Comparative Analysis of the Complete Chloroplast Genomes in Toona sinensis and Toona ciliata: Phylogenetic Relationship of Toona

Youli Li
South China Agricultural University
Min Gu
South China Agricultural University
Jianna Lin
South China Agricultural University
Huier Jiang
South China Agricultural University
Xingcui Xiao
Sichuan Academy of Forestry Sciences
Wei Zhou (wzhou@scau.edu.cn)
South China Agricultural University

Article

Keywords: Toona sinensis, Toona ciliata, chloroplast complete genome, phylogenetic evolution

Posted Date: July 8th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1766876/v1

Abstract

Taking the chloroplast genomes of *Toona sinensis* (GenBank:OK572965) and *Toona ciliata* (GenBank:OK572964.1) as the research objects, the structural characteristics and phylogenetic relationship of their chloroplast (cp) genomes were researched. Here, we sequenced the complete chloroplast genome sequence using Illumina HiSeq high-throughput sequencing technology via a combination of annotations and simultaneously compared their cp genome structure features. Codon usage bias was aligned with the previously reported *Azadirachta indica* (KF986530), together with IR boundary analysis with other investigated relatives, and complete genome sequences were compared to each other to construct maximum likelihood (ML) phylogenetic trees. The results showed that (1) the chloroplast genomes of *Toona sinensis* and *Toona ciliata* exhibited typical quadripartite and circular structures that were relatively conserved in genomic structure and the synteny of gene order; both *Toona sinensis* and *Toona ciliata* annotated 125 genes, including 88 protein-coding genes (PCGs), 29 transfer RNAs (tRNAs) and 8 ribosomal RNAs (rRNAs). (2) Fifteen and 13 palindromic repeats, 10 and 15 forward repeats and a single complementary repeat were detected in the chloroplast genomes of the two species. At the same time, 61 and 55 SSRs were identified, respectively, which were primarily located in the coding interval. There was a significant codon preference in the genome, with codons that end in A and T used much more frequently than synonymous codons. (3) IR boundary and sequence divergence analyse found some candidate fragments to identify the selected related species, such as *trnH, ycf1*,etc. (4) A phylogenetic tree was constructed using the maximum likelihood (ML) method, and the results revealed that *Toona sinensis* and *Toona ciliata* clustered into a single independent branch with high support. This investigation has established a theoretical foundation for plant identification, molecular marker development, conservation efforts, population dynamics and further study of phylogenetic processes. The availability of chloroplast genomes provides valuable genetic information for accurate identification, taxonomy, and phylogeny and contributes to the exploration and utilization of plants.

Introduction

Toona is an important genus in the Meliaceae family that is widespread in China and abundant in germplasm, containing 4 species with 3 varieties[1], of which the species with the greatest economic and ornamental value are *Toona sinensis* and *Toona ciliata*[2,3]. *Toona sinensis* is a woody plant native to China that has a rich nutritional value. The roots, stems and leaves of *Toona sinensis* can be used for medicinal purposes and have good effectiveness[4,5,6]. *Toona* ciliata is not only a valuable timber species but is also known as the Chinese mahogany (*Swietenia mahagoni*) and is a national class II key protected wild plant[7,8,9,10,11]. The economic value of this genus is soaring, with great potential for exploitation and utilization[12,13,14,15]. In the past, few scholars have systematically studied the Chinese *Toona*, and little research has been done on other species of *Toona*, especially their origin, classification, and cytogenetics[16,17]. The taxonomic information on the *Toona* in FRPS is not yet comprehensive, e.g., there is confusion between species, so it is necessary to systematically and comprehensively analyze the genetics of *Toona*[18,19]. The development through genetic sequencing and the widespread use of

systems along with the construction of phylogenetic trees have made it possible to explore the *Toona* in further depth from a molecular point of view, in addition to morphological classification[20]. The purpose of this systematic study is to clarify the germplasm resources of the *Toona* to provide a foundation for the exploitation of its medicinal, edible and timber resources while avoiding confusion; at the same time, the evolutionary relationships among the species of the genus can be explored, which is not only of great practical value but also of great scientific significance.

