Genomes of Complete Chloroplast of Valuable Toona Species: Basic Features, Comparative Analysis and Phylogenetic Evolution in Meliaceae

李诱理[1, 2] 刘轩哲[1, 2] 蒋慧儿[1, 2] 罗贞[1, 2] 杨昊[1, 2] 陈晓阳[1, 2*], 周玮[1, 2*]

（1. 华南农业大学林学与风景园林学院 广州 510642； 2. 广东省林木种质资源创新开发与利用重点实验室 广州 510642）

Abstract: 【Objective】*Toona* is a highly economically valuable genus which has an extensive range of genetic resources. Effective identification of germplasm resources is crucial for carrying out breeding work because of the underexploitation of its resources and the high degree of similarity of its physical traits. Chloroplast genome (cp) genome is typified by evolutionary slowness and structural conservation. *Toona ciliata* (TC) and *Toona sinensis* (TS) were chosen from qualitied germplasm resources for high-throughput sequencing, plus publically available data of *Toona ciliata* var. *Pubescens* (TCv) to enable researchers to examine the properties of the various cp genomes of Toona essential species.【Method】With the tools of Cp-omics, we conducted genome mapping, repetitive sequence, codon preference, and snp/indel in 3 highly valuable species, combined with NCBI downloads of the complete genomes of *Toona* species to conduct IR/SC boundary and high marginal region analysis, maff sequence alignment, and phylogenetic research. 【Result】Consequently, the cp genome was 159139-159618 bp in length, with a typical cp tetrad structure, all with a genomic GC content of nearly 37%, and encoding a total of 125-138 functional genes. The number of long repeat sequences are not equivalent, mostly distributed in Intergenic Space, and in simple sequence repeats at 54-60bp, dominated by single and dinucleotide repeats. Codon Usage bias has A/T preference, whilst 31 codons had RSCU values greater than 1, MILC and SCUO were nearby 0.565151-0.573785 and 0.300127-0.303655 and were corroborative. In reference to *T.sinensis*, *T.ciliata* var. *pubescen* and *T.ciliata* were excavated to 113 and 115 snp, 151 and 127 indel, respectively. Within the IR/SC border, despite a certain consistency between T.ciliata and several variants, there are still more variants due to the expansion and contraction of single copy regions. The variant intervals were focused on non-coding regions, mutation hotspots such as *petB-petD, trnC-petN-psbM, trnT-trnF-ndhJ, ndhC-trnM, trnS-trnR, trnC-petN-trnD* and *ndhF-rpl32-trnL* were found, nucleotide diversity analysis screened for highly variable interspecies intervals such as c*csA-ndhD*, 7 SNP loci were selected in variant fragments as markers that could be applied for species identification, the phylogenetic tree of the *Toona* showed close affinities. 【Conclusion】This work is of great significance for the development and identification of germplasm resources, breeding of superior species, systematic classification and evolutionary relationships of *Toona*.

Keywords: *Toona*; Chloroplast Genome; SNP; Indel; Comparative analysis; Phylogenetic analysis

1. Introduction

The *Toona* genus is an economically valuable plant of the Meliaceae family, represented mainly by TS, TC and TCv. They are both precious fast-growing species, with TS a multifunctional species that combines timber, food, medicinal and landscaping functions, while TC and TCv have considerable economic value in the timber industry and are internationally renowned as "Chinese Mahogany" for their beautiful wood grain[i],[ii],[iii],[iv]. The phenotypic characteristics between TC and its variety TCv are extremely similar and hard to distinguish, and their phenotypic characteristics are easily altered by the environment and are unstable[v],[vi]. Besides, pests and diseases has plagued TC forest trees, which have a great impact on their economic value. Due to environmental changes, low natural regeneration rate and over-exploitation, its population is gradually decreasing. Under the Red Book of Chinese Plants, the species belongs to the endangered category and is classified as a Class II endemic Chinese wild plant under national priority protection[vii],[viii]. Although Toona germplasm resources are abundant and widely distributed, with many local species and wild resources, the exploitation of its resources is scarce at this stage, so carrying out effective identification of germplasm resources is crucial to carry out breeding work[ix],[x],[xi],[xii].

Currently, the taxonomy of *Toona* is focused on morphological aspects[xiii],[xiv]. Genetic sequencing and the widespread systematic availability together with the construction of phylogenetic trees may potentially enable the exploration of *Toona* in further molecular insight besides morphological classification[xv]. Their analysis based on molecular laminate level by high-throughput sequencing technology is of high interest for the conservation and identification of endangered species. Wang[xvi] mined the evolutionary points of distinction between *T.sinensis* and *T.ciliata* based on the whole genome level. Li[xvii] aligned TS and TC based on the mitochondrial genome and obtained that they have high similarity in the mitochondrial genome. However, the gene-based differentiation analysis of *T.calata* var. has not been reported.

As of the three DNA genomes in plants with uniparental inheritance (maternal), the cp genome is haploid, has a highly conserved genomic structure with moderate sequence variation, convenient sequencing and analyzing, for significantly reduced gene flow compared to the biparentally inherited nuclear genes[xviii]. Its relatively homogeneous genome structure and genome sequence integrity have made it widely accepted in biological research[xix]. Cp genome provides more suitable material for species identification, plant evolution, and development analysis[xx],[xxi]. In addition, they can reveal remarkable divergences in sequence and structural variation within species, especially at the IR boundaries of chloroplast genomes, where there is some amplification and cross-species variation, which is valuable for investigating the evolution of green plants and revealing relationships[xxii],[xxiii].

Previously, several studies were reporting the cp genome of associated species. Liu et al.[xxiv],[xxv] obtained the basic characteristics of the cp genome of the *Toona* plant by high-throughput sequencing, and established phylogenetic trees to show that the 3 species were closely related. However, the works were limited to the fundamental characterization or systematic evolutionary analysis of individual species, while the sequence alignment between interspecies to explore the distinctive gene fragments to provide an effective approach for species identification has not been studied yet[xxvi],[xxvii]. Furthermore, this series of examinations will provide theoretical support for the germplasm source protection or identification of *Toona*[xxviii].

Have assembled and annotated the cp complete genome sequences of Ts and TC, both high-quality species from China. Moreover, we downloaded the published TCv. Aiming to comprehensively explore the basic features of the cp genomes of these three representative plants within *Toona*, we also filtered out the interspecific hyper-variable contigs by sequence matching, which can be applied in the identification of woody plants. Eventually, the systematic relationships of *Toona* relatives are discussed. In this study, the future development of specific DNA barcodes and the conservation of germplasm resources will be contributed to future in-depth work.

2. Results

2.1. Genome Features

A typical tetrameric double-stranded circular structure of the cp genomes of the 3 *Toona* plants consisted of 2 single-copy sequences in a reverse-duplicated pair and divided into four parts (Figure 1). Cp genomes of TS, TC, and TCv with total lengths of 156139 bp, 159618 bp, and 160737 bp, respectively, where the major differences were in the Large single copy (LSC) region, while the small single copy (SSC) and two Inverted (IR) regions differed by less than 100 bp (Table 1).

Table 1. Genomic characteristics of the cp genome of 3 *Toona* plants.

	T.ciliata(OK572964)	*T.sinensis*(OK572965)	*T.ciliata.*var(MZ926838)
Total	159618	159139	160737
Large	87101	86707	88137

single copy (LSC, bp)			
Inverted repeat (IR, bp)	27059	27025	26982
Small single copy (SSC, bp)	18399	18382	18636
GC%	37.89	37.9	37.49
Total genes	125	125	138
Protein coding genes	88	88	90
tRNA	29	29	40
rRNA	8	8	8
Genes of unknown function		ycf1* ycf2* ycf15*	ycf3**,ycf4,ycf2(x2),ycf15(x2),ycf1

Note: * represents a repeated gene (indicating that there is more than one gene on the chloroplast).

Within the GC content, TC and TS contained 37.89% and 37.9%, respectively, while TCv had 37.49%. 88 protein-coding genes (PCG), 29 tRNA, and 8 rRNA are annotated in TC and TS, whereas TCv had 88 PCG, 40 tRNA, and 8 rRNA. In terms of functional genes, TC and TS were significantly related, yet TCv differed, most notably by the presence of two more types of hypothetical genes, ycf3, and ycf4 (Annotated functional genes are detailed in Appendices A).

2.2. Codon Usage bias (CUB)

The codon usage (CU) of the 3 cp genomes were all coded for 10 amino acids by the common 64 codons (Figure 2). Leucine (Leu) (10.616%-10.67%) had the highest ratio and cysteine (Cys) (1.149%-1.160%) the least, in agreement with the majority of the cp genome amino acid coding in green plants[xxix]. Statistics on the frequency of codon usage in the coding region (CDS) of the cp genome of the 3 *Toona* plants revealed that 27,844 codons together encode amino acids in TS, while TC consists of 27,834 and TCv has 26,450 (Appendices B). Leu was the furthest frequent amino acid in the codons encoded 2966 times in TS and 2970 and 2808 times in TC and TCv, respectively; Cys was the least frequently encoded amino acid, encoding only 320,232 and 305 times in TS, TC, and TCv, respectively. These results were coherent with those in Figure 2.

Table 2. SCUO and MILC profiling of 3 *Toona* plants.

Species	MILC	SCUO	r	p
T.ciliata.var(MZ926838)	0.573785	0.303655	0.475928	0.00000278
T.ciliata(OK572964)	0.565151	0.300127	0.4485537	0.00001318
T.sinensis(OK572965)	0.567064	0.301368	0.4674666	0.000004998

In the RSCU codon base composition of the protein-coding gene sequence (CDS) of the 3 *Toona*, it was concluded that the CDS sequences of the genes all have 64 codons, 61 of which encode 20 amino acids and the other 3 codons were stop codons (Stp). There were extreme similarities in RSCU among the three species, and the RSCU values of each codon did not differ much. Among them, 31 codons had RSCU values greater than 1, 31 codons had RSCU values less than 1, and 2 codons had RSCU values of 1 (Figure 3, all specific values for codon preference analysis

can be detailed in Appendices B).

