence times, we prepared three necessary files: a setup file for running the program, a sequence alignment file, and a phylogenetic tree file including fossil calibration points. The fasta to phy format conversion was performed on the cp genome sequence alignment file used in the phylogenetic analysis. The phylogenetic tree file was obtained using IQ-Tree v2.1.2. Four fossil calibration points were chosen to limit each node in accordance with earlier studies: The node of the branch clustered by the *Atalantia* species was set at 5.3–23.3 million years ago (Ma) (Soomro et al. 2021), the node of the branch containing the *Citrus reticulata* was set at 8 Ma (Xie et al. 2013), the node of the branch containing the *C. aurantium* was set at 1.8–5.3 Ma (Fischer and Butzmann 1998), and the node of the outgroups (*Toddalia asiatica* and *Zanthoxylum madagascariense*) was set at 35–40 Ma (Gregor 1989). In the setup file, the key clock and model parameters were set to correlated rates and HKY85, respectively. The first 2000 iterations were eliminated as "burn-in," and samples were collected every 10 iterations before 25,000 were collected. The default settings were used for the remaining parameters.

Results

The features of various *Atalantia* species

The mapping results showed that the coverage of the assembly of the complete cp genomes was relatively homogeneous (Online Resource 2). The complete cp genomes of various *Atalantia* species were circular, double-stranded structures. All *Atalantia* species exhibited a typical quadripartite structure, in which the LSC region and the SSC region were separated by two IR regions (Fig. 1 and Online Resource 3). The cp genome of *A. monophylla* was the largest (160,341 bp) among *Atalantia* species, followed by *A. roxburghiana* (160,288 bp). *A. racemosa* had the shortest cp genome (159,557 bp). The length was 27,049–27,154 bp in the IR regions. The LSC region was longer, ranging from 87,165 to 87,768 bp, while the SSC region was smaller, ranging from 18,319 to 18,457 bp. The GC content of the complete cp genomes was 38.4–38.5%. The GC content of the LSC, SSC, and IR regions was 36.8–36.9%, 33.3–33.5%, and 42.9%, respectively (Table 1). Despite their different cp genome lengths, the genetic analysis showed some similarities among *Atalantia* species. Herein, we identified 127–133 functional genes, including 83–88 protein-coding genes, 35–38 tRNA genes, and 8 rRNA genes, in the complete cp genomes of *Atalantia* species (Table 1). These genes can be generally categorized into four groups: genes involved in photosynthesis; self-replicating genes; other genes; and genes with unidentified activities (Table 2). In the cp genomes of the *Atalantia* species analyzed, the *acc*D and *inf*A genes became potential pseudogenes (Online Resource 3).

The contraction and expansion of IR regions

The structure of genes around IR regions and junction sites was identified using the IRscope. The four vertical lines in Fig. 2 represent the JLB (IRb/LSC), JSB (IRb/SSC), JSA (SSC/IRa), and JLA (IRa/LSC) junctions. Near the IR boundaries, the genes *rps*3, *rpl*22, *rps*19, *ndh*F, *ycf*1, *rps*2, *trn*H, and *psb*A were detected. The *ndh*F genes of *A. ceylanica*, *A. monophylla*, *A. racemose*, and *A. buxifolia* crossed the IRb and SSC regions. The *ndh*F genes of *A. roxburghiana* and *A. kwangtungensis* were all located in the SSC region. The functional *ycf*1 genes found at the JSA junction of all six species had a relatively consistent length. The length of the *ycf*1$^{\Psi}$ pseudogenes at the JSB junction varied from 1045 to 1098 bp. The length and location of the *rpl*22 genes were relatively consistent at the JLB junction. Interestingly, the *rpl*22 gene at the JLA junction of *A. monophylla* was shorter than the *rpl*22 gene at the JLB junction. The *rpl*22 genes in other *Atalantia* species (*A. ceylanica*, *A. racemosa*, *A. roxburghiana*, *A. buxifolia*, and *A. kwangtungensis*) overlapped with the *trn*H genes in the LSC region.

Codon usage bias and selective pressure analysis

Codon usage preference and relative synonymous codon usage (RSCU) were analyzed based on protein-coding genes in the cp genomes of *Atalantia* species. A heatmap (Fig. 3) based on the RSCU results for six *Atalantia* species was constructed to facilitate data comprehension. The evolutionary tree classified 64 codons into two major branches based on RSCU values. One branch comprised 20 codons with RSCU values greater than 1.38, while the other contained the remaining codons. These six *Atalantia* species displayed high similarity in their codon use frequencies (Online Resource 4). Codons ending with an A or T base had higher coding rates. Except for CTA (leucine), ATA (isoleucine), and TGA (stop codon), codons ending with A or T had RSCU > 1, while codons ending with C or G had RSCU < 1. Although some variations were observed, most amino acids have at least two synonymous codons, with Arg, Leu, and Ser having a total of six. Both tryptophan (UGG) and methionine (AUG) had RUSC values of 1. Based on RSCU values, *A. buxifolia* and *A. kwangtungensis* formed one branch, and the other four species (*A. monophylla*, *A. ceylanica*, *A. racemosa*, and *A. roxburghiana*) formed the other branch.

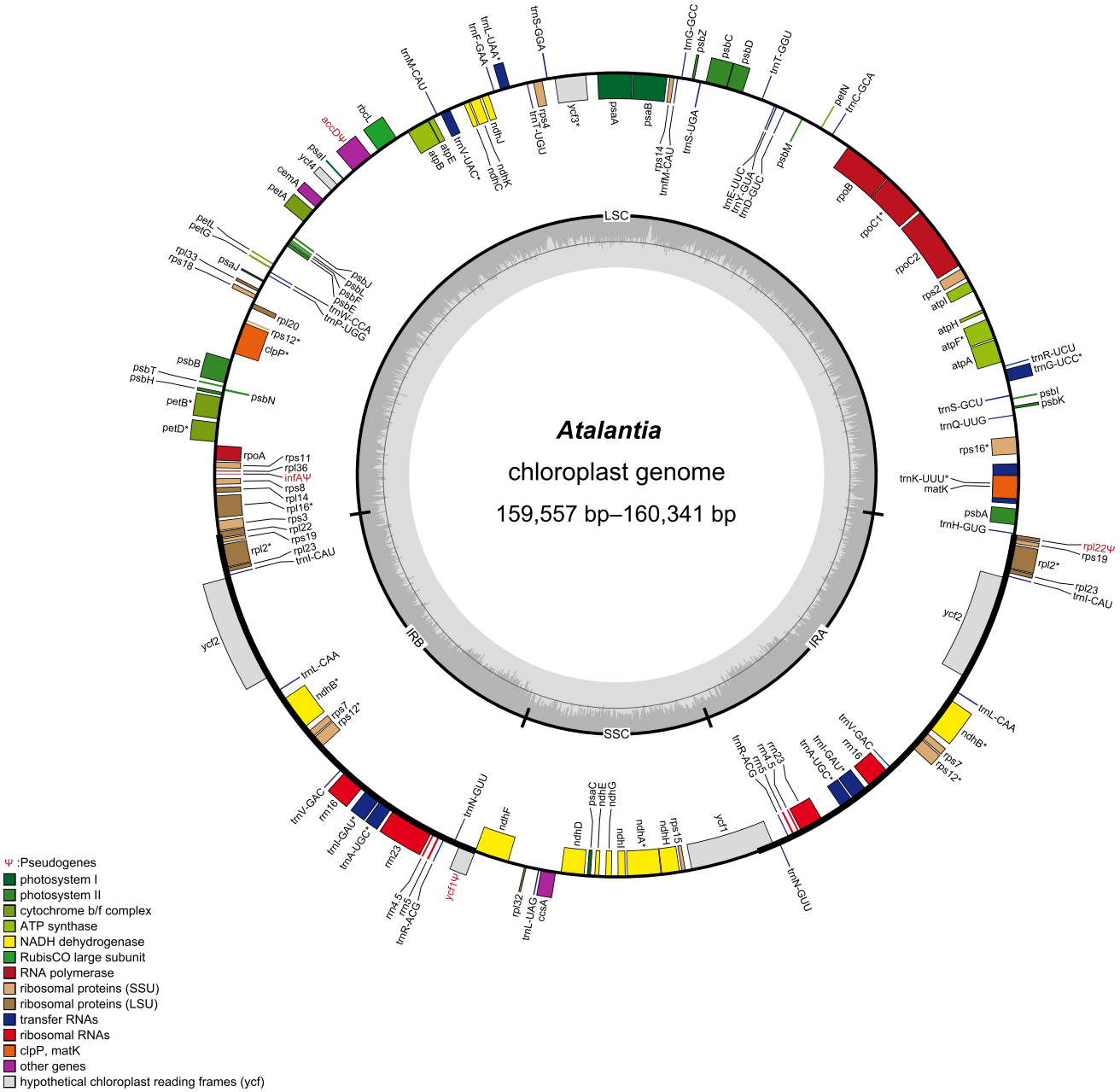

Fig. 1 The circular map of the *Atalantia* chloroplast genomes. On the inner and outer sides of the circle, genes are displayed. Genes are transcribed clockwise in the inner circle and counterclockwise in the outer circle, respectively. The lower left corner displays many colored genes, each of which serves a different purpose. Dark-grey and light-grey inner circles, which represent GC and AT contents, respectively, are divided. The large single copy (LSC), the small single copy (SSC), and the inverted repeat (IRA and IRB) regions are also represented. The partial outer circles reveal the gene losses in GenBank files for *A. buxifolia* and *A. kwangtungensis*

In analyzing selective pressure, the Ka/Ks ratio, which represents the ratio of non-synonymous substitutions to synonymous substitutions, is used to evaluate whether there is selective pressure on protein-coding genes. Using *A. roxburghiana* as a reference, the Ka/Ks values of protein-coding genes in the cp genomes of *Atalantia* species were analyzed (Online Resource 5). Results indicated that most genes had a Ka/Ks ratio of less than 1, which aligned with the conservativeness of the cp genomes. Particularly, all *A. monophylla* genes had Ka/Ks values lower than 1. The *rpl*20 genes of *A. racemosa* and *A. kwangtungensis* had the highest Ka/Ks value (1.55) among *Atalantia*