Chloroplasts are the essential organelles for photosynthesis in green plants and are responsible for the manufacture of organic material and the storage of energy[21]. Accompanying intensive biotechnology evolution, the chloroplast genome structure has been discovered to reveal the origin or evolution of species in general terms in addition to their relatedness to various species[22]. The cp genome is one of three DNA genomes in plants. It is characterized by a conserved architecture, stability, low uniparental inheritance, moderate sequence variation and convenient sequencing analysis and offers relatively consecutive sequences, moderate variation and limited genetic recombination rates in comparison to the nuclear and mitochondrial genomes. The cp genome is one of three DNA genomes in plants. The chloroplast genome is one of the three DNA genomes in plants. It is characterized by a conserved structure, high stability, moderate uniparental inheritance, moderate sequence variability and ease of sequencing, and limited genetic recombination rates in comparison to the nuclear and mitochondrial genomes[23]. By virtue of its relatively homogeneous genome structure and integrity of the genome sequence, it has been broadly accepted in the field of biological research, presenting a worthwhile source of information on evolutionary biology and constituting a powerful tool for resolving plant phylogenies[24]. In particular, at the IR boundary of the chloroplast genome, there is a certain amount of expansion and variation across species, which is of great value in researching the evolution of green plants and revealing relationships. Therefore, the cp genome is widely used for germplasm identification, molecular breeding, phylogenetic and genetic diversity research, etc. It is the primary tool in molecular biology investigations[25].

In this study, the cp complete genome sequences of two species, *oona sinensis* (GenBank:OK572965) and *Toona ciliata* (GenBank:OK572964.1), were assembled, annotated and structurally analyzed, aiming to comprehensively explore the overall characteristics of the cp genomes of these two species within *Toona* and to screen the interspecific high variation sequences, which can be employed in the identification of the two woody plants. Finally, the taxonomic classification of relatives of the *Toona* are discussed and can provide a theoretical basis for exploring their evolution with respect to the genetics and essential germplasm resources of the related taxa.

Materials And Methods

1.1 Plant material, DNA library construction and High-throughput sequencing

Toona sinensis was selected from Pingxiang, Guangxi(Longitude: 106.75E, Latitude: 22.12 N), while *Toona ciliata* was selected from Baoshan, Yunnan(Longitude: 99.14531E, Latitude: 25.11868N), both of

which were subjected to seedling trials before this study and were found suitable for growth in Guangzhou, Guangdong. (Note: WeiZhou has carried out a detailed identification of the plant material.

The seed trial forest is located at the teaching and research base of South China Agricultural University, Guangzhou, China, with a geographical location of N23°16', E113°37'.) The DNA samples of aromatic *Toona sinensis* and *Toona ciliata* were sent to Guangzhou Ruike Gene Technology Co., Ltd., for testing in terms of purity, concentration and fragment distribution range. The samples with quality control results of Class B or above were qualified and utilized as the materials for this experiment. A noncontact ultrasonic breaker Bioruptor Pico was used to randomly break the genome into fragments with a length of approximately 500 bp to construct a small-fragment genomic DNA library. The library construction procedure was carried out according to the instructions of the "VAHTSTM Universal DNA Library Prep Kit for Illumina®" from Vazyme Biotech Co. Following library construction, the libraries were quality-controlled by qPCR and Agilent 2100 Bioanalyzer.

Qualified DNA libraries were sequenced using the Illumina HiSeq 4000 High-throughput sequencing platform with a sequencing strategy of PE150 (Pair-End 150), and the amount of sequencing data per sample was not less than 4 Gb. The original offline sequences obtained by the Illumina HiSeq 4000 sequencing platform were called Raw Reads. The data processing was completed through the process of removing low-quality sequences and decontaminating connectors to obtain high-quality sequences, which were called clean reads. All subsequent analyses were based on clean reads. The judgment principle for removing low-quality sequences was as follows: when the N content in any sequencing read exceeded 10% of the number of bases in the read, the paired reads were removed; when the number of low-quality (Q<=5) bases contained in a sequenced read surpassed 50% of the number of bases in that read, the paired reads were removed.