In order to gain insight into the relationship between CUB and gene expression, SCUO and MILC values were calculated. The mean SCUO value for all 3 species was approximately 0.30, which is lower than the median of 0.5, reflecting that most of the genes in their cp genomes have low CUB and very similar CU; the MILC values associated with gene expression levels were 0.565151-0.573785, embodying an adequate expression level of the 3 cp genes. The connection between gene expression levels and CUB was additionally examined. We noticed a statistically positive connection of SCUO and MILC values (Table 2), further proposing that gene expression levels might be impacted by CUB.

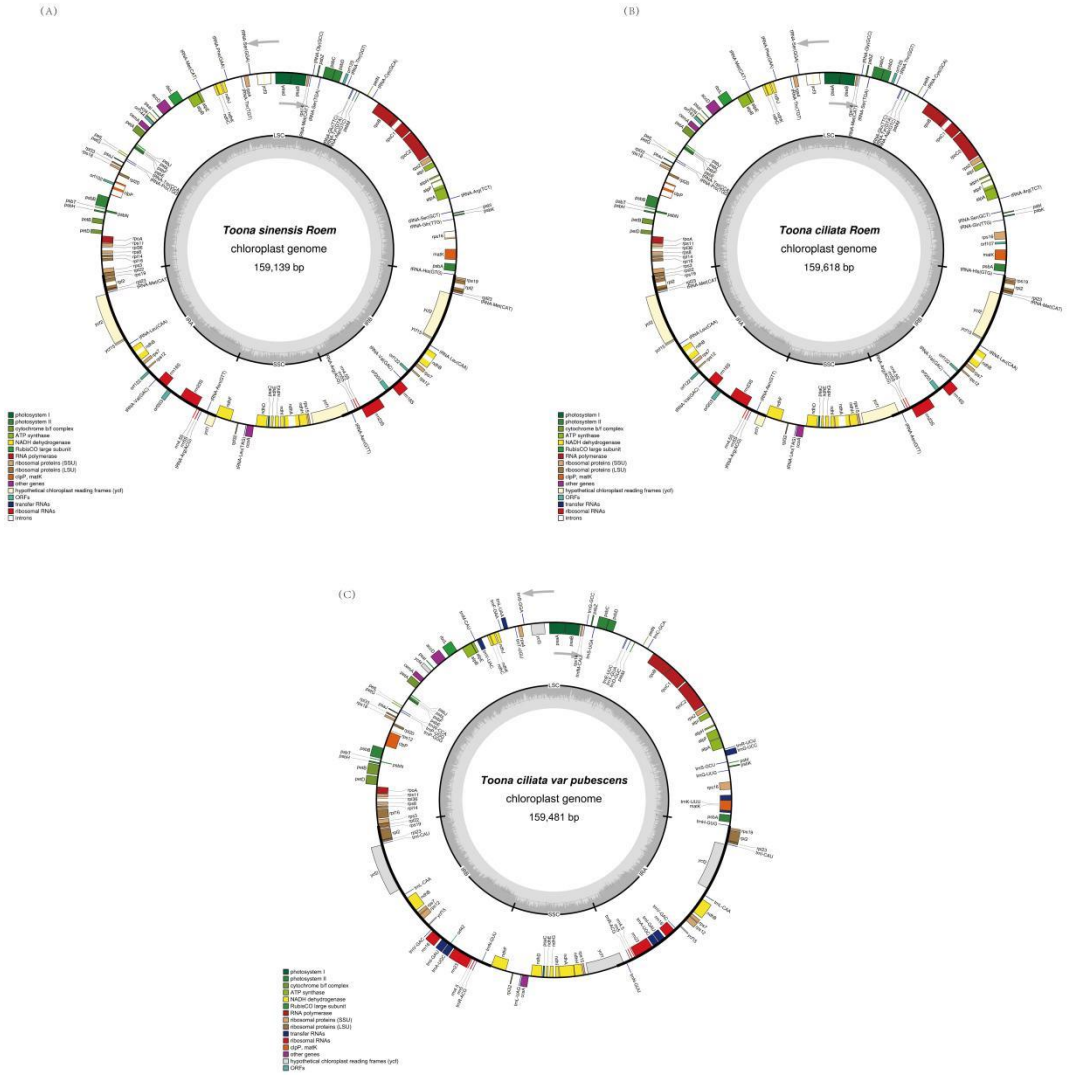

Figure 1. Maps of the cp genomes of the 3 *Toona*. (A)TS,(B)TC,(C)TCv.

The outside of the circles means the genes are reverse transcribed and the inside of the circles are in the clockwise direction. There are two IR regions denoted by the thick black lines on the external circle. And the dark grey graphic of the internal nucleus stands for GC content.

2.3. Repeated Sequences and SSRs Polymorphism

2. 3. 1. Repetitive Sequences

Based on the parameters set for the long repeat analysis of the 3 *Toona* cp genomes, the sequences with 30-60bp length in the short copy region are distributed as in Figure 4. In the 3 genomes, 2 palindromic (P) and 2 forward (F)

repeats were identified, with 58bp and 48bp p-repeats and 36bp and 30bp F-repeats, however, TC had an extra 30bp P-repeat and 41bp F-repeat compared to TS and TCv.

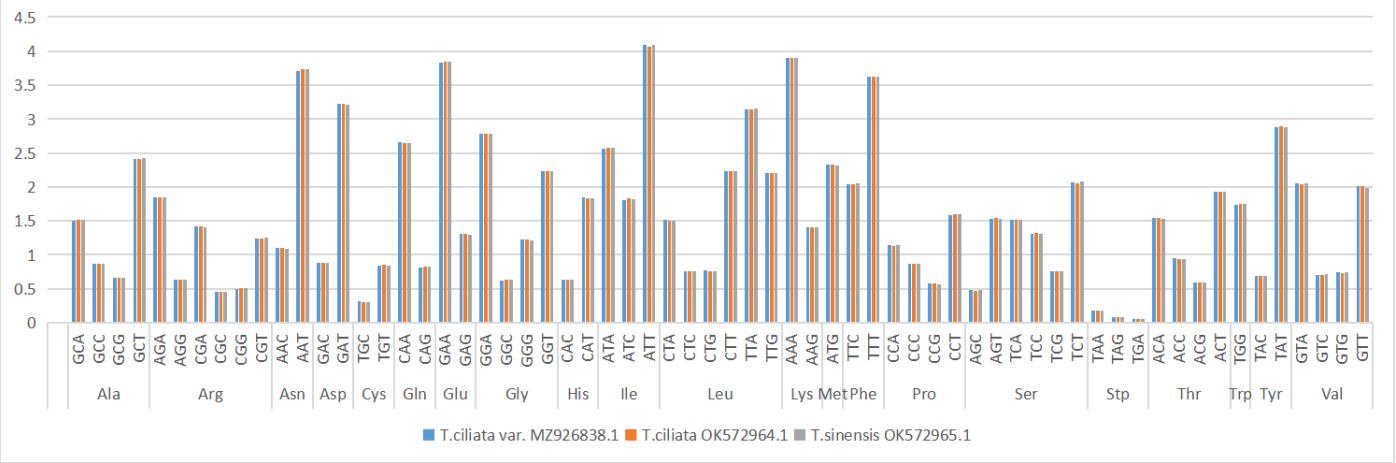

Figure 2. Histograms of codon usage frequencies (%) in the chloroplast genomes of 3 *Toona* species.

The values in the below statistics table are the frequencies (%) of the codons encoded in the horizontal coordinates of the corresponding bars.

Note: Stp indicates termination codon.

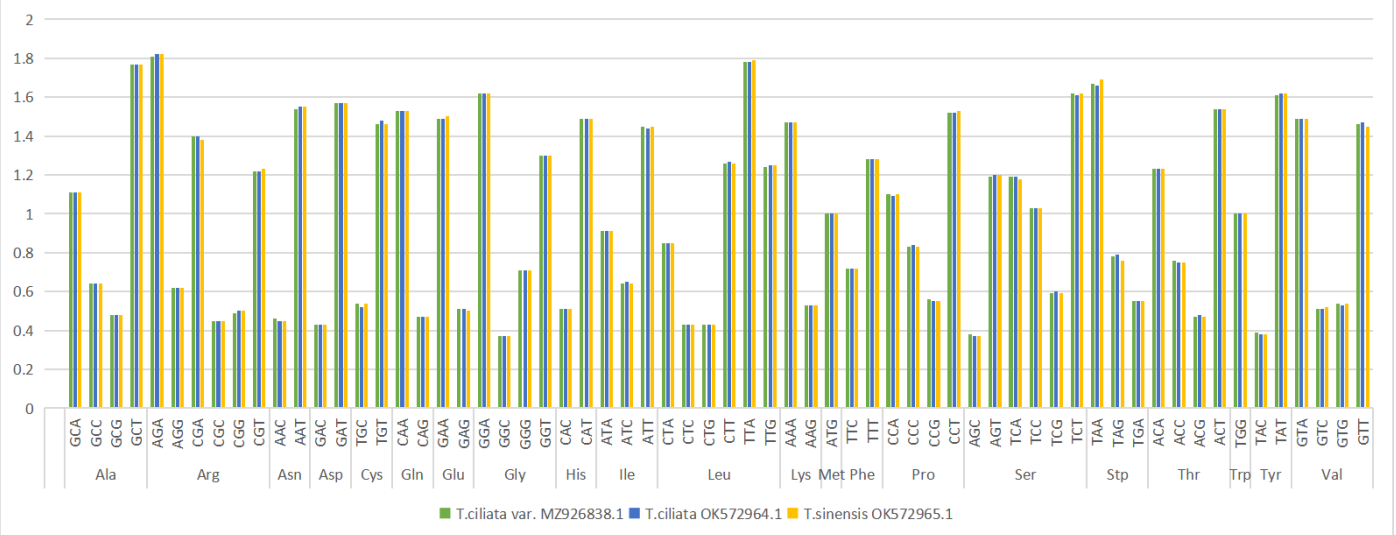

Figure 3. Histograms of RSCU in the chloroplast genomes of 3 *Toona* species.