Table 1 Summary of complete chloroplast genomes of various *Atalantia* species

Genome features	*Atalantia ceylanica*	*Atalantia monophylla*	*Atalantia racemosa*	*Atalantia roxburghiana*	*Atalantia buxifolia*	*Atalantia kwangtungensis*
Genome size (bp)	1,59,647	1,60,341	1,59,557	1,60,288	1,60,056	1,60,248
LSC size (bp)	87,165	87,768	87,086	87,695	87,561	87,483
SSC size (bp)	18,384	18,319	18,335	18,421	18,359	18,457
IR size (bp)	27,049	27,127	27,068	27,086	27,068	27,154
Total GC content (%)	38.5%	38.5%	38.5%	38.5%	38.5%	38.4%
GC content in LSC (%)	36.8%	36.8%	36.9%	36.9%	36.8%	36.8%
GC content in SSC (%)	33.5%	33.4%	33.4%	33.4%	33.4%	33.3%
GC content in IR (%)	42.9%	42.9%	42.9%	42.9%	42.9%	42.9%
Number of genes[unique]	133 [113]	133 [113]	132 [113]	132 [113]	128 [109]	127 [109]
Protein genes [unique]	87 [79]	88 [79]	87 [79]	87 [79]	84 [76]	83 [76]
tRNA genes [unique]	38 [30]	37 [30]	37 [30]	37 [30]	36 [29]	35 [29]
rRNA genes [unique]	8 [4]	8 [4]	8 [4]	8 [4]	8 [4]	8 [4]
Accession numbers in GenBank	NC_065396	NC_065397	NC_065398	NC_065399	NC_063578	MH329190

Table 2 Genes in the chloroplast genome of *Atalantia* species

Category	Gene group	Gene name
Photosynthesis	Subunits of photosystem I	*psa*A, *psa*B, *psa*C, *psa*I, *psa*J
	Subunits of photosystem II	*psb*A, *psb*B, *psb*C, *psb*D, *psb*E, *psb*F, *psb*H, *psb*I, *psb*J, *psb*K, *psb*L, *psb*M, *psb*N, *psb*T, *psb*Z
	Subunits of NADH dehydrogenase	*ndh*A*, *ndh*B*(2), *ndh*C, *ndh*D, *ndh*E, *ndh*F, *ndh*G, *ndh*H, *ndh*I, *ndh*J, *ndh*K
	Subunits of cytochrome b/f complex	*pet*A, *pet*B*, *pet*D*, *pet*G, *pet*L, *pet*N
	Subunits of ATP synthase	*atp*A, *atp*B, *atp*E, *atp*F*, *atp*H, *atp*I
	Large subunit of rubisco	*rbcL*
Self-replication	Proteins of large ribosomal subunit	*rpl*14, *rpl*16*, *rpl*2*(2), *rpl*20, *rpl*22, *rpl*23(2), *rpl*32, *rpl*33, *rpl*36
	Proteins of small ribosomal subunit	*rps*11, *rps*12**(2), *rps*14, *rps*15, *rps*16*, *rps*18, *rps*19(2), *rps*2, *rps*3, *rps*4, *rps*7(2), *rps*8
	Subunits of RNA polymerase	*rpo*A, *rpo*B, *rpo*C1*, *rpo*C2
	Ribosomal RNAs	*rrn16*(2),*rrn23*(2),*rrn4.5*(2),*rrn5*(2)
	Transfer RNAs	*trn*A-UGC*(2), *trn*C-GCA, *trn*D-GUC, *trn*E-UUC, *trn*F-GAA, *trn*G-GCC, *trn*G-UCC*, *trn*H-GUG, *trn*I-CAU(2), *trn*I-GAU*(2), *trn*K-UUU*, *trn*L-CAA(2), *trn*L-UAA*, *trn*L-UAG, *trn*M-CAU, *trn*N-GUU(2), *trn*P-UGG, *trn*Q-UUG, *trn*R-ACG(2), *trn*R-UCU, *trn*S-GCU, *trn*S-GGA, *trn*S-UGA, *trn*T-GGU, *trn*T-UGU, *trn*V-GAC(2), *trn*V-UAC*, *trn*W-CCA, *trn*Y-GUA, *trnf*M-CAU
Other genes	Maturase	*mat*K
	Protease	*clp*P**
	Envelope membrane protein	*cem*A
	Acetyl-CoA carboxylase	*acc*D
	c-type cytochrome synthesis gene	*ccs*A
	Translation initiation factor	*inf*A
Genes of unknown function	Conserved hypothetical chloroplast ORF	*ycf*1(2), *ycf*2(2), *ycf*3**, *ycf*4

*Gene with one introns; **gene with two introns; (2):number of copies of multi-copy genes

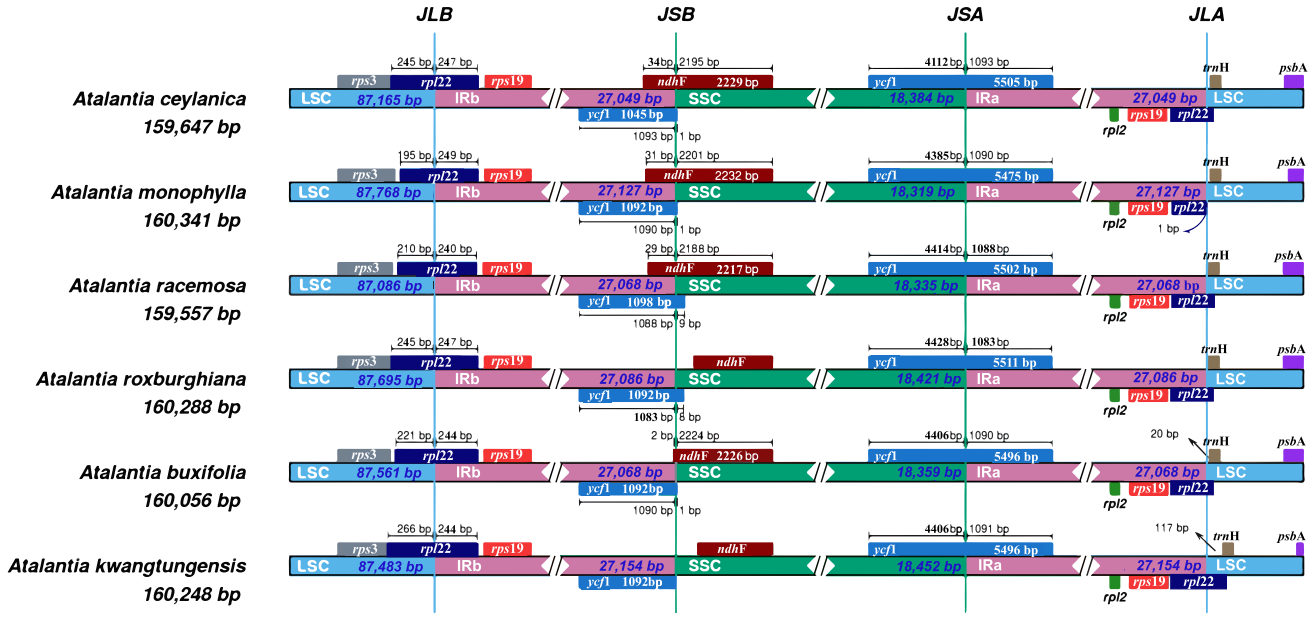

Fig. 2 The contraction and expansion of the inverted repeat/single copy (IR/SC) junctions. In the chloroplast genomes of *Atalantia* species, the junctions of the IR, short single copy (SSC), and large single copy (LSC) regions were compared. The LSC/IRb, IRb/SSC, SSC/IRa, and IRa/LSC junctions are represented by loci JLB, JSB, JSA, and JLA, respectively

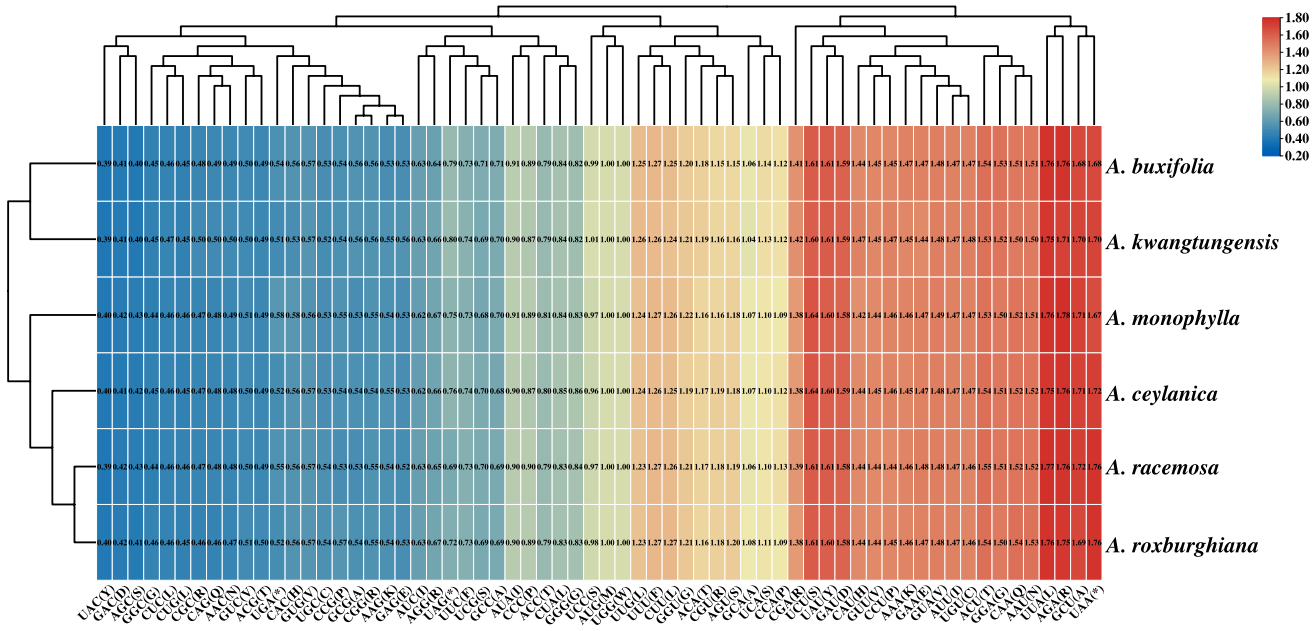

Fig. 3 The heat map for the relative synonymous codon usage (RSCU) analysis of *Atalantia* species. The RSCU values of 64 codons were used as the basis for tree clustering. As the red color deepens, the RSCU value increases. As the blue tint gets darker, the RSCU value decreases. Each column represents a different codon. Each row represents a distinct species of *Atalantia*

species. The Ka/Ks values for the *mat*K genes in *A. ceylanica*, *A. racemosa*, and *A. kwangtungensis* were greater than 1. Moreover, the *ndh*A gene in *A. ceylanica*, the *ndh*G gene in *A. kwangtungensis*, and the *rbc*L and *rpl*22 genes in *A. buxifolia* had Ka/Ks values greater than 1.