1.2 Assembly and Annotation

The generated clean reads were assembled by splicing employing the SPAdes v.3.5.0[26]tool. Since the integrity of the chloroplast genome assembly is influenced by the content of chloroplast DNA in the total DNA sample in terms of sequence complexity, various approaches, such as high-throughput and PCR+conventional sequencing, should be combined to ensure that the complete cyclic genome sequence is obtained as far as possible to ensure the quality of subsequent information analysis. CpGAVAS[27], DOGMA[28] and ORF Finder were deployed to annotate the chloroplast genome. Mapping was performed by the visualization software Organellar Genome DRAW[29]. To annotate more precisely, the preliminary results of the annotation were checked against the reported chloroplast genomes of the related species *Azadirachta indica* (KF986530) by applying the method of Blastn and Blastp (https://blast.ncbi.nlm.nih.gov/Blast.cgi) for coding proteins, introns and spacers, and protein-coding genes, and differences in the genome, protein-coding gene, intron and spacer sequences were assessed with DnaSP 5.10[30]. Genomic sequences were calibrated in MAFFT v5[31] and adjusted manually when necessary. For the protein-coding gene sequences, introns and spacer sequences, each gene or fragment was edited using the ClustalW multiple proofing option in BioEdit v7.0.9.0[32] software.

According to the requirements of NCBI, the verified genome sequence and gene annotation in a sequence file were submitted to NCBI directly by emailing the SQN file in the submit2ncbi directory as an attachment to the official NCBI email address. (gb-sub@ncbi.nlm.nih.gov). The final accession numbers of the NCBI were obtained as Toona sinensis (GenBank:OK572965 https://www.ncbi.nlm.nih.gov/nuccore/OK572965) and Toona ciliata GenBanK:OK572964.1 https://www.ncbi.nlm.nih.gov/nuccore/OK572964.1).

1.3 Chloroplast Genomics Analysis

RSCU

The extent of codon usage in the chloroplast genomes of *Toona sinensis* and *Toona ciliata* was analyzed with reference to the formula mentioned in the Sharp PM literature[33] .

GC -skew

The GC content was calculated by Bioedit v7.2.5[34] software, and the GC/AT offset rate statistics for *Toona sinensis* (OK572965) and *Toona ciliata* (OK572964) were carried out with mVISTA software, the GC offset rates of the three species were compared based on the GC offset results for *Azadirachta indica* published on NCBI.

Expansion and Contraction of IR boundary

The IR region is highly conserved in the chloroplast genome; however, there is some variation at the IR/SC boundary. The expansion and contraction of IR region boundaries are thought to be primarily responsible for the alterations of the chloroplast genome size in angiosperms. Comparative analysis of the annotated chloroplast genomes of *Toona ciliata var. Pubescens* (MZ926838.1), *Toona ciliata* (OK572964), *Toona sinensis* (OK572965), *Carapa guianensis* (NC_037442.1), *Cedrela odorata* (NC-037251.1), and *Xylocarpus rumphii* (NC_038199.1) was conducted. The annotated information of the chloroplast genome was then used to calculate the IR boundary data then map the boundary differences between the IR and SC regions of the simple chloroplast genome structure. Mapping was based on the statistical results and the use of IRscope[35](https://irscope. Shinyapps.io/irapp/) to plot the IR boundary comparison.

Divergence Hotspot Regions

We submitted the chloroplast complete genome location and sequence information of *Toona ciliata var. Pubescens* (MZ926838.1), *Toona ciliata* (OK572964), *Toona sinensis* (OK572965), *Carapa guianensis* (NC_037442.1), *Cedrela odorata* (NC-037251.1), and *Xylocarpus rumphii* (NC_038199.1) to mVISTA. The analysis was performed on the website using MAFFT to align the chloroplast genomes of these plants and by using the sliding window in DNAspv5[36] to perform bias analysis of the gene sequences.