The values in the statistical tables below are the RSCU values corresponding to the horizontal coordinates of the bars.

Note: Stp indicates termination codon.

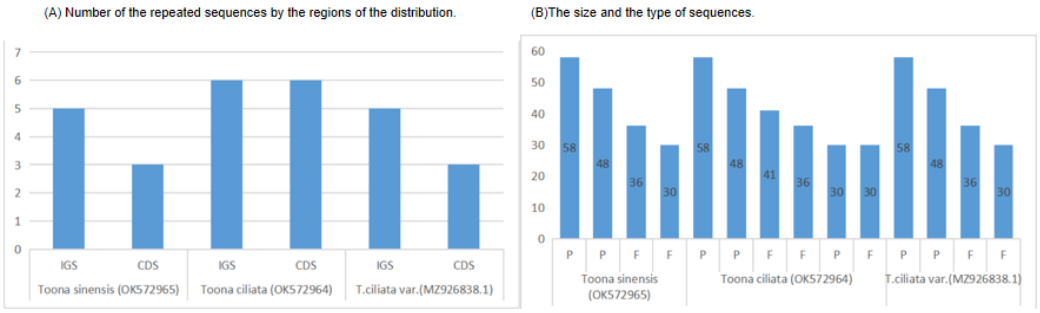

Figure 4. Statistics of long repeated sequences in 3 *Toona*. (A) Number of the repeated sequences by the regions of the distribution,(B)The size and the type of sequences.

Note: IGS represents Intergenic Space; CDS represents Coding Region. F means forward(direct) match; P means palindromic match.

Among these repeats, TS and TCv have 5 start positions (62.5%) in the Intergenic spacers (IGS) and 3 (37.5%) in the Coding Region (CDS), while 3 genes in the CDS were *ycf2*(2), *ndhA*. In contrast, TC had an additional 1 start position in the IGS and 3 in the CDS, resulting in 3 additional CDS genes, *petD*(2), *ndhF* (Figure 4A, specific values of repeat sequences are detailed in Appendices C).

2. 3. 2. *SSRs*

To investigate the distribution and divergence of SSRs among *Toona* species, we examined SSR in cp genomes. In 3 cp genomes, single nucleotides composed of A/T and dinucleotides consisting of AG/CT were detected to dominate the repeat sequence types of SSRs, separately. Mono- and di-nucleotides were AT enriched, with 60, 54, and 56 A/T repeats for TS, TC, and TCv, respectively, along with one AG/CT repeat for each (Table 3). It is indicated that strong AT levels are present in the SSR, which is consistent with the results on the GC content of the genome and the degree of codon usage.

Table 3. Types and numbers of SSRs in the Cp genome.

SSR Type	Repeat Unit	*T.sinensis*	*T.ciliata*	*T.ciliata* var.
Mono	A/T	60	54	56
Di	AG/CT	1	1	1

2.4. IR/SC boundary contraction and expansion

On six species of the genus *Toona*, we conducted IR boundary analysis, and we discovered that, with the exception of one variety of TC, *Toona fargesii* (NC-037251.1), which occured obviously large differences and the genes were altered substantially (production of *rrn16*, *trnI*, and ycf1 genes, deletion of *rps3*, *rpl22, rps19*, *rpl2* and *trnH* genes, same for the *rps3, rpl22, rps19, ycf1, ndhF, rpl2, rps19, trnH,* and *PsbA* genes distributed and fairly conserved across the borders of the other five species. On the five slightly conserved cp IR/SC boundaries, *rpl22* straddles the LSC/IRb; the ndhF of *T.ciliata* var. *henryi*, *T.ciliata* var. *yunnanensis* and *T.ciliata* lies entirely on the SSC, 24 bp from the IRb/SSC boundary. on the other hand, *T.ciliata* var.*pubescens* and *T.sinensis* had a 7bp and 9bp protrusion in the IRb region; the SSC/IRa and IRa/SSC boundaries were distributed with the *ycf1* and *trnH* genes, respectively, and the length of the *trnH* varied among the 5 species. Overall, high similarity was maintained in *T.ciliata* and its 3 varieties, *T.ciliata* var. *Henryi, T.ciliata* var. *yunnanensis* and *T.ciliata* var. *pubescens*.

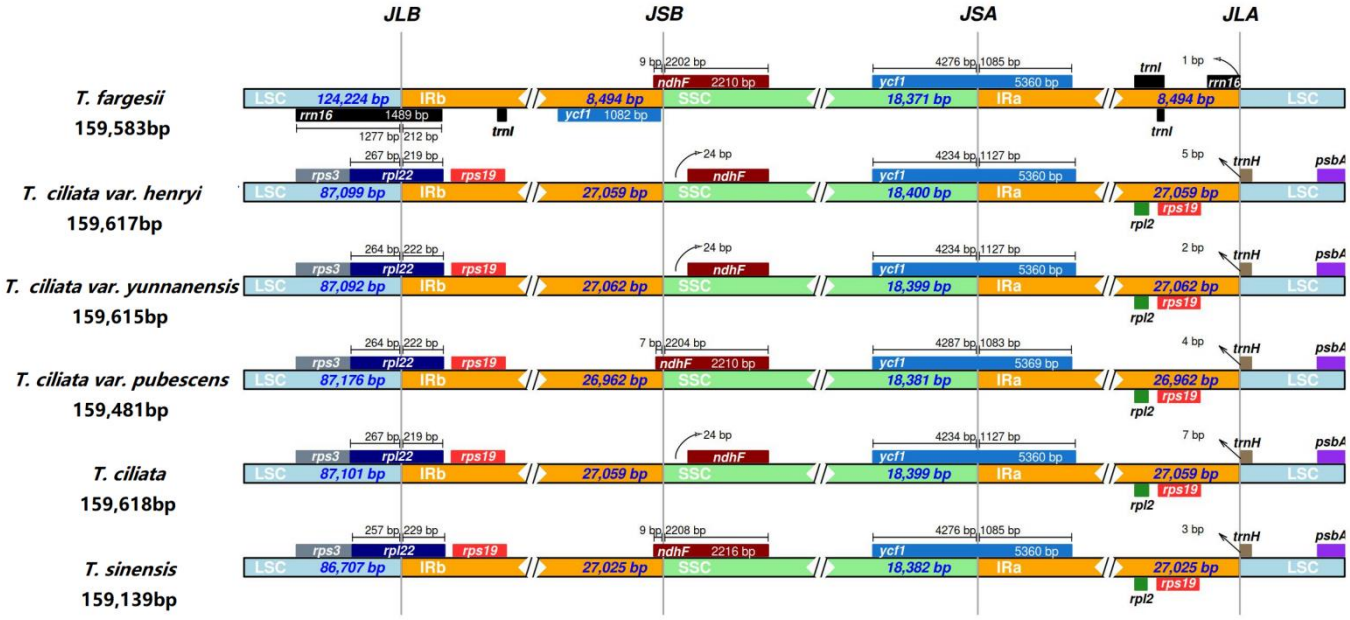

Figure 5. Comparison of expansion and contraction of IR boundaries.

The genes on the positive strand are shown to be transcribed above the corresponding tracks from right to left, whereas the negative strand genes are depicted below the tracks from left to right. The arrows denote the distance from the side of the gene to the LSC/IR boundary.

2.5. *Sequence variation analysis*

With the cp genome of *T.sinensis* as the reference sequence, we matched it with the other five species. In these plants, the IR region of the cp genome was more well-conserved than the LSC and SSC regions, the CDS was more conservative than the No-Coding Sequence (CNS), and the degree of IGS variation exceeded the level of gene region (GS) variation. IGS with *petB-petD*, *trnC-petN-psbM*, *trnT-trnF-ndhJ, ndhC-trnM, trnS-trnR, trnC-petN-trnD,* and *ndhF-rpl32-trnL* were the high-variable areas; the protein-coding gene region was conserved, whereby only *ycf1* and *ycf3* were relatively significant; the four rRNA genes (*rrn16*, *rrn5*, *rrn16,* and *rrn23*) were the best conserved (Figure 5).