Analysis of SSRs and long repeats

In the *Atalantia* cp genomes, there were SSRs ranging from 86 (*A. racemosa*) to 103 (*A. monophylla*) (Online Resource 6). SSRs were identified and classified as mononucleotide,

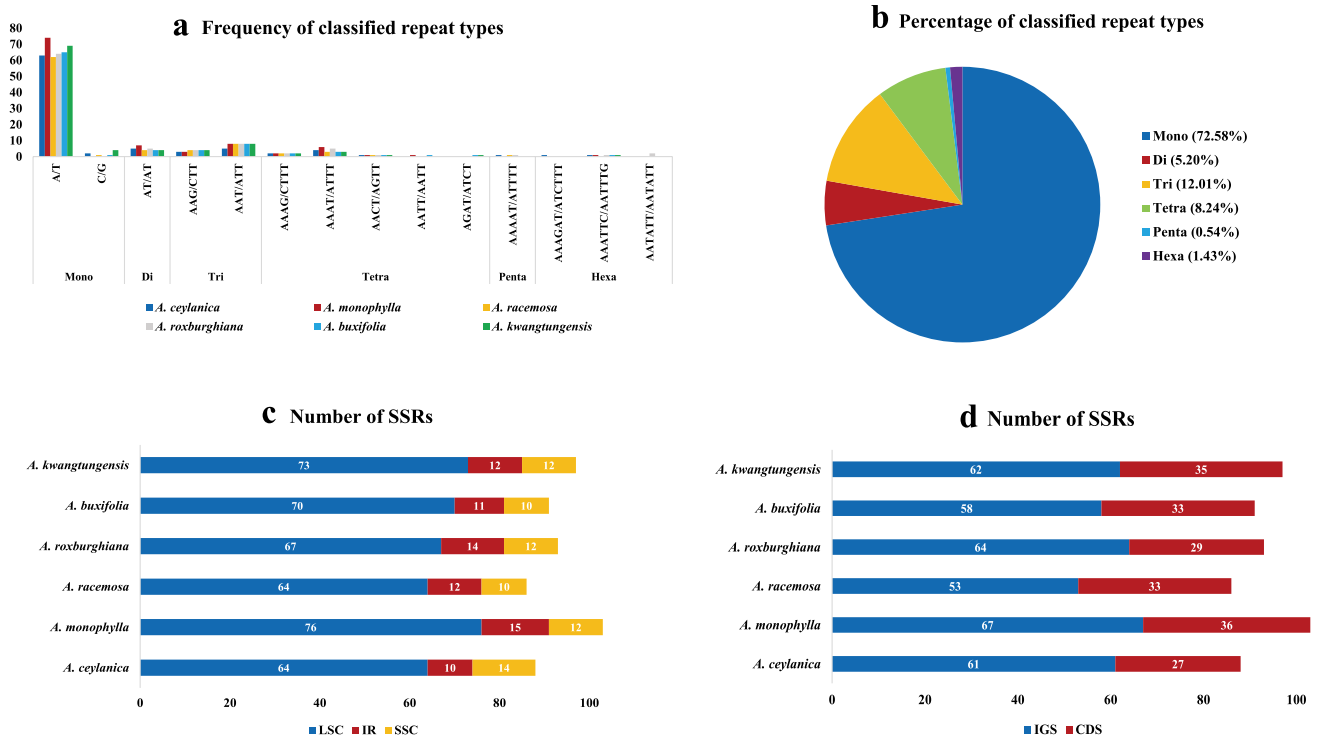

Fig. 4 Analysis of simple sequence repeats (SSRs) in *Atalantia* chloroplast genomes. **a** Number of six SSRs. **b** Type of shared SSRs among the chloroplast genomes. **c** Number of SSRs in LSC, SSC, and IR regions. **d** Number of SSRs in the coding regions (CDS) and intergenic region (IGS)

dinucleotide, trinucleotide, tetranucleotide, and pentanucleotide repeats in the cp genomes of *Atalantia* species (Fig. 4a). Most mononucleotide repeats were A/T motifs, while all dinucleotide repeats were AT/TA motifs. Mononucleotide repeats accounted for approximately 72.58% of all SSRs, followed by trinucleotide repeats (12.01%), tetranucleotide repeats (8.24%), dinucleotide repeats (5.20%), and hexanucleotide repeats (1.43%). Pentanucleotide repeats were rare (0.54%) (Fig. 4b). The distribution of SSRs in the LSC region was higher (72.04–76.92%) in comparison with the IR (11.36–15.05%) and the SSC regions (10.99–15.91%) (Fig. 4c). Based on the analysis of the SSR locations, intergenic sequences (IGS) had more SSRs than coding sequences (CDS) (Fig. 4d).

Herein, all repeats in the cp genomes of various *Atalantia* species were counted with repeat units greater than 30 bp. The total number of long repeat sequences in *Atalantia* cp genomes ranged from 34 (*A. roxburghiana*) to 58 (*A. ceylanica*) (Online Resource 7). Forward (F), palindromic (P), reverse (R), and complement (C) repeats were identified among these long repeat sequences. Forward and palindromic repeats were much more abundant than reverse and complement repeats in various *Atalantia* cp genomes (Fig. 5a). The majority of long repeat sequences were 30–35 bp in length, followed by 36–41 bp (Fig. 5b). The LSC region had a higher proportion of long repeat sequences than the IR and SSC regions (Fig. 5c). The IGS had much longer repeat sequences than the CDS in the cp genomes of various *Atalantia* species (Fig. 5d).

Comparative analysis of cp genomes in *Atalantia* species

We employed mVISTA to explore the divergence in cp genomes among *Atalantia* species (Fig. 6). The LSC and SSC regions of the cp genome displayed the highest divergence, while the IR regions exhibited very little. Furthermore, non-coding regions showed greater divergence than coding regions. The rRNA genes in *Atalantia* cp genomes showed little variation. Several intergenic regions in the six *Atalantia* cp genomes displayed the highest levels of divergence, including *trn*H-GUG–*psb*A, *rps*16–*trn*Q-UUG, *trn*S-GCU–*trn*G-UCC, *atp*F–*atp*I, *psb*M–*trn*D-GUC, *ndh*C–*trn*V-UAC, *atp*B–*rbc*L, *acc*D–*psa*I, and *ycf*4–*cem*A. Additionally, the *ndh*F and *ycf1* genes diverged in their coding regions.

The Pi values for *Atalantia* cp genomes observed a range of 0 to 0.105, with an average value of 0.007 (Online Resource 8). Interestingly, the Pi values were notably low in the IR regions, but the SSC and LSC regions showed significant divergence (Fig. 7). The LSC and SSC regions contained highly variable regions (*trn*H-GUG–*psb*A, *pet*N–*psb*M, and *psb*M–*trn*D-GUC), as well as genes (*psb*M,

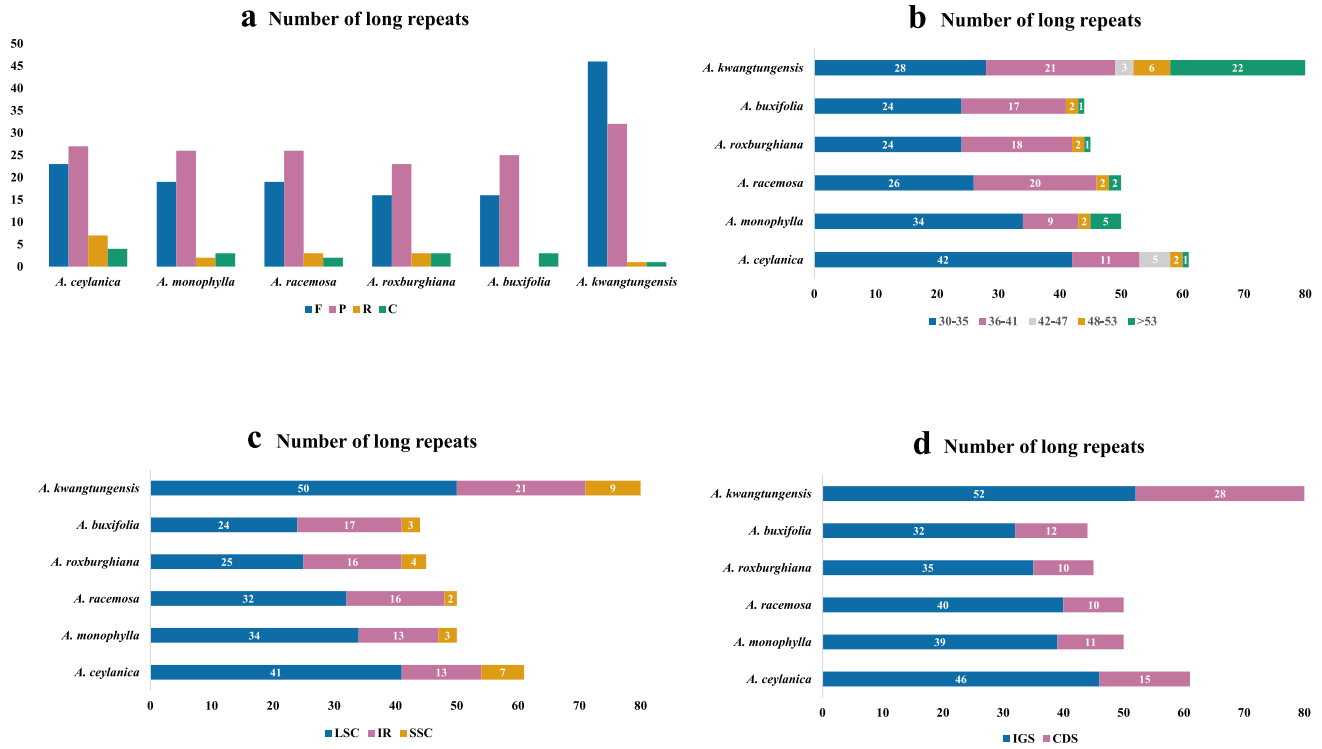

Fig. 5 Analysis of long repeats in *Atalantia* chloroplast genomes. **a** Number of different long repeats. **b** Number of long repeats by length. **c** Number of long repeats in LSC, SSC, and IR regions. **d** Number of long repeats in the coding regions (CDS), intergenic region (IGS)

*trn*D-GUC, *rpl*32, *ndh*F, and *ycf*1). In contrast, no variable sites were identified within the IR regions.