Repetitive Sequences

The chloroplast genome sequences were analyzed with MISA software. Parameters were set as follows: single nucleotide repeats ≥ 10; dinucleotide repeats ≥ 6; trinucleotide, tetranucleotide, pentanucleotide and hexanucleotide repeats ≥ 5; and the minimum distance between the 2 SSRs was set to 100 bp. Analysis of the chloroplast long repeat >30 bp utilized REPuter[37]](http://bibiserv.techfak.uni-bielefeld.de/reputer/).

Phylogenetic Analysis

The cp genome sequence was downloaded from NCBI for 39 close relatives of *Toona*, including *Phellodendron amurense* (KY707335), *Toona ciliata* (OK572964) and *Toona sinensis* (OK572965). Their complete chloroplast genome sequence was used in RAxML8.1.5[38]] (https://sco.h-its.org/exelixis/web/software/raxml/index.html) application. We employed *Francoa sonchifolia* (Y047638.1) as the outgroup and then we applied the ML method (maximum likelihood method). The bootstrap value was 1000.

Results & Analysis

2.1 Genome Features

The chloroplast whole-genome lengths of *Toona sinensis* (OK572965) and *Toona ciliata* (OK572964) varied little, 159139 bp and 159,618 bp, with the large single copy region (LSC) at 87101 bp and 86707 bp and the small single copy region (SSC) at 18399 bp and 18382 bp, respectively(**Fig. 1**). Both cp genomes were annotated with 125 genes, of which 88 were protein-coding genes, 29 were tRNAs and 8 were rRNAs (**Table.** 1). The chloroplast genomes of *Toona sinensis* and *Toona ciliata* contained approximately similar GC contents, 37.9% and 37.89%, respectively. The IR region was the highest at 42.76%, the SSC was the lowest at 32.26% and 32.25%, and the LSC was in the middle at 36.06% and 36.05%, respectively (Appendix A). Among the four functional classifications of the chloroplast genomes of the two plants, there were 105 types, including photosynthesis-related genes (photosynthesis), self-replication genes (self-replication), other genes (other genes), unknown function genes (unknown) and 46, 51, 5, and 3 functional genes, respectively. Of the four major functional classifications of the chloroplast genomes of both species, 105 types of genes were identified, including photosynthesis-related genes, self-replication genes, other genes, and unknown functional gene species and categories, which were 46, 51, 5, and 3, respectively. Photosynthesis-related genes were mainly found in subunits of NADH dehydrogenase, the large subunit of Rubisco, subunits of photosystem II, subunits of photosystem I, subunits of ATP synthase, subunits of ATP synthase, and subunits of ATP synthase. The self-replicating genes are the small subunit of the ribosome, large subunit of the ribosome, DNA-dependent RNA polymerase, ribosomal RNA genes, and transfer RNA genes in the genome; other genes include c*emA, ccsA, accD, clpP,* and *matK*; the unknown genes are mainly *ycf1**, *ycf2*, and *ycf15* (Appendix B, C).

Table 1. Complete genome characteristics of chloroplasts from three plants.

Summary of complete chloroplast genomes for three species

	Toona ciliata OK572964	*Toona sinensis* OK572965	*Azadirachta indica* KF986530
Total	159618	159139	160737
Large single copy (LSC, bp)	87101	86707	88137
Inverted repeat (IR, bp)	27059	27025	26982
Small single copy (SSC, bp)	18399	18382	18636
GC%	37.89	37.9	37.49
Total	125	125	131
Protein coding genes	88	88	86
tRNA	29	29	37
rRNA	8	8	8

Note: ORF genes uncounted, Protein coding genes, tRNAs, rRNAs are counted quantity-wise, regardless of whether they are the same gene or not.