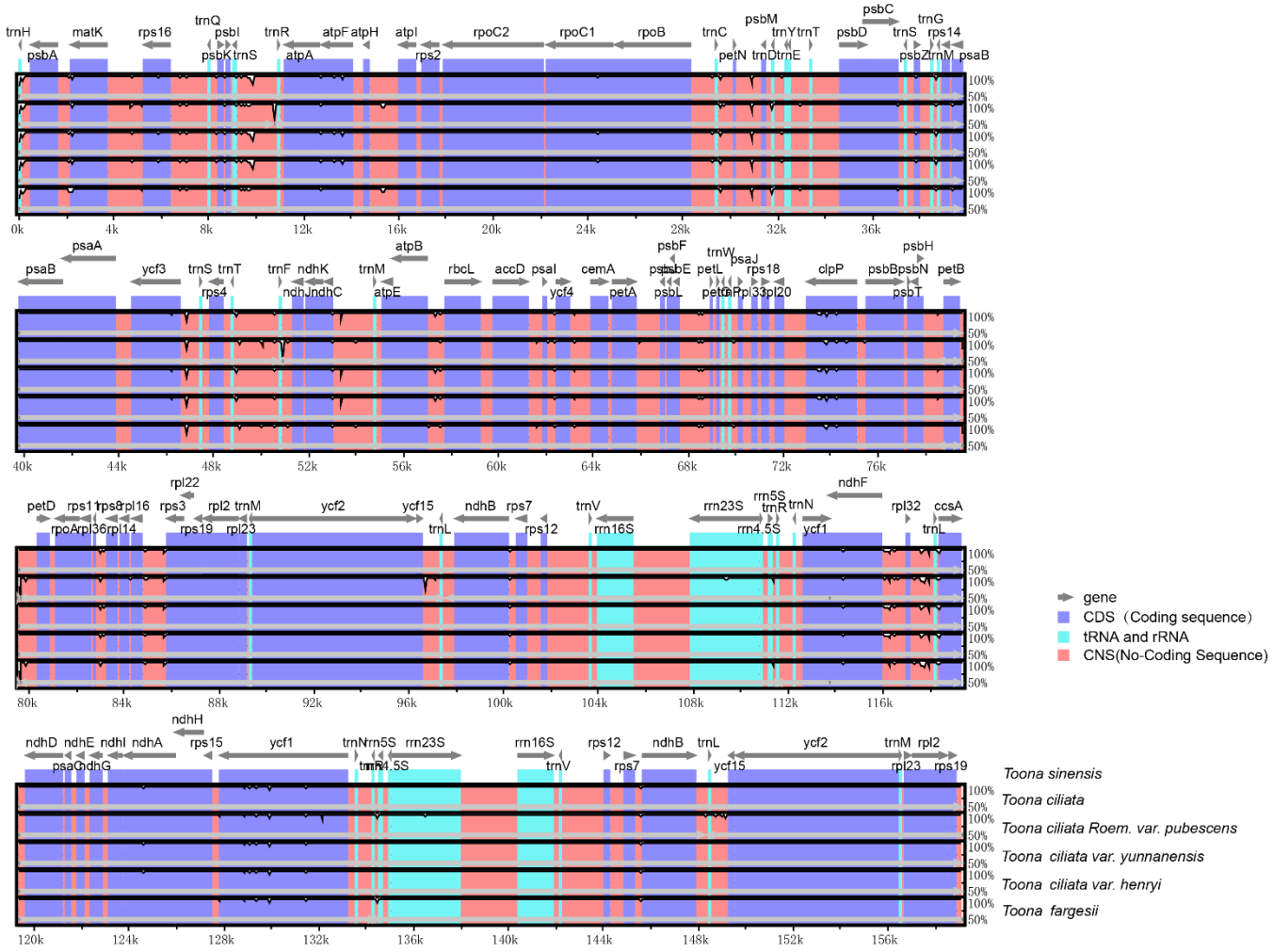

Figure 6. Sequence reassimilation among cp genomes of the *Toona*.

The vertical scale, ranging from 50% to 100%, indicates the percentage of identity calculated in sliding windows. The horizontal axis indicates the coordinates within the chloroplast genome. Different colors correspond to the types of genome regions. Blue: Regions coding for proteins; Pink: Non-coding Regions; Light blue: Regions coding for tRNAs and rRNA.

Additionally, we determined the nucleotide polymorphism (Pi) values of the genomic sequences to examine the degree of sequence variation across *Toona* cp genomes of which 10 pi is greater than or equal to 0.015 (Appendices E and Figure 6). Similar to the earlier findings, single copy (LSC and SSC) areas were more variable than IR regions because they were less conserved than other regions. In the IGS area, certain sections displayed diversity, including *ndhF-rpl32-trnL*, *matK-rps16*, *trnH-psbA*, *ycf2-trnL*, *trnL-ycf2*, and *ccsA-ndhD*, where the greatest Pi value across chloroplast genomes was 0.02133. These fragments can be explored to find prospective molecular markers that can be used to study phylogeny and genetic variation.

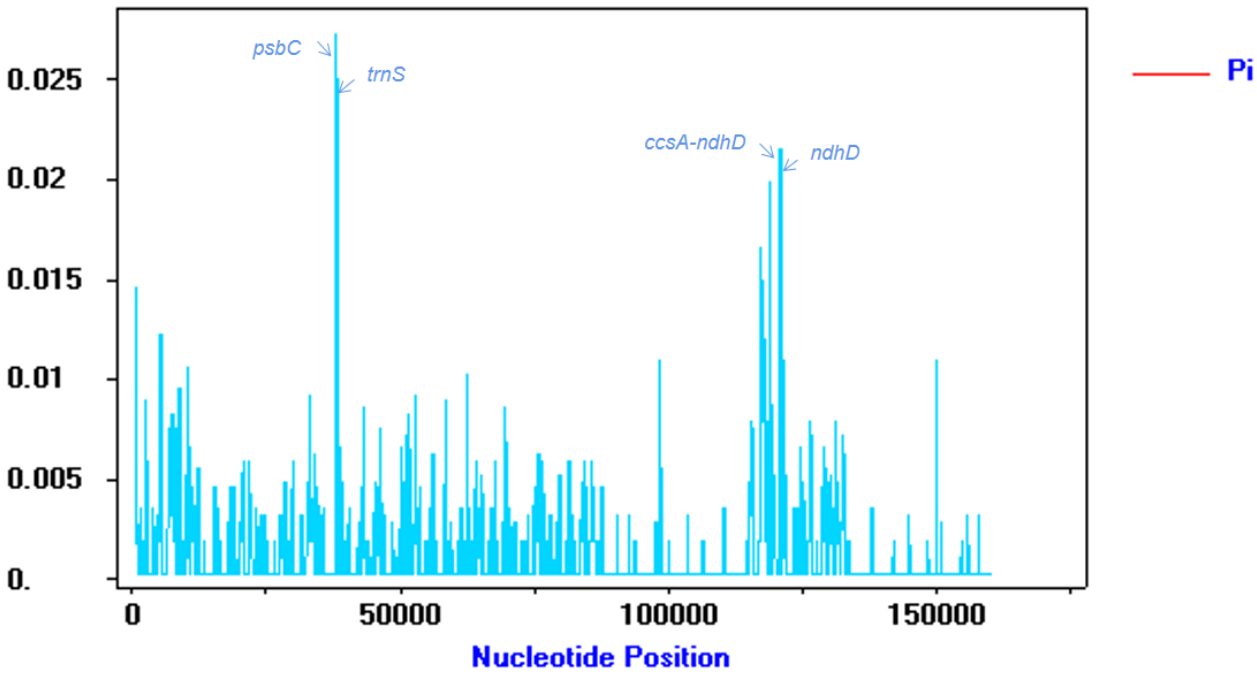

Figure 7. Genome-wide nucleotide polymorphism analysis of cp in *Toona*. x-axis = midpoint of a window located; y-axis = pi level per window.

Note: Only the gene information of the region with pi > 0.02 is indicated.

2.6. *Intraspecific analyses*

The SNP/Indel sites of the TCv and TC chloroplast genomes were evaluated with the cp genome of TS as a starting point. TCv had 113 SNPs, of which 66 were synonymous (S) type and 47 were nonsynonymous (N); TC had 115 SNP loci, of which 60 were synonymous (S) type and 55 were nonsynonymous (N). Except for 2 and 4 SNPs in the IR area and 30 and 32 SNPs on the IR/SC boundary, all of the SNPs in TCv and TC were distributed in the SC region (refer to Table 4 for extra details on SNP locus distribution and data type statistics).

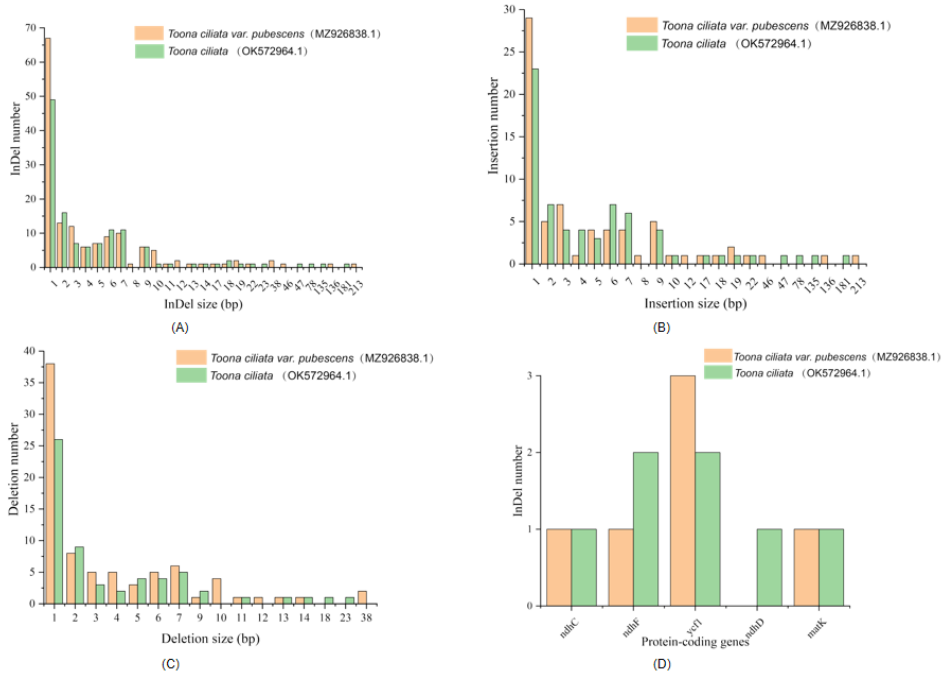

Figure 8.Base insertion and deletion statistics of 3 cp genomes. Took TS (OK572965.1) as the standard sequence and enumerated the base insertions or deletions in the cp genomes of TC (OK572964.1) and TCv (MZ926838.1). (A) Indel Information, (B) Insertion Information, (C)Deletion Image, (D) Protein-coding genes associated with Indel.

In the indel site counts for TCv and TC (Figure 8), there were 151 (70 Insertion and 81 Deletion) and 127 indels (67 Insertion and 60 Deletion), respectively. Likewise the PCG region's indel counts showed that the *ycf1* gene was a hotspot that caused PCG indel variation, with 6 for TCv and 7 for TC, specifically, with the *ndhC*, *ndhF*, *ycf1*, and *matK* genes (plus one additional *ndhD* gene for TC) (Indel data and proteomic coding regions are detailed in Appendices E). All the SNPs and indels will be potential genetic markers to facilitate species identification.