Phylogenetic analysis and divergence time estimation

The ML and BI trees of Aurantioideae species, constructed from the complete cp genomes, exhibited strong support (Fig. 8). All ML bootstrap values were higher than 77, and all Bayesian posterior probability (PP) values were greater than 0.999. The phylogenetic trees can be divided into three main branches. The basal branch consisted of species from the Clauseneae tribe, specifically the genera *Glycosmis*, *Clausena*, *Murraya*, and *Merrillia*. The second branch comprised various *Atalantia* species, formulating two sub-branches. The first sub-branch contained *A. ceylanica* and *A. roxburghiana*, and the second sub-branch included *A. kwangtungensis*, *A. buxifolia*, *A. racemosa*. Finally, the last main branch contained *Citrus* species.

The results of the divergence time estimation indicated that the Aurantioideae species diverged during the Cenozoic era, as illustrated in Fig. 9. The most recent common ancestor of Aurantioideae was estimated to have existed approximately 22.76 Ma (19.71–25.25 Ma, 95% highest posterior density (HPD)). The divergence of the *Atalantia* and *Citrus* species occurred at approximately 17.09 Ma (14.01–20.2 Ma, 95% HPD), while the *Citrus* species diverged around 10.07 Ma (7.30–13.05 Ma, 95% HPD). The common ancestor of the *Atalantia* species diverged approximately 14.26 Ma (10.98–17.52 Ma, 95% HPD). The latest divergence event was observed at approximately 0.37 Ma (0.11–0.72 Ma, 95% HPD) in *A. buxifolia* and *A. kwangtungensis*. *A. racemosa* and *A. monophylla* exhibited divergence at approximately 9.40 Ma (5.64–13.13 Ma, 95% HPD) and 12.66 Ma (9.22–15.97 Ma, 95% HPD), respectively. *A. ceylanica* and *A. roxburghiana* diverged around 8.26 Ma (4.78–11.99 Ma, 95% HPD).

Discussion

Characteristic and comparative analysis of the cp genomes

In this study, we added the complete cp genome resources of four *Atalantia* species. The length of the cp genomes in *Atalantia* species ranged from 159,557 to 160,341 bp, which is similar to that of *Fortunella* and *Citrus* species in the Rutaceae (Wang et al. 2021, 2022; Shi et al. 2023b). Several factors, such as pseudogenization, intergenic

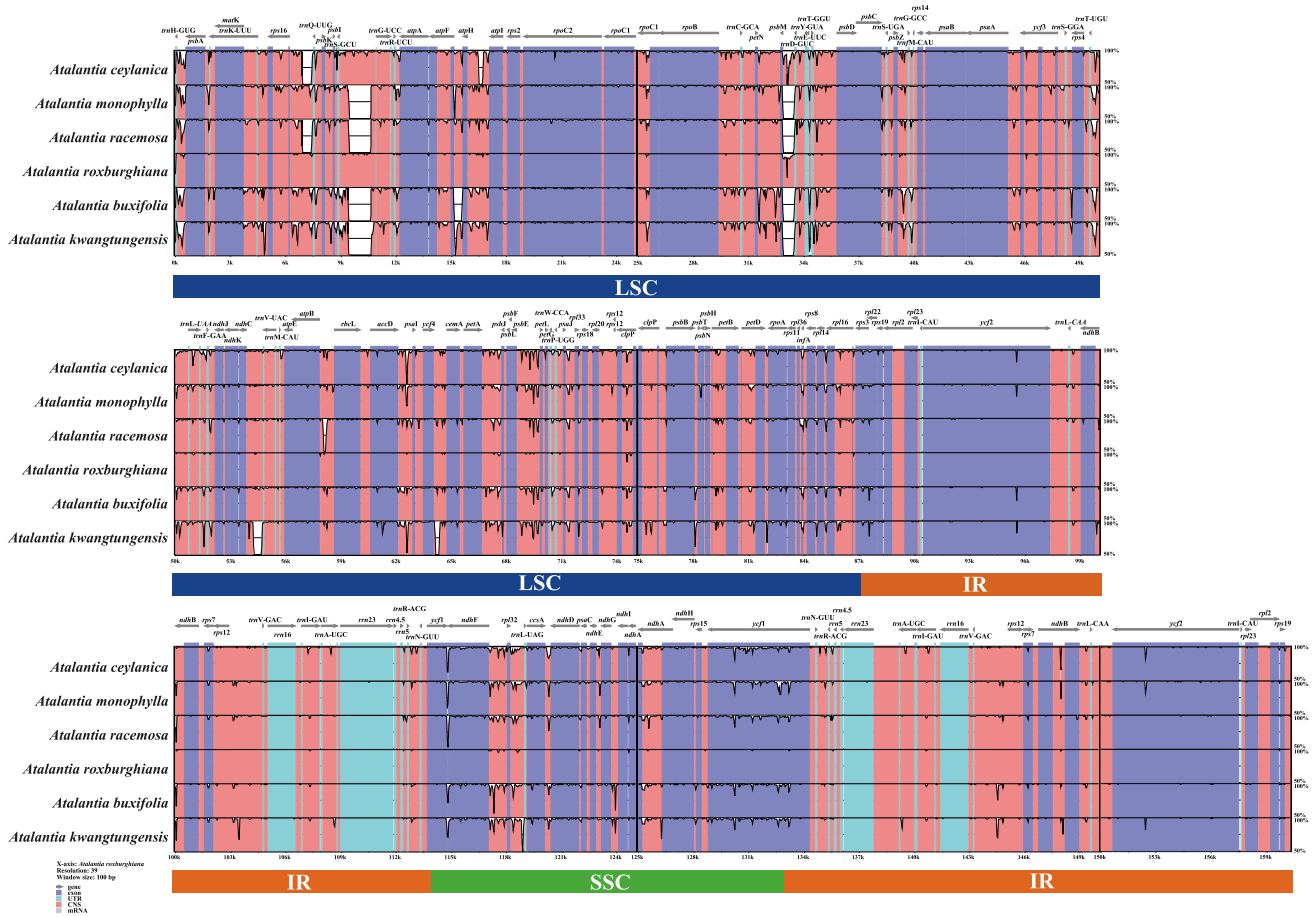

Fig. 6 Analysis of the whole chloroplast genomes of *Atalantia* species. The mVISTA was used to carry out the sequence analysis of the *Atalantia* species using *A. roxburghiana* as the reference sequence. Dark gray arrows show the direction of specific genes, while the y-axis shows the percentage of identity (50–100%). Pink bars show non-coding sequences (CNS), purple bars show exons, blue bars show RNA, and gray bars show mRNA

region divergence, and the contraction and extension of IR regions, can affect the length of cp genomes (Liu et al. 2019; Yang et al. 2021; Shi et al. 2023a). Notably, the presence of an internal stop codon (TGA) in the *inf*A genes of *Atalantia* species renders them potential pseudogenes. Likewise, the *inf*A gene in *Citrus* species is frequently pseudogenized, and the nuclear copy substitutes for its function (Millen et al. 2001; Bausher et al. 2006; Su et al. 2014; Shi et al. 2023b). Furthermore, the *acc*D gene in *A. kwangtungensis* is shorter than in other *Atalantia* species (Online Resource 9), with an internal stop codon (TAA) leading to the elimination of the zinc-binding domain and pseudogenization (Bilder et al. 2006). This is the first report of pseudogenization in the *acc*D gene among Rutaceae species (Zhao et al. 2021; Sun et al. 2021). In the initial stage of fatty acid synthesis, the *acc*D gene encodes the subunit of the rate-limiting enzyme acetyl-CoA carboxylase (ACC) (Sasaki et al. 1993; Sasaki and Nagano 2004). Compensating mechanisms are required to explain the pseudogenization of the *acc*D gene in the cp genome (Xu and Wang 2021). Some studies suggest that the *acc*D gene in the cp genome migrates to the nuclear genome, which then returns it to perform functional roles in the cp genome (Rousseau-Gueutin et al. 2013; Liu et al. 2016). Alternatively, the gene encoding the eukaryotic acetyl-CoA carboxylase receives a DNA sequence encoding a cp transition signal, potentially eliminating the need for the *acc*D gene (Konishi et al. 1996; Sasaki and Nagano 2004). However, further studies are needed to determine the mechanism responsible for the *acc*D gene in *A. kwangtungensis* being a potential pseudogene without affecting normal function.