2.2 Codon Usage bias

The chloroplast genomes of *Toona sinensis* and *Toona ciliata* displayed the highest number of codons encoding Leu, with 2966 and 22,970 codons, accounting for 10.65% and 9.96% of all codons, respectively, among which 6 codons were synonymous, which is the largest variety of synonymous codons. Spt had the fewest codons, with 95 codons, resulting in 0.3% and 0.32% of all codons. Excluding the termination codon, the TGAs encoding Ile for ATT and Stp had the maximum and minimum number of occurrences as follows: 1132 and 22 for *Toona sinensis*, respectively, versus 1132 and 23 for *Toona ciliata* (Appendix D and E). In the set of synonymous codons, the number of codons ending in A/T was higher than the number ending in G(U)/C, where RSCU is also relatively more frequent. Among the total codons, the GC content only accounted for 37.9%, while codons ending in G/C accounted for 27.1%; within a group of synonymous codons, codons with RSCU > 1 all terminated in A/T; moreover, the value of RSCU was relatively higher than that of codons ending in G/C (**Fig. 2** and **Fig. 3**). It can be visually shown that among the synonymous codons, the frequency of usage of those ending in A/T is quite significant, which fully demonstrates the strong A/T preference.

2.3 IR region borders

Diverging degrees of variation in the contraction and expansion of the boundaries of the IR and SC regions determine the length and variety of genomes[39].In the evolutionary journey of plants, there has

been a major role for chloroplast genome diversity in angiosperms[40,41].

Although the chloroplast genome of *Toona* is relatively stable and conserved in terms of sequence length, genome composition and GC content, there is diversity in the transition regions of the four border regions. All six species show high conservation at the LSC and IRb boundaries, all with the *rpl22* gene, while at the IRb-SSC boundary, *Toona ciliata var. Pubescens* (MZ926838.1) lacks the ycf1 gene, while *Toona ciliata* (OK572964) has the *ndhF* gene located entirely in the SSC interval; except for *Cedrela odorata* (NC-037251.1), whose *Ycf1* gene straddles SSC and IRa, the SSC and IRa junction gene of the other five species was *ycf1*. The most significant variation was at the junction of IRa and LSC for *Carapa guianensis* (NC-037442.1), *Cedrela odorata* (NC-037251.1), and *Toona ciliata var. Pubescens* (MZ926838.1), with 2 bp, 5 bp and 4 bp *trnH* genes, respectively; for *Carapa guianensis* (NC-037442.1) there was a 6 bp DDH8 gene, while *Toona ciliata* (OK572964) and *Toona sinensis* (OK572965) had 5 bp and 1 bp tRNA genes, respectively(**Fig. 4**).

2.4 GC Skew Rate

In the *Toona sinensis* and *Toona ciliata* cp genomes, the AT skew was concentrated in the nonprotein coding regions, reaching a maximum in the 23S region with an AT skew of -0.1802; in the 16S region, the AT skew was -0.1318 and -0.1304, respectively; in the 5S and 4.5S regions, their skew rates were consistent, at -0.069 and -0.1373, respectively. In the protein coding and tRNA regions, the AT-skew of both was not significant, -0.017 and -0.011, so it can be concluded that the AT-skew of *Toona sinensis* and *Toona ciliata* was reflected in the nonprotein coding regions (Appendix F).

2.5 Genome Sequence Divergence

Analysis of the chloroplast genome sequence diversity of the six species with mVISTA revealed that the conserved noncoding sequences (CNS) were more diverse, the untranslated regions (UTR) were more conserved, and the exon region (exon) was between the UTR and CNS (**Fig. 5**).

Sequences in the major regions were in high agreement among the six species, demonstrating the evolutionary conservation of chloroplast genomes among CNS relatives. Although chloroplast genomes are more conserved among related species, there are still regions of moderate to high divergence, with the vast majority in the nonprotein-coding regions, tRNA-Ser(GCT), tRNA-Arg(TCT), tRNA-Cys(GCA), tRNA-Ser(GGA), tRNA-Arg(TRC), tRNA-Phe(GAA), and tRNA-Leu(TAG) regions with high divergence, and these regions with high variability can be used to design specific DNA barcodes. In the blue protein coding region, the encoded genes are more divergent than tRNAs, among which *matK, rps16, rpoC2, ycf3, Clp2, rrn23,* and *ndhF* are more variable, and these genes can be used to investigate the specific function of the species.