2.7. Validation of second-generation genome-wide sequencing

We evaluated several second-generation genome sequencing data from their variant fragments or loci to establish the generalisability, i.e. to confirm the applicability of the variant fragments or loci screened in the three Toona species for molecular marker creation. The entire cp genomes of four people each from TS, TC and TCv were examined by mafft v7.490 [xxx] to build a formative tree and grouping arrangement, as portrayed in "Operating Instructions" file in Appendices F. In view of the consequences of the arrangement, a self-composed Python program was utilized to identify three intraspecific indistinguishable yet interspecific conflicting loci.

According to the ML construction tree (Figure 9) and sequence comparative analysis ("Alignment" file) of the 12 cp genomes, these 135 differential loci were detected amongst the 3 species ("diff.snp" file for statistical results), these divergent loci efficiently distinguished 3 species classes among 12 individuals, with considerable variances in the interspecific and consistency within intraspecific. While the polymorphic areas (pi 0.01) were enumerated in Table 5, a total of seven snp loci were filtered, and a simple primer design revealed that all can separate 3 *Toona* species in the high-mutational regions. Above of them might serve as alternative variant loci for the creation of molecular markers.

2.8. Maximum-likelihood (ML) Phylogeny

To clarify the affinities of plants within the *Toona* and their developmental position in the Meliaceae, the ML developmental tree will be constructed by choosing 22 *Toona* plants and 11 other Meliaceae species, with *Citrus limon* (MTV80608.1) as the outgroup.

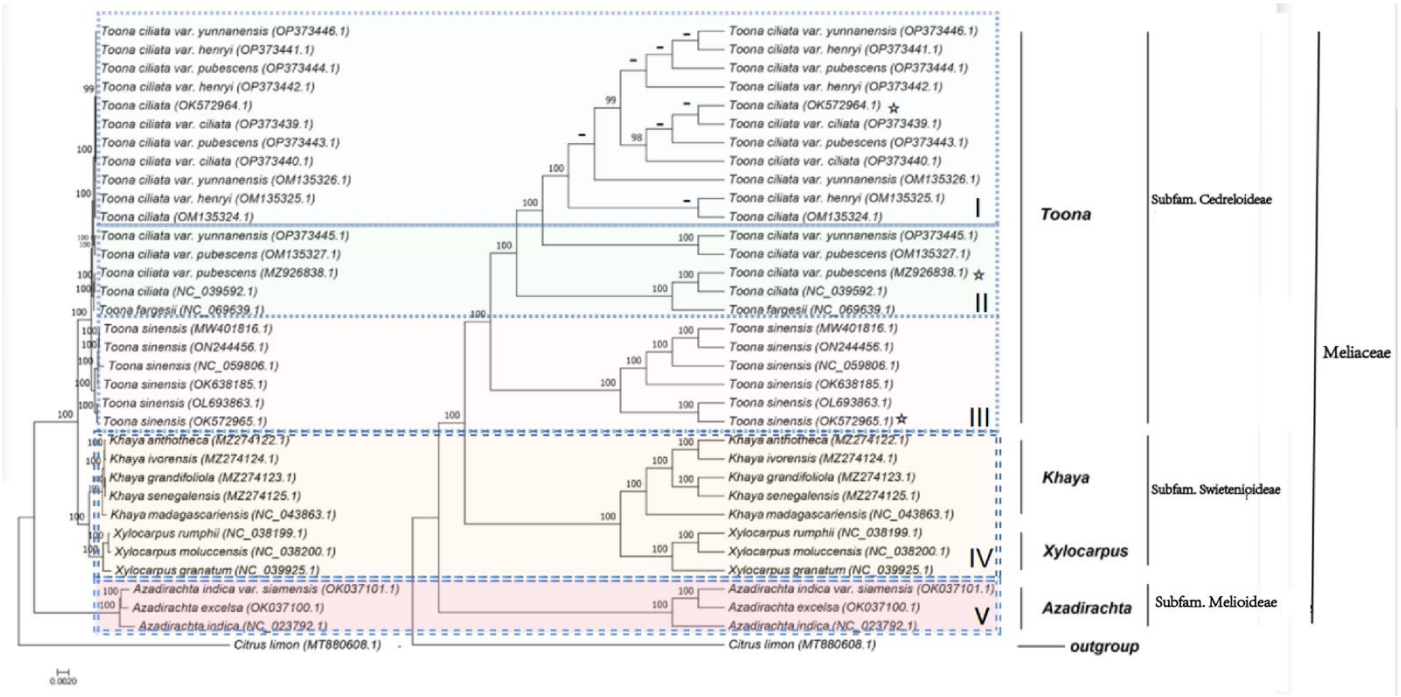

Figure 9. ML tree based on 34 genome-wide cp sequences.

33 species are selected representative plants from 3 subgroups in the Meliaceae family, and *Citrus limon* is the outgroup, all the registration numbers are displayed in the figure. A support value based on 1000 replications is indicated on each node.

Note: "-" means the bootstrap value (BP) is less than 50%. Class I and II are categorized into a branch, the main division of which is based on whether the BP (The value over 50%) is valid.

In accordance to the results (Figure 9), 33 species together form a topologically developed tree with 41 nodes, 34 of which have a bootstrap value (bp) of 98%. Simultaneously, in the genus *Toona*, six *T.sinensis* clustered into one large branch (class III), all with 100% support, and the remaining 16 *T.ciliata* and its varieties clustered into another branch (with very low-value bp values in the Class I and 100% bp values in all Class II). *T.fargesi*i (NC-069639.1) was proved to form a sister single branch with *T.ciliata* var. *yunnanensis* and a branch where *T.ciliata* clustered in the Class II section (100% support). the branches of Toona, especially I and II, are very short in length, with little variation between apparent species and no evolutionary support. It is notable that the three subgroups of the Meliaceae: the *Toona* group (I+II+III) belonging to the subfam. Cedreloideae are more related to the sister groups of *Khaya* and *Xylocarpus* of the subfam. Swietenioideae (IV) than to the *Azadirachta* group of the subfam.Meliaceae (Class V).

3. Discussion

3.1. cp features and structural differences in the cp genomes

A significant degree of structural similarity exists across the three species, according to the first structural analysis of the complete genomes of TC, TS, and TCv. The gene map is a typical tetrad structure, which is typical of most higher green plants[xxxi]. Among the functional genes encoded, TS and TC are consistently expressed, whereas TCv has 13 extra tRNAs as well as the ycf3 and ycf4 genes. tRNA is essential in protein synthesis in chloroplasts, which is linked to the evolutionary features of the genome of TCv[xxxii]; it is necessary to confirm the relationship between the Ycf3 and Ycf4 genes and the photosystem I (PSI) of TCV through gene function since these genes encode the Ycf3 and Ycf4 proteins, which are engaged in photosynthesis and are a regulatory component of photosynthesis in plants[xxxiii],[xxxiv],[xxxv].

The CUB in the Cp genome has been extensively exploited to investigate the evolutionary characteristics of numerous plant species[xxxvi]. Codons are crucial core components connecting amino acids, proteins, and genetic material in living organisms[xxxvii]. For the investigation of protein expression and its accompanying functions, the analysis of codon preferred usage bias yields trustworthy data. In this work, preferred codons for TC, TS, and TCv were screened using the codon usage frequency and RSCU statistics, and it was discovered that CUB consistently had an A/T preference and a propensity to encode Leu among the three plants; SCUO and MILC statistics can further show the high level of genes in the cp and how they are influenced by CUB[xxxviii],[xxxix], transgenic studies that were conducted by choosing the optimum codons to increase the expression efficiency of e external genes, which previously provides information for genetically improving transgenes and heritability predictions based on codon optimization.

3.2. Cp SSRs

SSRs have high species-level polymorphism, making them ideal molecular markers for population genetic and evolutionary studies. A number of studies have been conducted on three Toona to analyse SSRs based on population traits and genetic diversity. For example, Liu et al. carried out segregation and epistasis analyses of SSRs for the endangered TCv population and analysed its central and peripheral populations by polymorphic SSR primer design, concluding that population genetic distance was not linked to geographical distance[xl],[xli]. In addition, cp SSR markers are often used for the analysis of cytogenetic patterns, e.g. Wills DM et al. used markers from cpSSR to study the organelle genetic patterns of *Helianthus annuus* in some hybrids[xlii]. We identified 60, 54 and 56 SSRs in the cp genomes of TS, TC and TCv, respectively, which are highly polymorphic and have important implications for the subsequent development of

molecular markers for cp SSRs.

3.3. Potential molecular markers based on cp DNA

While Xiao et al[xliii]. utilized mtDNA and nrDNA for ITS markers to determine the level of divergence of TC in phylogeny, McPherson et al[xliv]. employed silicio cp DNA for assembly and SNP detection and exploited variation in CP for molecular ecology of TC. Snps and indels were discovered in TC and TCv (TS was used as the control) cp DNA, also validated by 4 individual screens on 3 different plants in this study. The snp loci selected between individuals in the high-variance interval were successfully differentiated between the three species by primer validation and the *matK* gene is frequently a specific sequence for the development of markers and barcodes for cpDNA. However, in *Crepidiastrum denticulatum*[xlv] and *Acacia* [xlvi], snp-based molecular markers inferred from the *matK* and *rbcL* regions were successfully developed and tested; snp and cp indel markers were created for the matK genes in the PCG region of the cp DNA to help identify between the 3 *Carduss* plants[xlvii] and *Cruciata* (Gentianaceae)[xlviii]. The snps/Indels of TC and TCv might potentially be useful genetic markers for species identification based on the above-mentioned accomplishments.

3.4. *Variation Alignment*

Fluctuations in the size of the cp genome and the structural stability and evolution are significantly influenced by the expansion and contraction of the boundaries of the IR region[xlix]. Although the IR region is highly conserved and structural changes are not significant, the SC regions of *Toona* plants is predominantly contracted and expanded, producing boundary variation. the TC variant, *T.fargesii* (NC-037251.1), undergoes significant variation, perhaps for reasons such as *Adiantum malesianum* (Pteridaceae), in contrast to other comparative genomics analysis of six *Adiatum* species, due to a 6876 bp long rpoB-trnD-GUC intergenic spacer (IGS), causing topological tension at the LSC/IRb boundary to the point of large alterations[l], but also requires multiple samples for sequencing and calibration model analysis, while TS, TC, TCv and the other two variants of TC are still conserved within each other.