The contraction and expansion of the IR regions can result in a shift in gene location (Song et al. 2022c). The *rpl*22 genes in the *Atalantia* species were pseudogenized at the JSA junction. In *A. monophylla*, the *rpl*22 gene at the JSA junction was truncated due to an early stop codon (TAG). Additionally, in other *Atalantia* species (*A.*

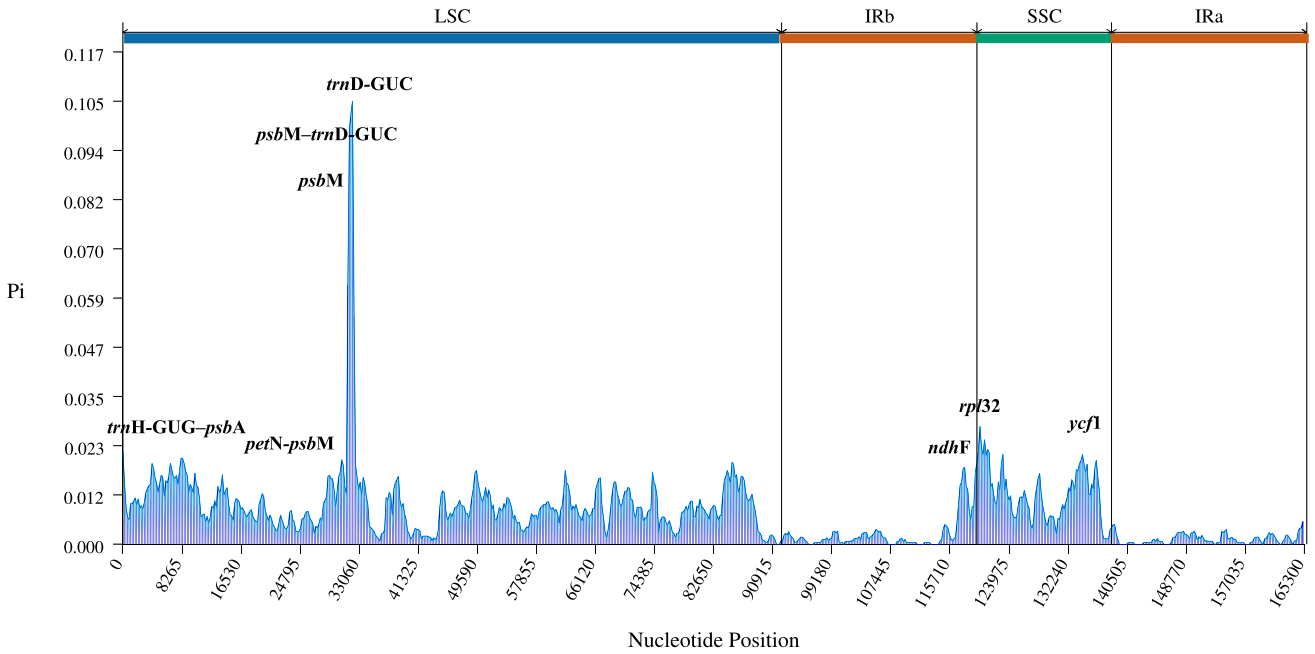

Fig. 7 Nucleotide diversity (Pi) among the whole chloroplast genomes of *Atalantia* species was analyzed using a sliding window method. The base position along the sequences is shown on the x-axis, and the Pi value is shown on the y-axis, correspondingly

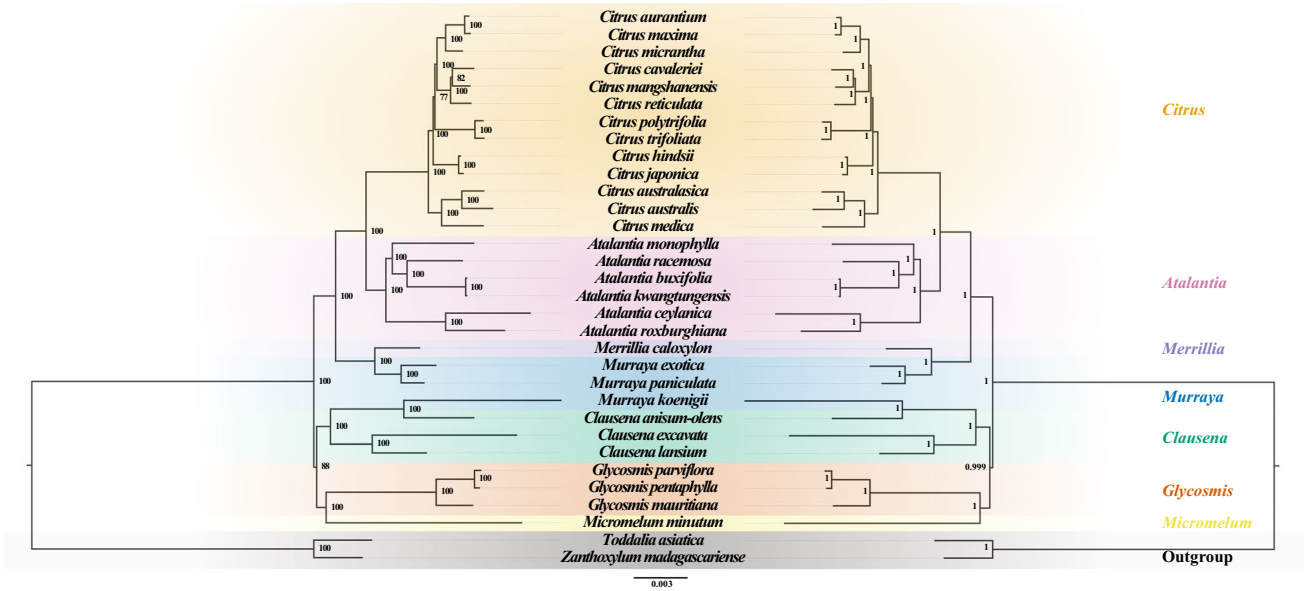

Fig. 8 Phylogenetic trees constructed with the complete chloroplast genomes of *Atalantia* species using the maximum likelihood (ML) and Bayesian inference (BI) methods. *Toddalia asiatica* and *Zanthoxylum madagascariense* were used as outgroups. Support values are indicated by the numbers above the nodes

ceylanica, *A. racemosa*, *A. roxburghiana*, *A. buxifolia*, and *A. kwangtungensis*), the *rpl*22 genes were pseudogenized at the JLA junction because they overlapped with the *trn*H genes. The expansion of the IR region may result in the overlap of the *rpl*22 genes with the *trn*H genes at the JLA junction, thereby causing pseudogenization (Liu et al. 2023). These IR boundary variations found in *Atalantia* cp genomes resemble those in *Citrus* species (Wang et al. 2021; Shi et al. 2023b).

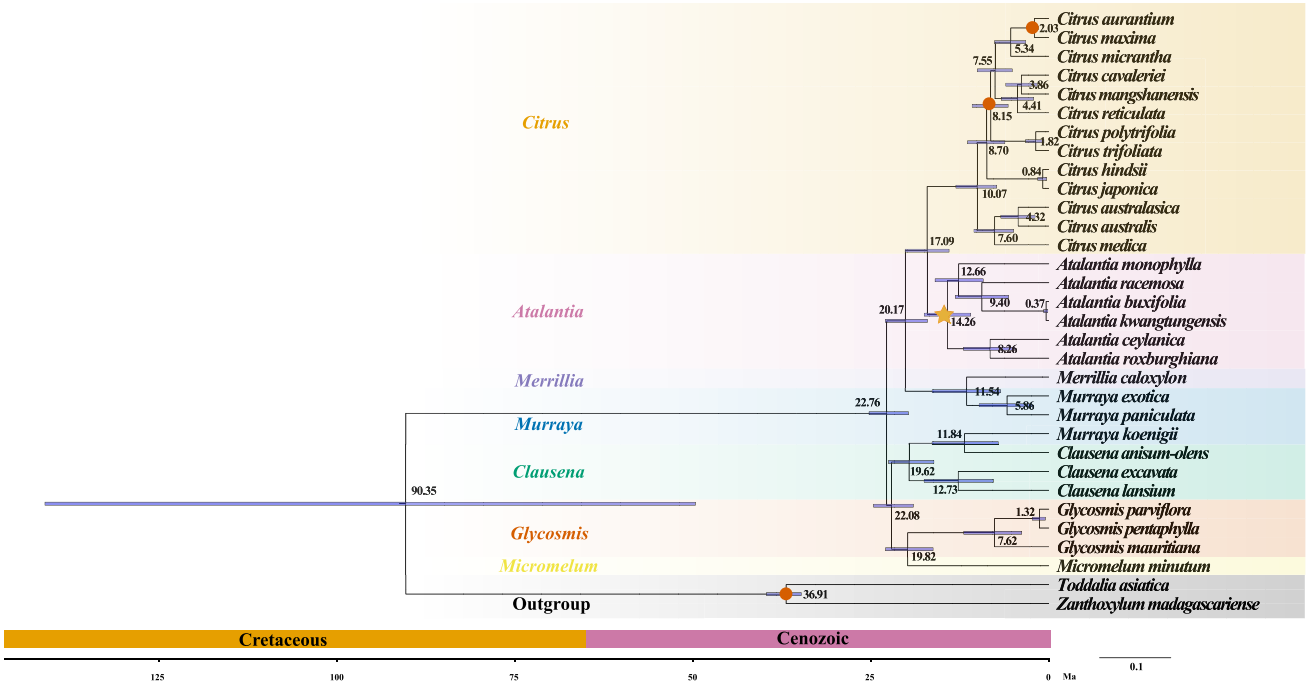

Fig. 9 Divergence time estimation based on chloroplast genome sequences. Each node shows the divergence times, and the blue bars provide the 95% greatest posterior density interval for each node age

Based on selective pressure analysis, several genes in *Atalantia* species, including *mat*K, *ndh*G, *rpl*20, and *rpl*22, have undergone positive selection during evolution. Heteroplasmy, which causes competition among diverse cytoplasmic genomes, can lead to considerable intra- and inter-individual selection of cp genome genes (Carbonell-Caballero et al. 2015). The cp genomes of *Citrus* species show a high level of heteroplasmy (Carbonell-Caballero et al. 2015). The mapping results suggest the potential existence of heteroplasmy in the *Atalantia* species (Online Resource 10). The presence of positively selected genes in *Atalantia* species, like in *Citrus* species, might contribute to their adaptation to the environment by conferring certain advantages (Carbonell-Caballero et al. 2015; Shi et al. 2023b).