The high-variance regions of the cp genome were targeted for screening suitable fragments for molecular marker and DNA barcode development. *petB-petD, trnC-petN-psbM, trnT-trnF-ndhJ, ndhC-trnM, trnS-trnR, trnC-petN-trnD* were detected in mvista. The sliding window analysis resulted in the identification of 10 highly variable regions with a Pi greater than 0.015, which can be considered as potential molecular markers for *Pyrrosiae Folium* and other plants[li]. The presence of seven valid snp marker loci could be tentatively demonstrated, the validity of these fragments for plant identification, however, remains to be further verified.

3.5. Phylogenetic location and classification of the Toona species

By the construction of the ML tree of the cp complete genome, *Toona* is classified in the subfam. Cedreloideae of the Meliaceae into a Class III where the TS species are clustered, and a Class I+II where TC and all the varieties form, which is consistent with the division in FRPS 43. The nodes of Class III all have 100% support, and the species variation and evolution of TS is low and relatively stable, so it can be clearly classified; the Class II can obtain definite positional relationships, but in the branch there is a very poor support rate due to information limitations etc. In the Class I, the species relationships of the re-clade cannot be determined for the time being, mainly attributed to the phenomenon of mixing among the lineages of TC, and the uniparental inheritance of the cp genome, which is unable to explain issues such as hybrid variants and gradual gene infiltration, etc[lii]. More cp genomes of TC and its variants should be available for future *Toona* cp phylogenetic analyses. In contrast, the overall evolutionary tree could illustrate that *Toona* plant chloroplast genomes are evolutionarily conserved, with minimal evolution and narrow genetic diversity.

The present phylogenetic analysis may offer the possibility that chloroplast genomes may have utility for *Toona* phylogeny and species identification going forward.

4. Conclusion

We carried out a chloroplast genome analysis of the *Toona* that differs from the nuclear and mitochondrial

genomes in the following aspects: (1) Selected high-qualitys TS and TC for complete genome sequencing of cp can further enrich the GenBank database and contribute genetic species information to the cp genome research of the valuable Class II endangered *T.cilata* and the versatile economic *T.sinensis*. (2) Mapping of TS, TC and TCv and genomic information, such as SSRs, SNP and Indel provide candidate molecular markers for the molecular identification of TC and TCv. (3) Screening of highly variable regions and marker loci in *Toona* based on cp comparative genomics, providing a database for plant identification and DNA bar-coding probes. (4) Developmental tree construction, although the cp genomic data did not effectively resolve the phylogeny among species of taxa within branches (*T.ciliata* and its varieties), it can address the phylogenetic relationships among major branches (groups), and has meaningful implications for the description of specific species, providing the means to classify species of the *Toona*, which is not yet clearly delineated, and the species' phylogenetic position in the system.

5. Materials and Methods

5.1. *Plant material*

The cp sequence of TCv was downloaded from the NCBI (accession number: MZ926838). TS and TC were sampled from high-quality single plants of the Toona resource nursery at the site of South China Agricultural University (23°N, 113°E), both are excellent tree species that have been tested and selected for adaptation to growth in Guangzhou.

5.2. *DNA extraction, library construction, and high-throughput sequencing*

Extracted DNA from TS and TC leaves by CTAB[liii]. Screening of high-quality DNA by Nanodrop, 1% (w/v) agarose gel electrophoresis.

In order to generate no less than 4Gb of data per sample, the DNA library was sequenced by Illumina HiSeq 4000 high-throughput sequencing platform with Pair-End 150 sequencing strategy. A clean read was used as the basis for all subsequent analyses.

5.3. *Genome assemble and annotation*

Multiple iterations of Velvet were run in different pairwise combinations of Kmer values (51, 61, 71, 81, 91) and expected coverage values (50, 100, 200, 500, 1000, 2000) by splicing clean reads with the Velvet version 1.1.06 [liv] tool. Each run, the mineral coverage is set to 10% of the expected coverage and the scaffold is off. The plastid genome of *Azadirachta indica* (GenBank accession number KF986530) was targeted as the query sequence by using BlastN. For each genome, contigs from the best three assemblies, which maximized the average length of chloroplast contigs, were aligned manually, and the consensus of these three assemblies was taken as the final sequence.

Initially, CpGAVAS[lv], and DOGMA[lvi] were utilized to annotate the cp genome. The *Azadirachta indica* cp genome sequence was then searched with BlastN, and manually proofread for start and stop codons, and intron and exon boundaries. Predicted the differences in the genome, protein-coding genes, introns, and spacer sequences with DnaSP 5.10[lvii]. For protein-coding gene sequences, each gene or fragment was edited by ClustalW multiple alignment options in BioEdit v7.0.9.0[lviii]. The tRNAscan-SE[lix] was performed to identify tRNA genes. The mapping was performed by the visualization software Organellar Genome DRAW[lx]. Annotated cp genome sequences were deposited in GenBank under accession numbers OK572964 and OK572965.

5.4. *Cp structural features analysis*

The GC contents of the three *Toona* plants were calculated by the Nucleotide Composition function in Bioedit v7.2.5[lxi]. SSRs were analyzed by MISA[lxii], and the minimum number of repeats of mono-, di-, tri-, tetra-, penta, and hexanucleotide was set to 10, 6, 5, 5, 6, and 5, in that order[lxiii], repeated 3 times. Long repeat sequences were analyzed with REPuter[lxiv], repeats were not less than 30 bp, and no longer than 60 bp, and the similarity between the two repeats was over 90%. The frequency of codon usage and RSCU values were calculated with MEGA 11[lxv]. Both synonymous codon usage order (SCUO) and measures independent of length and composition (MILC) were calculated

with coRdon 1.18.0, default parameters correlation analysis of SCUO and MILC values was performed utilizing Karl Pearson's correlation analysis method[lxvi].

5.5. Genome comparison and nucleotide variation analysis

Download the cp complete genome of *Toona ciliate* var. *yunnanensis* (OM135326.1), *Toona ciliate* var. *henryi* (OM135325.1), and *Toona fargesii* (NC_069639.1) from NCBI together with the 3 genomes for comparative analysis. IRscope[lxvii] was utilized to assess IR expansion and contraction. MAFFT v7.490 was applied to compare cp genome sequences and identify variation hotspots. DnaSP v5.1[lxviii] was performed to calculate nucleotide polymorphism (Pi) among the cp genomes, the sliding window length was set to 200bp and the step size was 60bp. With TS(OK5729.65.1) as control and the parameters set to default coefficients, was applied maff v7.490 to perform multiple sequence alignment and record the snp and indel loci of TC and TCv. Download the cp entire genomes of TC, TCV and TC from NCBI, in addition to the three in this review, and arrangement match them by maff and python programming.

5.6. Phylogenetic Analysis

31 cp sequences of Meliaceae were downloaded from NCBI (Detailed plant information in the "taxonomy.xls" file), added to the 3 sequences in this work, performed the alignment of individual genes among multiple species by MUSCLE v.3.8.31[lxix], and then integrated the genes of each species in a certain order to form the protein-coding gene sequence set of the species, followed by the sequences for further analysis. Nucleic acid models were tested based on the selected sequence DNA using jModelTest 2.1.7[lxx], Prottest3.2[lxxi] was used for the amino acid model test, and the best model for tree construction was selected as the minimum value of AIC (Akaike Information Criterion).

RAxML8.1.5[lxxii] was adopted to construct the phylogenetic tree using the ML method. The optimal model for constructing the tree was GTR+I+G.

Declaration

Ethics approval and consent to participate

All samples were adopted for the total experiment. No specific permits are required for sample collection in this study. We comply with the IUCN Policy Statement on Research Involving Species at Risk of Extinction and the Convention on the Trade in Endangered Species of Wild Fauna and Flora.

Each member of the team declared that has read the relevant institutional, national, and international guidelines and legislation before the commencement of the experiment and had complied with each of the legislative requirements during the experiment.

The identified samples were preserved in South China Agricultural University Herbarium(CANT), where the voucher for TS was 33208 and the voucher for TC was 33209 (Note: Professor Xiaoyang Chen and Teacher WeiZhou have carried out a detailed identification of the plant material).

Consent for publication

Not applicable.

Competing interests

The authors declare no conflict of interest.

References

[i] Mirza AU, Khan MS, Kareem A, Nami SAA, Bhat SA, Mohammad A, Singh P, Nishat N. Biomediated synthesis, characterization, and biological applications of nickel oxide nanoparticles derived from Toona ciliata, Ficus carica and Pinus roxburghii. Bioprocess Biosyst Eng. 2021 Jul;44(7):1461-1476. doi: 10.1007/s00449-021-02528

-4. Epub 2021 Apr 5. PMID: 33818638.

[ii] Yan Y, Min Y, Min H, Chao C, Ying Q, Zhi H: n-Butanol soluble fraction of the water extract of Chines e toon fruit ameliorated focal brain ischemic insult in rats via inhibition of oxidative stress and inflammat ion. *Journal of Ethnopharmacology* 2014, 151(1):176-182.

[iii] Gautam A, Jhade D, Ahirwar D, Sujane M, Sharma GN. Pharmacognostic evaluation of toona ciliata bar k. J Adv Pharm Technol Res. 2010 Apr;1(2):216-20. PMID: 22247848; PMCID: PMC3255442.

[iv] Zhang PP, Cui ZR, Li Y, Heng L, Wei SS, Wang XB, Yang MH, Luo J, Kong LY. Ciliatasecones A-C, three rearranged limonoids from Toona ciliata var. yunnanensis. Org Biomol Chem. 2019 Jan 16;17(3):555-560. doi: 10.1039/c8ob02736f. PMID: 30574992.

[v] Ndlovu MJ, Bagla VP, Mokgotho MP, Makgatho ME, Matsebatlela TM. Potential anticancer activity of Ac etone extracts of *Toona cilliata*, *Seriphium plumosum* and *Schkuhria pinnata* on HeLa cervical cancer cells. Afr H ealth Sci. 2021 Jun;21(2):663-672. doi: 10.4314/ahs.v21i2.23. PMID: 34795721; PMCID: PMC8568207.

[vi] Beseni BK, Matsebatlela TM, Bagla VP, Njanje I, Poopedi K, Mbazima V, Mampuru L, Mokgotho MP. Poten tial Antiglycation and Hypoglycaemic Effects of *Toona ciliata* M. Roem. and *Schkuhria pinnata* Lam. Thell. Crude Extracts in Differentiated C2C12 Cells. Evid Based Complement Alternat Med. 2019 Jan 21;2019:5406862. doi: 10.1155/2019/5406862. PMID: 30805018; PMCID: PMC6363240.

[vii] Li P, Shang Y, Zhou W, Hu X, Mao W, Li J, Li J, Chen X. Development of an efficient regeneration syste m for the precious and fast-growing timber tree *Toona ciliata*. Plant Biotechnol (Tokyo). 2018;35(1):51-58. doi: 1 0.5511/plantbiotechnology.18.0130a. Epub 2018 Mar 25. PMID: 31275037; PMCID: PMC6543735.

[viii] Brunt C, Read J, Sanson GD. Changes in resource concentration and defence during leaf development in a tough-leaved (Nothofagus moorei) and soft-leaved (Toona ciliata) species. Oecologia. 2006 Jul;148(4):583-92. Epub 2006 May 19. PMID: 16710660.

[ix] Yang HL, Kuo YT, Vudhya Gowrisankar Y, Lin KY, Hsu LS, Huang PJ, Lin HC, Hseu YC. The Leaf Extrac ts of *Toona sinensis* and Fermented Culture Broths of *Antrodia camphorata* Synergistically Cause Apoptotic Cell Death in Promyelocytic Leukemia Cells. Integr Cancer Ther. 2020 Jan-Dec;19:1534735420923734. doi: 10.1177/1 534735420923734. PMID: 32618215; PMCID: PMC7336824.

[x] Yuk HJ, Lee YS, Ryu HW, Kim SH, Kim DS. Effects of Toona sinensis Leaf Extract and Its Chemical Cons tituents on Xanthine Oxidase Activity and Serum Uric Acid Levels in Potassium Oxonate-Induced Hyperu ricemic Rats. Molecules. 2018 Dec 9;23(12):3254. PMID: 30544886; PMCID: PMC6321014.

[xi] Hanif S, Alsaiari M, Ahmad M, Sultana S, Zafar M, Rozina, Harraz FA, Alharbi AF, Abahussain AAM, Ah mad Z. Membrane reactor based synthesis of biodiesel from Toona ciliata seed oil using barium oxide nano catalyst. Chemosphere. 2022 Dec;308(Pt 3):136458.

[xii] Mirza AU, Khan MS, Kareem A, Nami SAA, Bhat SA, Mohammad A, Singh P, Nishat N. Biomediated synt hesis, characterization, and biological applications of nickel oxide nanoparticles derived from Toona ciliata, Fi cus carica and Pinus roxburghii. Bioprocess Biosyst Eng. 2021 Jul;44(7):1461-1476. doi: 10.1007/s00449-021-0252 8-4. Epub 2021 Apr 5. PMID: 33818638.

[xiii] Wu H, Song Z, Niu X, Liu J, Jiang J, Li Y. Classification of *Toona sinensis* Young Leaves Using Machine Le arning and UAV-Borne Hyperspectral Imagery. Front Plant Sci. 2022 Jun 28;13:940327. doi: 10.3389/fpls.2022.9 40327. PMID: 35837456; PMCID: PMC9274089.

[xiv] Liu J, Jiang JM, Chen YT. Genetic diversity of central and peripheral populations of Toona ciliata var. pub escens, an endangered tree species endemic to China. Genet Mol Res. 2014 Jun 17;13(2):4579-90. doi: 10.4238/ 2014.June.17.10. PMID: 25036507.

[xv] Zhao H, Ren L, Fan X, Tang K, Li B: Identification of putative flavonoid-biosynthetic genes through transcr iptome analysis of Taihe Toona sinensis bud. *Acta Physiologiae Plantarum* 2017, 39(6).

[xvi] Wang X, Xiao Y, He ZH, Li LL, Lv YW, Hu XS. Evolutionary Divergence between *Toona ciliata* and *Toona s inensis* Assayed with Their Whole Genome Sequences. Genes (Basel). 2022 Oct 5;13(10):1799. doi: 10.3390/gene s13101799. PMID: 36292684; PMCID: PMC9602400.

[xvii] Li Y, Gu M, Liu X, Lin J, Jiang H, Song H, Xiao X, Zhou W. Sequencing and analysis of the complete m itochondrial genomes of Toona sinensis and Toona ciliata reveal evolutionary features of Toona. BMC Genom ics. 2023 Feb 1;24(1):58. doi: 10.1186/s12864-023-09150-6. PMID: 36726084; PMCID: PMC9893635。

[xviii] Heinze B. Sequencing of Complete Chloroplast Genomes. Methods Mol Biol. 2021;2222:89-105. doi: 10.1007 /978-1-0716-0997-2_5. PMID: 33301089.

[xix] Daniell H, Jin S, Zhu XG, Gitzendanner MA, Soltis DE, Soltis PS. Green giant-a tiny chloroplast genome w ith mighty power to produce high-value proteins: history and phylogeny. Plant Biotechnol J. 2021 Mar;19(3):4 30-447. doi: 10.1111/pbi.13556. Epub 2021 Feb 22. PMID: 33484606; PMCID: PMC7955891.

[xx] Shidhi PR, Nadiya F, Biju VC, Vijayan S, Sasi A, Vipin CL, Janardhanan A, Aswathy S, Rajan VS, Nair A S. Complete chloroplast genome of the medicinal plant *Evolvulus alsinoides*: comparative analysis, identification of mutational hotspots and evolutionary dynamics with species of Solanales. Physiol Mol Biol Plants. 2021

Aug;27(8):1867-1884. doi: 10.1007/s12298-021-01051-w. Epub 2021 Aug 25. PMID: 34539121; PMCID: PMC8405790.

[xxi] Park I, Song JH, Yang S, Kim WJ, Choi G, Moon BC. *Cuscuta* Species Identification Based on the Morphology of Reproductive Organs and Complete Chloroplast Genome Sequences. Int J Mol Sci. 2019 Jun 3;20(11):2726. doi: 10.3390/ijms20112726. PMID: 31163646; PMCID: PMC6600609.

[xxii] Kaewdaungdee S, Sudmoon R, Tanee T, Lee SY, Chaveerach A. Chloroplast Genome Analysis for Genetic Information and Authentication in Five *Barleria* Species. Genes (Basel). 2022 Sep 22;13(10):1705.

[xxiii] Asaf S, Khan AL, Khan A, Al-Harrasi A. Unraveling the Chloroplast Genomes of Two *Prosopis* Species to Identify Its Genomic Information, Comparative Analyses and Phylogenetic Relationship. Int J Mol Sci. 2020 May 6;21(9):3280.

[xxiv] Lu Z, Dong X, Fan Y, Liu W, Dai J, Han X, Liu J. Complete chloroplast genome of *Toona ciliata* Roem. var. *pubescens* (Franch.) Hand.-Mazz (Meliaceae), 'Chinese mahogany'. Mitochondrial DNA B Resour. 2022 Mar 16;7(3):495-497.

[xxv] Xin GL, Liu JQ, Liu J. Complete chloroplast genome of an endangered tree species, *Toona ciliata* (Sapindales: Meliaceae). Mitochondrial DNA B Resour. 2018 Jun 8;3(2):663-664. doi: 10.1080/23802359.2018.1476074.

[xxvi] Yang Z, Li L, Chen CH, Zhang YY, Yang Y, Zhang P, Bao GH. Chemical composition and antibacterial activity of 12 medicinal plant ethyl acetate extracts using LC-MS feature-based molecular networking. *Phytochem Anal*. 2022 Apr;33(3):473-489.

[xxvii] Li, P.; Zhan, X.; Que, Q.; Qu, W.; Liu, M.; Ouyang, K.; Li, J.; Deng, X.; Zhang, J.; Liao, B.; Pian, R.; Chen, X. Genetic Diversity and Population Structure of *Toona Ciliata* Roem. Based on Sequence-Related Amplified Polymorphism (SRAP) Markers. *Forests* 2015, *6*, 1094-1106.

[xxviii] Liu J, Jiang JM. Sampling strategies for natural Toona ciliata populations. Genet Mol Res. 2016 Oct 17;15(4).

[xxix] Somaratne Y, Guan DL, Wang WQ, Zhao L, Xu SQ. The Complete Chloroplast Genomes of Two Lespedeza Species: Insights into Codon Usage Bias, RNA Editing Sites, and Phylogenetic Relationships in Desmodieae (Fabaceae: Papilionoideae). Plants (Basel). 2019 Dec 31;9(1):51.

[xxx] Katoh K, Rozewicki J, Yamada KD. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. Brief Bioinform. 2019 Jul 19;20(4):1160-1166. PMID: 28968734; PMCID: PMC6781576.

[xxxi] Daniell H, Jin S, Zhu XG, Gitzendanner MA, Soltis DE, Soltis PS. Green giant-a tiny chloroplast genome with mighty power to produce high-value proteins: history and phylogeny. Plant Biotechnol J. 2021 Mar;19(3):430-447.

[xxxii] Mohanta TK, Khan AL, Hashem A, Allah EFA, Yadav D, Al-Harrasi A. Genomic and evolutionary aspects of chloroplast tRNA in monocot plants. BMC Plant Biol. 2019 Jan 22;19(1):39.

[xxxiii] Naver H, Boudreau E, Rochaix JD. Functional studies of Ycf3: its role in assembly of photosystem I and interactions with some of its subunits. Plant Cell. 2001 Dec;13(12):2731-45. doi: 10.1105/tpc.010253. PMID: 11752384; PMCID: PMC139485.