Highly variable regions exist in cp genomes and can potentially distinguish closely related species or genera (Zhou et al. 2020). Based on mVISTA results and Pi values, we identified several highly divergent regions (*trn*H-GUG–*psb*A, *rps*16–*trn*Q-UUG, *trn*S-GCU–*trn*G-UCC, *atp*F–*atp*I, *pet*N–*psb*M, *psb*M–*trn*D-GUC, *ndh*C–*trn*V-UAC, *atp*B–*rbc*L, *acc*D–*psa*I, and *ycf*4–*cem*A) and genes (*psb*M, *trn*D-GUC, *rpl*32, *ndh*F, and *ycf*1). These highly variable regions and genes could be valuable for phylogenetic analysis and species identification. However, evaluating which of these regions can be used for phylogenetic analysis will require further investigation.

Phylogenetic relationships and divergence history

The taxonomy of the Aurantioideae species is challenging (Ranade et al. 2009), and while cp gene fragments have helped explore the phylogeny of Aurantioideae species (Bayer et al. 2009; Penjor et al. 2010, 2013), there are some differences between the phylogenetic analyses of Aurantioideae species based on cp gene fragments and those based on RAD-Seq (Nagano et al. 2018). Further investigation is needed to clarify the phylogenetic relationships among *Atalantia* species. This study employed ML and BI trees based on the cp genome to determine the position of the genus *Atalantia* within the Aurantioideae (Fig. 8). The genera *Atalantia* and *Citrus* are sister taxa based on phylogenetic trees of complete cp genomes. Previous studies have noted similarities in chemical makeup and botanical characteristics between the two genera (Bacher et al. 1999; Araújo et al. 2003; Wang et al. 2017). Moreover, the cp genomes of *Atalantia* species in this study share similar characteristics with *Citrus* species (Shi et al. 2023b). It has been hypothesized that *Atalantia* species have a polyphyletic origin and evolutionary history along with *Citrus* species (Araújo et al. 2003). Additionally, *Merrillia* and *Murraya* species are closely related to *Atalantia* species based on phylogenetic trees. The phylogenetic analysis based on the complete cp genome shows that *A. monophylla* and *A. buxifolia* exhibit a closer relationship, whereas *A. ceylanica*

and *A. roxburghiana* are more closely related. These findings are consistent with previous studies based on cp gene fragments (Penjor et al. 2013; Schwartz et al. 2015). However, a phylogenetic study of Aurantioideae species based on RAD-Seq indicates that *A. monophylla* and *A. roxburghiana* are more closely related, while *A. ceylanica* is more closely related to *A. buxifolia* (Nagano et al. 2018). The discrepancies observed in phylogenetic analyses using the complete cp genome and RAD-Seq may have resulted from the different evolutionary histories of cp and nuclear DNA (Nagano et al. 2018; Zhou et al. 2022). These discrepancies may arise from cp capture or reticulate evolution, commonly observed in *Citrus* and *Actinidia* species (Chat et al. 2004; Curk et al. 2014; Liu et al. 2017b; Nagano et al. 2018). Further study analyzing the whole genome, including the nuclear genome, would be required to overcome these discrepancies.

Based on the complete cp genome, we estimated divergence times for Aurantioideae species (Fig. 9). The crown of the Aurantioideae subfamily diverged around 22.76 Ma, slightly earlier than the estimated time of divergence based on cp gene fragments, which was about 19.80 Ma (Pfeil and Crisp 2008). The results of the *Citrus* species divergence time estimates differ from a previous study that used the nuclear genome for time divergence estimates (Wu et al. 2018). This discrepancy is primarily due to differences in the phylogenies based on the nuclear and cp genomes (Wu et al. 2018). Nevertheless, the age of the root of the genus *Citrus* is estimated at approximately 10.07 Ma in our study, similar to the previous estimate of 8.00 Ma (Wu et al. 2018). The median time of divergence for the genera *Citrus* and *Atalantia* on The Time Tree website (http://timetree.org/) is approximately 17.90 Ma, based on five different studies (Salvo et al. 2010; Appelhans et al. 2012; Manafzadeh et al. 2014; Koenen et al. 2015; Guo et al. 2021), which is similar to our study of about 17.09 Ma. Our analysis based on the complete cp genome showed that the genus *Atalantia* diverged around 14.26 Ma, while a multi-locus study estimated the divergence to be at 4.70 Ma (Schwartz et al. 2015). The discrepancy is likely due to different fossil constraints, and this study makes use of newly discovered Miocene era fossilized wood from the *Atalantia* genus (Soomro et al. 2021). Herein, *A. monophylla* diverged from other *Atalantia* species at about 12.66 Ma. *A. monophylla* is primarily distributed in India and Pakistan (Soomro et al. 2021) and is the most common and abundant species in the southern Indian Eastern Ghats (Prasad and Kumari 2019). *A. racemosa*, also found in southern India (Ranade et al. 2009), diverged from other *Atalantia* species at about 9.40 Ma. *A. buxifolia* and *A. kwangtungensis*, two species from southern China (Shi et al. 2014; Zhu et al. 2018), most recently diverged at about 0.37 Ma. *A. ceylanica* and *A. roxburghiana* diverged at around 8.26 Ma. *A. ceylanica* is primarily found in Sri Lanka and southern India (Fernando and Soysa 2014), while *A. roxburghiana* is primarily found in Southeast Asia (Esser 2021). These time divergence estimates based on the complete cp genome for *Atalantia* species provide valuable information for future studies of the genus. However, further validation of the whole genome is required to confirm the accuracy of these estimates.

Conclusions

The complete cp genomes of four *Atalantia* species were sequenced and de novo assembled in this study. The cp genomes of *Atalantia* species showed a typical quadripartite structure. *Atalantia* species have conserved cp genomes similar to *Citrus* species. *Atalantia* species prefer amino acids with A/U-terminal codons. The *rpl*22 genes were pseudogenized at the JSA junction. Certain mechanisms have been evolved to compensate for the pseudogenization of the *acc*D and *inf*A genes in the *Atalantia* chloroplast genomes. Several genes in *Atalantia* species, including *mat*K, *ndh*G, *rpl*20, and *rpl*22, have undergone positive selection during evolution. SSRs, long repeat sequences, and some highly variable loci have been found in the cp genomes of the *Atalantia* species and have the potential to be utilized as markers for population genetics and phylogenetic analyses. The complete cp genome was used to investigate the phylogenetic relationships and divergence times among *Atalantia* species. These findings can provide useful information for potential future studies of the genus *Atalantia*.

Information on Electronic Supplementary Material

Online Resource 1. Taxonomic position, species name and NCBI accession numbers of Aurantioideae species in the phylogenetic trees.
Online Resource 2. The results of mapping the reads to the assembled *Atalantia* complete cp genome sequences.
Online Resource 3. The individual chloroplast genome maps for each *Atalantia* species.
Online Resource 4. Statistics on the relative synonymous codon usage (RSCU) values of different *Atalantia* species.
Online Resource 5. Statistics of Ka/Ks values obtained from selection pressure analysis of other *Atalantia* species, using *Atalantia* roxburghiana as a reference.
Online Resource 6. Statistics regarding the number, type, region, and location of simple sequence repeats (SSRs) in diverse *Atalantia* chloroplast genomes.
Online Resource 7. Statistics regarding the type, number, region, and location of long repeats in diverse *Atalantia* chloroplast genomes.
Online Resource 8. Statistics on the nucleotide diversity (Pi) values of the *Atalantia* chloroplast genome, as well as the corresponding gene regions and their locations.
Online Resource 9. The pseudogenization of the *acc*D gene in *A. kwangtungensis*.

Online Resource 10. Information on heterozygous variants present in *Atalantia* species, including Phred-scaled quality scores, genotypes, and likelihood statistics for each genotype.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00606-023-01868-w.

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

Author contributions WBS helped in conceptualization, data curation, investigation, methodology, project administration, software, visualization, writing—review & editing, writing—original draft. WCS contributed to conceptualization, data curation, formal analysis, investigation, project administration. YZ was involved in data curation, formal analysis, validation. CS performed conceptualization, data curation, funding acquisition, project administration, supervision, writing—original draft. SW helped in data curation, funding acquisition, project administration. All authors contributed to the article and approved the submitted version.

Data availability In the GenBank database, the data supporting the findings of this study can be found at https://www.ncbi.nlm.nih.gov/, under accession numbers NC_065396–NC_065399.

Declarations

Conflicts of interest The authors declare that the work reported in this manuscript does not present any conflicts of interest for them.

References

Alves MN, Lopes SA, Raiol-Junior LL et al (2021) Resistance to 'Candidatus Liberibacter asiaticus', the Huanglongbing Associated Bacterium, in sexually and/or graft-compatible *citrus* relatives. Frontiers Pl Sci 11:617664. https://doi.org/10.3389/fpls.2020.617664

Appelhans MS, Keßler PJA, Smets E et al (2012) Age and historical biogeography of the pantropically distributed Spathelioideae (Rutaceae, Sapindales). J Biogeogr 39:1235–1250. https://doi.org/10.1111/j.1365-2699.2012.02686.x

Bacher M, Brader G, Hofer O, Greger H (1999) Oximes from seeds of *Atalantia ceylanica*. Phytochemistry 50:991–994. https://doi.org/10.1016/S0031-9422(98)00638-4

Baskar K, Sudha V, Nattudurai G et al (2018) Larvicidal and repellent activity of the essential oil from *Atalantia monophylla* on three mosquito vectors of public health importance, with limited impact on non-target zebra fish. Physiol Molec Pl Pathol 101:197–201. https://doi.org/10.1016/j.pmpp.2017.03.002

Bausher MG, Singh ND, Lee SB et al (2006) The complete chloroplast genome sequence of *Citrus sinensis* (L.) Osbeck var "Ridge Pineapple": Organization and phylogenetic relationships to other angiosperms. BMC Pl Biol 6:1–11. https://doi.org/10.1186/1471-2229-6-21

Bayer RJ, Mabberley DJ, Morton C et al (2009) A molecular phylogeny of the orange subfamily (Rutaceae: Aurantioideae) using nine cpDNA sequences. Amer J Bot 96:668–685. https://doi.org/10.3732/ajb.0800341

Beier S, Thiel T, Münch T et al (2017) MISA-web: a web server for microsatellite prediction. Bioinformatics 33:2583–2585. https://doi.org/10.1093/bioinformatics/btx198

Bilder P, Lightle S, Bainbridge G et al (2006) The structure of the carboxyltransferase component of acetyl-CoA carboxylase reveals a zinc-binding motif unique to the bacterial enzyme. Biochemistry 45:1712–1722. https://doi.org/10.1021/bi0520479

Bolger AM, 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 LA (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

Brudno M, Do CB, Cooper GM et al (2003) LAGAN and Multi-LAGAN: efficient tools for large-scale multiple alignment of genomic DNA. Genome Res 13:721–731. https://doi.org/10.1101/gr.926603

Carbonell-Caballero J, Alonso R, Ibañez V et al (2015) A phylogenetic analysis of 34 chloroplast genomes elucidates the relationships between wild and domestic species within the genus *Citrus*. Molec Biol Evol 32:2015–2035. https://doi.org/10.1093/molbev/msv082

Chat J, Jáuregui B, Petit RJ, Nadot S (2004) Reticulate evolution in kiwifruit (*Actinidia*, Actinidiaceae) identified by comparing their maternal and paternal phylogenies. Amer J Bot 91:736–747. https://doi.org/10.3732/ajb.91.5.736

Chen C, Chen H, Zhang Y et al (2020) TBtools: an integrative toolkit developed for interactive analyses of big biological data. Molec Pl 13:1194–1202. https://doi.org/10.1016/j.molp.2020.06.009

Chukaew A, Ponglimanont C, Karalai C, Tewtrakul S (2008) Potential anti-allergic acridone alkaloids from the roots of *Atalantia monophylla*. Phytochemistry 69:2616–2620. https://doi.org/10.1016/j.phytochem.2008.08.007

Curk F, Ancillo G, Garcia-Lor A et al (2014) Next generation haplotyping to decipher nuclear genomic interspecific admixture in *Citrus* species: analysis of chromosome 2. BMC Genet 15:1–19. https://doi.org/10.1186/s12863-014-0152-1

Darriba D, Taboada GL, Doallo R, Posada D (2012) JModelTest 2: more models, new heuristics and parallel computing. Nature Meth 9:772. https://doi.org/10.1038/nmeth.2109

Das AK, Swamy S (2016) Antioxidant activity and determination of bioactive compounds by GC-MS in fruit methanol extracts—a comparative analysis of three *Atalantia* species from south India. J Appl Pharm Sci 6:130–134. https://doi.org/10.7324/JAPS.2016.60220

de Araújo EF, de Queiroz LP, Machado MA (2003) What is *Citrus*? Taxonomic implications from a study of cp-DNA evolution in the tribe Citreae (Rutaceae subfamily Aurantioideae). Organisms Diversity Evol 3:55–62. https://doi.org/10.1078/1439-6092-00058

Dierckxsens N, Mardulyn P, Smits G (2017) NOVOPlasty: de novo assembly of organelle genomes from whole genome data. Nucl Acids Res 45:1–9. https://doi.org/10.1093/nar/gkw955

Dobrogojski J, Adamiec M, Luciński R (2020) The chloroplast genome: a review National Center for Biotechnology Information. Acta Physiol Pl 42:1–13. https://doi.org/10.1007/s11738-020-03089-x

Esser H-J (2021) Taxonomic notes on the Rutaceae of Thailand. Thai Forest Bull 49:27–31. https://doi.org/10.20531/TFB2021.49.1.02

Fernando CD, Soysa P (2014) Total phenolic, flavonoid contents, in-vitro antioxidant activities and hepatoprotective effect of aqueous leaf extract of *Atalantia ceylanica*. BMC Complement Altern Med 14:395. https://doi.org/10.1186/1472-6882-14-395

Fischer TC, Butzmann R (1998) *Citrus meletensis* (Rutaceae), a new species from the Pliocene of Valdarno (Italy). Pl Syst Evol 210:51–55. https://doi.org/10.1007/BF00984727

Frazer KA, Pachter L, Poliakov A et al (2004) VISTA: Computational tools for comparative genomics. Nucleic Acids Res 32:273–279. https://doi.org/10.1093/nar/gkh458

Green MR, Sambrook J (1972) Agarose gel electrophoresis. Cold Spring Harbor Protoc 124:7–19. https://doi.org/10.1101/pdb.prot100404

Gregor HJ (1989) Aspects of the fossil record and phylogeny of the family *Rutaceae (Zanthoxyleae, Toddalioideae)*. Pl Syst Evol 162:251–265. https://doi.org/10.1007/BF00936920

Greiner S, Lehwark P, Bock R (2019) OrganellarGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes. Nucl Acids Res 47:W59–W64. https://doi.org/10.1093/nar/gkz238

Grosser JW, Mourao-Fo FAA, Gmitter FG et al (1996) Allotetraploid hybrids between *citrus* and seven related genera produced by somatic hybridization. Theor Appl Genet 92:577–582. https://doi.org/10.1007/BF00224561

Guo X, Fang D, Sahu SK et al (2021) Chloranthus genome provides insights into the early diversification of angiosperms. Nat Commun 12:1–14. https://doi.org/10.1038/s41467-021-26922-4

Guo C, Luo Y, Gao L et al (2022) Phylogenomics and the flowering plant tree of life. J Integr Plant Biol 65:299–323. https://doi.org/10.1111/jipb.13415

Herrero R, Asíns MJ, Carbonell EA, 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

Huelsenbeck JP, Ronquist F (2001) MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics 17:754–755. https://doi.org/10.1093/bioinformatics/17.8.754

Jin JJ, Bin YuW, Yang JB et al (2020) GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. Genome Biol 21:1–31. https://doi.org/10.1186/s13059-020-02154-5

Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Molec Biol Evol 30:772–780. https://doi.org/10.1093/molbev/mst010

Kearse M, Moir R, Wilson A et al (2012) Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28:1647–1649. https://doi.org/10.1093/bioinformatics/bts199

Koenen EJM, Clarkson JJ, Pennington TD, Chatrou LW (2015) Recently evolved diversity and convergent radiations of rainforest mahoganies (Meliaceae) shed new light on the origins of rainforest hyperdiversity. New Phytol 207:327–339. https://doi.org/10.1111/nph.13490

Konishi T, Shinohara K, Yamada K, Sasaki Y (1996) Acetyl-CoA carboxylase in higher plants: most plants other than gramineae have both the prokaryotic and the eukaryotic forms of this enzyme. Pl Cell Physiol 37:117–122. https://doi.org/10.1093/oxfordjournals.pcp.a028920

Kurtz S, Choudhuri JV, Ohlebusch E et al (2001) REPuter: the manifold applications of repeat analysis on a genomic scale. Nucl Acids Res 29:4633–4642. https://doi.org/10.1093/nar/29.22.4633

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 H (2018) Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics 34:3094–3100. https://doi.org/10.1093/bioinformatics/bty191/4994778

Li D, Xing F (2016) Ethnobotanical study on medicinal plants used by local Hoklos people on Hainan Island, China. J Ethnopharmacol 194:358–368. https://doi.org/10.1016/j.jep.2016.07.050

Li H, Handsaker B, Wysoker A et al (2009) The sequence alignment/map format and SAMtools. Bioinformatics 25:2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Liu TJ, Zhang CY, Yan HF et al (2016) Complete plastid genome sequence of *Primula sinensis* (Primulaceae): structure comparison, sequence variation and evidence for *acc*D transfer to nucleus. PeerJ 4:e2101. https://doi.org/10.7717/peerj.2101

Liu Y, Li D, Zhang Q et al (2017) Rapid radiations of both kiwifruit hybrid lineages and their parents shed light on a two-layer mode of species diversification. New Phytol 215:877–890. https://doi.org/10.1111/nph.14607

Liu Y, Tahir Ul Qamar M, Feng JW et al (2019) Comparative analysis of miniature inverted-repeat transposable elements (MITEs) and long terminal repeat (LTR) retrotransposons in six *Citrus* species. BMC Pl Biol 19:1–16. https://doi.org/10.1186/s12870-019-1757-3

Liu X, Chang E, Liu J (2021) Comparative analysis of the complete chloroplast genomes of six white oaks with high ecological amplitude in China. J Forest Res 32:2203–2218. https://doi.org/10.1007/s11676-020-01288-3

Liu Q, Gao Y, Dong W, Zhao L (2023) Plastome evolution and phylogeny of the tribe Ruteae (Rutaceae). Ecol Evol 13:1–20. https://doi.org/10.1002/ece3.9821

Manafzadeh S, Salvo G, Conti E (2014) A tale of migrations from east to west: the Irano-Turanian floristic region as a source of Mediterranean xerophytes. J Biogeogr 41:366–379. https://doi.org/10.1111/jbi.12185

Millen RS, Olmstead RG, Adams KL et al (2001) Many parallel losses of *infA* from chloroplast DNA during angiosperm evolution with multiple independent transfers to the nucleus. Pl Cell 13:645–658. https://doi.org/10.1105/tpc.13.3.645

Nagano Y, Mimura T, Kotoda N et al (2018) Phylogenetic relationships of Aurantioideae (Rutaceae) based on RAD-Seq. Tree Genet Genomes 14:1–11. https://doi.org/10.1007/s11295-017-1223-z

Pang X, Almaz B, Qi XJ et al (2020) Bioactivity of essential oil from *Atalantia buxifolia* leaves and its major sesquiterpenes against three stored-product insects. J Essent Oil-Bearing Pl 23:38–50. https://doi.org/10.1080/0972060X.2020.1729245

Penjor T, Anai T, Nagano Y et al (2010) Phylogenetic relationships of *Citrus* and its relatives based on *rbc*L gene sequences. Tree Genet Genomes 6:931–939. https://doi.org/10.1007/s11295-010-0302-1

Penjor T, Yamamoto M, Uehara M et al (2013) Phylogenetic relationships of *Citrus* and Its relatives based on *mat*K gene sequences. PLoS ONE 8:1–13. https://doi.org/10.1371/journal.pone.0062574

Pfeil BE, Crisp MD (2008) The age and biogeography of *Citrus* and the orange subfamily (Rutaceae: Aurantioideae) in Australasia and New Caledonia. Amer J Bot 95:1621–1631. https://doi.org/10.3732/ajb.0800214