[xxxiv] Krech, K.; Ruf, S.; Masduki, F.F.; Thiele, W.; Bednarczyk, D.; Albus, C.A.; Tiller, N.; Hasse, C.; Schöttler, M.A.; Bock, R. The Plastid Genome-Encoded Ycf4 Protein Functions as a Nonessential Assembly Factor for Photosystem I in Higher Plants. *Plant Physiol*. 2012, 159, 579–591.

[xxxv] Boudreau E, Takahashi Y, Lemieux C, Turmel M, Rochaix JD. The chloroplast ycf3 and ycf4 open reading frames of *Chlamydomonas reinhardtii* are required for the accumulation of the photosystem I complex. EMBO J. 1997 Oct 15;16(20):6095-104.

[xxxvi] Sundararajan A, Dukowic-Schulze S, Kwicklis M, Engstrom K, Garcia N, Oviedo OJ, Ramaraj T, Gonzales MD, He Y, Wang M, Sun Q, Pillardy J, Kianian SF, Pawlowski WP, Chen C, Mudge J. Gene Evolutionary Trajectories and GC Patterns Driven by Recombination in Zea mays. Front Plant Sci. 2016 Sep 22;7:1433.

[xxxvii] Pop C, Rouskin S, Ingolia NT, Han L, Phizicky EM, Weissman JS, Koller D. Causal signals between codon bias, mRNA structure, and the efficiency of translation and elongation. Mol Syst Biol. 2014 Dec 23;10(12):770.

[xxxviii] Chakraborty S, Yengkhom S, Uddin A. Analysis of codon usage bias of chloroplast genes in Oryza species : Codon usage of chloroplast genes in Oryza species. Planta. 2020 Sep 28;252(4):67.

[xxxix] Supek F, Vlahovicek K. Comparison of codon usage measures and their applicability in prediction of microbial gene expressivity. BMC Bioinformatics. 2005 Jul 19;6:182.

[xl] Liu J, Sun ZX, Chen YT, Jiang JM. Isolation and characterization of microsatellite loci from an endangered tree species, Toona ciliata var. pubescens. Genet Mol Res. 2012 Dec 17;11(4):4411-7. .

[xli] Liu J, Jiang JM, Chen YT. Genetic diversity of central and peripheral populations of Toona ciliata var. pubescens, an endangered tree species endemic to China. Genet Mol Res. 2014 Jun 17;13(2):4579-90.

[xlii] Wills DM, Hester ML, Liu A, Burke JM. Chloroplast SSR polymorphisms in the Compositae and the

mode of organellar inheritance in Helianthus annuus. Theor Appl Genet. 2005 Mar;110(5):941-7.
[xliii] Xiao Y, Zhang XX, Hu Y, Wang X, Li P, He ZH, Lv YW, Chen XY, Hu XS. Phylogeography of *Toona cilia ta* (Meliaceae) Complex in China Inferred from Cytonuclear Markers. Genes (Basel). 2022 Dec 31;14(1):116.
[xliv] McPherson H, van der Merwe M, Delaney SK, Edwards MA, Henry RJ, McIntosh E, Rymer PD, Milner ML, Siow J, Rossetto M. Capturing chloroplast variation for molecular ecology studies: a simple next generati on sequencing approach applied to a rainforest tree. BMC Ecol. 2013 Mar 14;13:8.
[xlv] Do HDK, Jung J, Hyun J, Yoon SJ, Lim C, Park K, Kim JH. The newly developed single nucleotide polym orphism (SNP) markers for a potentially medicinal plant, Crepidiastrum denticulatum (Asteraceae), inferred fr om complete chloroplast genome data. Mol Biol Rep. 2019 Jun;46(3):3287-3297.
[xlvi] Ismail M, Ahmad A, Nadeem M, Javed MA, Khan SH, Khawaish I, Sthanadar AA, Qari SH, Alghanem S M, Khan KA, Khan MF, Qamer S. Development of DNA barcodes for selected *Acacia* species by using rbcL an d matK DNA markers. Saudi J Biol Sci. 2020 Dec;27(12):3735-3742.
[xlvii] Jung J, Do HDK, Hyun J, Kim C, Kim JH. Comparative analysis and implications of the chloroplast geno mes of three thistles (*Carduus* L., Asteraceae). PeerJ. 2021 Jan 14;9:e10687.
[xlviii] Zhou T, Wang J, Jia Y, Li W, Xu F, Wang X. Comparative Chloroplast Genome Analyses of Species in G entiana section Cruciata (Gentianaceae) and the Development of Authentication Markers. Int J Mol Sci. 2018 Ju l 5;19(7):1962.
[xlix] Alwadani KG, Janes JK, Andrew RL. Chloroplast genome analysis of box-ironbark Eucalyptus. Mol Phylog enet Evol. 2019 Jul;136:76-86.
[l] Gu X, Zhu M, Su Y, Wang T. A Large Intergenic Spacer Leads to the Increase in Genome Size and Sequenti al Gene Movement around IR/SC Boundaries in the Chloroplast Genome of Adiantum malesianum (Pteridace ae). Int J Mol Sci. 2022 Dec 9;23(24):15616.
[li] Yang CH, Liu X, Cui YX, Nie LP, Lin YL, Wei XP, Wang Y, Yao H. Molecular structure and phylogenetic analyses of the complete chloroplast genomes of three original species of Pyrrosiae Folium. Chin J Nat Med. 2020 Aug;18(8):573-581.
[lii] Javaid N, Ramzan M, Khan IA, Alahmadi TA, Datta R, Fahad S, Danish S. The chloroplast genome of Fars etia hamiltonii Royle, phylogenetic analysis, and comparative study with other members of Clade C of Brassi caceae. BMC Plant Biol. 2022 Aug 2;22(1):384.
[liii] Doyle J J, Doyle J L. A rapid DNA isolation procedure for small quantities of fresh leaf tissueJ. Phytochem Bull, 1987, 19(1): 11–15.
[liv] Zerbino DR, Birney E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. 2008 May;18(5):821-9.
[lv] Liu et al : CpGAVAS, an integrated web server for the annotation, visualization, analysis, and GenBank su bmission of completely sequenced chloroplast genome sequences. BMC Genomics 2012 13:715.
[lvi] Automatic annotation of organellar genomes with DOGMA. Wyman SK,Jansen RK, Boore JL,Bioinformatics 2004 20 (17) : 3252-3255.
[lvii] Rozas, J., Sánchez-DelBarrio, J. C., Messeguer, X., and Rozas, R. (2003).DnaSP, DNA polymorphism analyse s by the coalescent and other methods.Bioinformatics 19, 2496–2497.
[lviii] Gupta SK, Padmanabhan BR, Diene SM, Lopez-Rojas R, Kempf M, Landraud L, Rolain JM. ARG-ANN OT, a new bioinformatic tool to discover antibiotic resistance genes in bacterial genomes. Antimicrob Age nts Chemother. 2014;58(1):212-20.
[lix] Todd M. Lowe, Patricia P. Chan, tRNAscan-SE On-line: integrating search and context for analysis of transf er RNA genes, *Nucleic Acids Research*, Volume 44, Issue W1, 8 July 2016, Pages W54–W57.
[lx] Lohse M, Drechsel O, Bock R. OrganellarGenomeDRAW (OGDRAW): a tool for the easy generation of hi gh-quality custom graphical maps of plastid and mitochondrial genomes. Curr Genet. 2007 Nov;52(5-6):267 -74.
[lxi] Gupta SK, Padmanabhan BR, Diene SM, Lopez-Rojas R, Kempf M, Landraud L, Rolain JM. ARG-ANNO T, a new bioinformatic tool to discover antibiotic resistance genes in bacterial genomes. Antimicrob Agent s Chemother. 2014;58(1):212-20.
[lxii] Thiel T, Michalek W, Varshney RK, Graner A. Exploiting EST databases for the development and chara cterization of gene-derived SSR-markers in barley (Hordeum vulgare L.). Theor Appl Genet. 2003 Feb;106 (3):411-22.
[lxiii] Faircloth BC. msatcommander: detection of microsatellite repeat arrays and automated, locus-specific prime r design. Mol Ecol Resour. 2008 Jan;8(1):92-4.
[lxiv] Kurtz S, Choudhuri JV, Ohlebusch E, Schleiermacher C, Stoye J, Giegerich R. REPuter: the manifold applic ations of repeat analysis on a genomic scale. Nucleic Acids Res. 2001 Nov 15;29(22):4633-42.
[lxv] Tamura K, Stecher G, Kumar S. MEGA11: Molecular Evolutionary Genetics Analysis Version 11. Mol Biol Evol. 2021 Jun 25;38(7):3022-3027.

[lxvi] Karan Sareen,Bhavesh R Bhalja,Sanjay Srivastava,Yogendra Swarnkar,Rudra Prakash Maheshwari. Islanding Detection Technique based on Karl Pearson's Coefficient of Correlation for Distribution Network with High P enetration of Distributed Generations[J]. International Journal of Emerging Electric Power Systems,2018,19(3).

[lxvii] Amiryousefi A, Hyvönen J, Poczai P. IRscope: an online program to visualize the junction sites of chloro plast genomes. Bioinformatics. 2018 Sep 1;34(17):3030-3031.

[lxviii] Rozas J, Sánchez-DelBarrio JC, Messequer X, Rozas R. DnaSP, DNA polymorphism analyses by the c oalescent and other methods. Bioinformatics. 2003 Dec 12;19(18):2496-7.

[lxix] Edgar, R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput . Nuclei c Acids Res. 32(5):1792-1797.

[lxx] David Posada; jModelTest: Phylogenetic Model Averaging, Molecular Biology and Evolution, Volume 25, Iss ue 7, 1 July 2008, Pages 1253–1256.

[lxxi] Diego Darriba, Guillermo L. Taboada, Ramón Doallo, David Posada; ProtTest 3: fast selection of best-fit m odels of protein evolution, Bioinformatics, Volume 27, Issue 8, 15 April 2011, Pages 1164–1165.