Prasad PRC, Kumari JA (2019) Structure, composition and diversity of trees within the dry evergreen reserve forest of Kondapalli (Eastern Ghats, southern India). Biodivers Res Conserv 54:23–36. https://doi.org/10.2478/biorc-2019-0009

Puttick MN (2019) MCMCtreeR: functions to prepare MCMCtree analyses and visualize posterior ages on trees. Bioinformatics 35:5321–5322. https://doi.org/10.1093/bioinformatics/btz554

Qu XJ, Moore MJ, Li DZ, Yi TS (2019) PGA: a software package for rapid, accurate, and flexible batch annotation of plastomes. Pl Meth 15:1–12. https://doi.org/10.1186/s13007-019-0435-7

Ranade SA, Nair KN, Srivastava AP, Pushpangadan P (2009) Analysis of diversity amongst widely distributed and endemic *Atalantia* (family Rutaceae) species from Western Ghats of India. Physiol Molec Biol Pl 15:211–224. https://doi.org/10.1007/s12298-009-0025-7

Ranwez V, Douzery EJP, Cambon C et al (2018) MACSE v2: toolkit for the alignment of coding sequences accounting for frameshifts and stop codons. Molec Biol Evol 35:2582–2584. https://doi.org/10.1093/molbev/msy159

Rousseau-Gueutin M, Huang X, Higginson E et al (2013) Potential functional replacement of the plastidic acetyl-CoA carboxylase subunit (*acc*D) gene by recent transfers to the nucleus in some angiosperm lineages. Plant Physiol 161:1918–1929. https://doi.org/10.1104/pp.113.214528

Rozas J, Ferrer-Mata A, Sanchez-DelBarrio JC et al (2017) DnaSP 6: DNA sequence polymorphism analysis of large data sets. Mol Biol Evol 34:3299–3302. https://doi.org/10.1093/molbev/msx248

Salvo G, Ho SYW, Rosenbaum G et al (2010) Tracing the temporal and spatial origins of island endemics in the Mediterranean region: a case study from the *Citrus* family (*Ruta* L., Rutaceae). Syst Biol 59:705–722. https://doi.org/10.1093/sysbio/syq046

Sasaki Y, Nagano Y (2004) Plant acetyl-CoA carboxylase: structure, biosynthesis, regulation, and gene manipulation for plant breeding. Biosci Biotechnol Biochem 68:1175–1184. https://doi.org/10.1271/bbb.68.1175

Sasaki Y, Hakamada K, Suama Y et al (1993) Chloroplast-encoded protein as a subunit of acetyl-CoA carboxylase in pea plant. J Biol Chem 268:25118–25123. https://doi.org/10.1016/s0021-9258(19)74577-3

Sayers EW, Beck J, Bolton EE et al (2021) Database resources of the National Center for Biotechnology Information. Nucleic Acids Res 49:D10–D17. https://doi.org/10.1093/nar/gkaa892

Schwartz T, Nylinder S, Ramadugu C et al (2015) The origin of oranges: a multi-locus phylogeny of Rutaceae subfamily aurantioideae. Syst Bot 40:1053–1062. https://doi.org/10.1600/036364415X690067

Scora RW, Duesch G, England AB (1969) Essential leaf oils in representatives of the aurantioideae (Rutaceae). Amer J Bot 56:1094–1102. https://doi.org/10.1002/j.1537-2197.1969.tb09764.x

Shi C, Hu N, Huang H et al (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 M, Guo X, Chen Y et al (2014) Isolation and characterization of 19 polymorphic microsatellite loci for *Atalantia buxifolia* (Rutaceae), a traditional medicinal plant. Conservation Genet Resources 6:857–859. https://doi.org/10.1007/s12686-014-0224-6

Shi W, Song W, Chen Z et al (2023a) Comparative chloroplast genome analyses of diverse *Phoebe* (Lauraceae) species endemic to China provide insight into their phylogeographical origin. PeerJ 11:e14573. https://doi.org/10.7717/peerj.14573

Shi W, Song W, Liu J et al (2023b) Comparative chloroplast genome analysis of *Citrus* (Rutaceae) species : insights into genomic characterization, phylogenetic relationships, and discrimination of subgenera. Sci Hortic (Amsterdam) 313:111909. https://doi.org/10.1016/j.scienta.2023.111909

Shimada H, Sugiura M (1991) Fine structural features of the chloroplast genome: comparison of the sequenced chloroplast genomes. Nucl Acids Res 19:983–995. https://doi.org/10.1093/nar/19.5.983

Song W, Chen Z, He L et al (2022a) Comparative chloroplast genome analysis of wax gourd (*Benincasa hispida*) with three benincaseae species, revealing evolutionary dynamic patterns and phylogenetic implications. Genes (Basel) 13:461. https://doi.org/10.3390/genes13030461

Song W, Chen Z, Shi W et al (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 et al (2022c) Comparative analysis the complete chloroplast genomes of nine *Musa* species: genomic features, comparative analysis, and phylogenetic implications. Frontiers Pl Sci 13:1–15. https://doi.org/10.3389/fpls.2022.832884

Soomro N, Mangi JU, Panhwer M et al (2021) Anatomical characteristics of fossil wood collected from the Manchar Formation (Miocene), Thano Bula Khan, Sindh, Pakistan. Ital Bot 11:1–8. https://doi.org/10.3897/ITALIANBOTANIST.11.60344

Su HJ, Hogenhout SA, Al-Sadi AM, Kuo CH (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 K, Liu QY, Wang A et al (2021) Comparative analysis and phylogenetic implications of plastomes of five genera in subfamily Amyridoideae (Rutaceae). Forests 12:1–14. https://doi.org/10.3390/f12030277

Swingle W, Reece P (1967) The botany of *Citrus* and its wild relatives. Citrus Ind 1:190–430

Tamura K, Stecher G, Kumar S (2021) MEGA11: molecular evolutionary genetics analysis version 11. Molec Biol Evol 38:3022–3027. https://doi.org/10.1093/molbev/msab120

Tillich M, Lehwark P, Pellizzer T et al (2017) GeSeq—versatile and accurate annotation of organelle genomes. Nucl Acids Res 45:W6–W11. https://doi.org/10.1093/nar/gkx391

Trifinopoulos J, Nguyen LT, von Haeseler A, Minh BQ (2016) W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. Nucl Acids Res 44:W232–W235. https://doi.org/10.1093/NAR/GKW256

Wang D, Zhang Y, Zhang Z et al (2010) KaKs_Calculator 2.0: A toolkit incorporating gamma-series methods and sliding window strategies. Genom Proteom Bioinform 8:77–80. https://doi.org/10.1016/S1672-0229(10)60008-3

Wang X, Xu Y, Zhang S et al (2017) Genomic analyses of primitive, wild and cultivated *Citrus* provide insights into asexual reproduction. Nat Genet 49:765–772. https://doi.org/10.1038/ng.3839

Wang T, Kuang RP, Wang XH et al (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

Wang T, Chen LL, Shu HJ et al (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). Frontiers Pl Sci 13:867659. https://doi.org/10.3389/fpls.2022.867659

Wu GA, Terol J, Ibanez V et al (2018) Genomics of the origin and evolution of *Citrus*. Nature 554:311–316. https://doi.org/10.1038/nature25447

Xia X, Yu X, Fu Q et al (2022) Comparison of chloroplast genomes of compound—leaved maples and phylogenetic inference with other Acer species. Tree Genet Genomes 18:1–12. https://doi.org/10.1007/s11295-022-01541-2

Xie S, Manchester SR, Liu K et al (2013) *Citrus linczangensis* sp. n., a leaf fossil of Rutaceae from the late Miocene of Yunnan. China Int J Pl Sci 174:1201–1207. https://doi.org/10.1086/671796

Xu X, Wang D (2021) Comparative chloroplast genomics of *Corydalis* species (Papaveraceae): evolutionary perspectives on their unusual large scale rearrangements. Frontiers Pl Sci 11:1–16. https://doi.org/10.3389/fpls.2020.600354

Yang Z (2007) PAML 4: phylogenetic analysis by maximum likelihood. Molec Biol Evol 24:1586–1591. https://doi.org/10.1093/molbev/msm088

Yang M, Zhang X, Liu G et al (2010) The complete chloroplast genome sequence of date palm (*Phoenix dactylifera* L.). PLoS ONE 5:e12762. https://doi.org/10.1371/journal.pone.0012762

Yang K, You CX, Wang CF et al (2015) Chemical composition and bioactivity of essential oil of *Atalantia guillauminii* against three species stored product insects. J Oleo Sci 64:1101–1109. https://doi.org/10.5650/jos.ess15135

Yang X, Zhou T, Su X et al (2021) Structural characterization and comparative analysis of the chloroplast genome of *Ginkgo biloba* and other gymnosperms. J Forest Res 32:765–778. https://doi.org/10.1007/s11676-019-01088-4

Zhang YM, Han LJ, Yang CW et al (2022) Comparative chloroplast genome analysis of medicinally important *Veratrum* (Melanthiaceae) in China: insights into genomic characterization and phylogenetic relationships. Pl Diversity 44:70–82. https://doi.org/10.1016/j.pld.2021.05.004

Zhao K, Li L, Quan H et al (2021) Comparative analyses of chloroplast genomes from 14 *Zanthoxylum* species: identification of variable DNA markers and phylogenetic relationships within the genus. Frontiers Pl Sci 11:1–12. https://doi.org/10.3389/fpls.2020.605793

Zhou T, Zhu H, Wang J et al (2020) Complete chloroplast genome sequence determination of *Rheum* species and comparative chloroplast genomics for the members of Rumiceae. Pl Cell Rep 39:811–824. https://doi.org/10.1007/s00299-020-02532-0

Zhou W, Harris AJ, Xiang QY (2022) Phylogenomics and biogeography of *Torreya* (Taxaceae)—integrating data from three organelle genomes, morphology, and fossils and a practical method for reducing missing data from RAD-seq. J Syst Evol 60:1241–1262. https://doi.org/10.1111/jse.12838

Zhu Z, Wang J, Sun X et al (2018) Complete plastome sequence of *Atalantia kwangtungensis* (Rutaceae): an endemic "near threatened" shrub in South China. Mitochondrial DNA Part B 3:730–731. https://doi.org/10.1080/23802359.2018.1483764