2.6 SSR Polymorphism and Repetitive Sequences

Repetitive Sequences

The *Toona sinensis* chloroplast genome contains 29 pairs of duplicated genes, of which 10, 15, 3 and 1 are forward (F), palindromic (P), inverted (R) and complementary (C) duplicates, respectively, whereas the *Toona ciliata* chloroplast genome contains 30 pairs of duplicated genes, of which 13, 15, 1 and 1 are forward, palindromic, inverted and complementary duplicates, respectively. The majority of the duplicates were identified in gene intervals. Most repeats were located in the intergenic spacer (IGS) and intronic regions (intron). In the protein-coding regions (CDs), the *ycf2* protein-coding gene was a hotspot for repeats, with a few in the protein-coding genes *ycf1* and *orf122* and only one complementary repeat in the protein-coding region *orf122*. The longest repeats in the *Toona sinensis* chloroplast genome are 147 bp and 190 bp in *Toona ciliata*, both palindromic repeats with repeated units in the protein-coding region of *ycf1. ycf2* (CDs) and *ycf3* (intron) are hotspots for repeats, especially *ycf2*, in the *Toona sinensis* chloroplast genome. The ycf2 gene, in particular, has eight repeats in the *Toona sinensis* chloroplast genome, including two forward and six palindromic repeats, and five repeats in the *Toona ciliata* chloroplast genome, including one forward and four palindromic repeats. *ycf3* (intron) has four repeats in the *Toona sinensis* chloroplast genome, two forward and two palindromic repeats, and two palindromic repeats in the *Toona ciliata* chloroplast genome, and 4 repeats, 2 forward and 2 palindromic repeats, in the *Toona ciliata* chloroplast genome (Appendix G and H).

SSRs

A total of 61 SSRs were detected in the chloroplast genome of *Toona sinensis*, of which 60 were single nucleotides and the elements were A and T, with 23 A elements and 39 T elements. Meanwhile, 55 SSRs were examined in the *Toona ciliata* cp genome, with 23 A elements and 31 T elements, illustrating the AT preference in the base composition of the SSRs (**Table. 2**). A repeat consisting of a dinucleotide AG was identified in which the structure was repeated six times, occurring in the coding interval region. The duplicated bases were mainly found in coding spacer regions, with 40 occurrences in the chloroplast genome of *Toona sinensis* and 35 in *Toona ciliata*; in the intron region, 10 and 9 times, respectively, occurring mainly in the yfc3 gene; in the CDs region, both occurring 11 times, primarily in the ycf1 coding gene. Their SSRs were 10-17 bp in length, with the smallest being a single nucleotide consisting of an A base, occurring in the orf107 gene encoding the interval region with CDs (Appendix I and J).

Table 2. Statistical table of the chloroplast genome SSRs of *Toona sinensis* and *Toona ciliata*

	Toona sinensis OK572965				*Toona ciliata* OK572964			
	SSR	IGS	CDS	Intron	SSR	IGS	CDS	Intron
mono-nucleotide	60	39	11	10	54	34	11	9
di-nucleotide	1	1	0	0	1	1	0	0

2.7 ML Phylogeny Analysis

To determine the phylogenetic status of *Toona sinensis* (OK572965) and *Toona ciliata* (OK572964), *Francoa sonchifolia* (Y047638.1) was used as the outgroup, and 34 species of Meliaceae were selected, plus 2 species of Rutaceae, 1 species of Simaroubaceae and 1 species of Burseraceae to construct an evolutionary tree. Using the maximum likelihood method (ML) to reconstruct the plant phylogenetic relationship of the *Toona* group (**Fig. 6**) . It can be concluded from the figure that there are 36 nodes in total, 29 of which have a support rate of 100%, and the same family is clustered together. *Toona* and *Cedrela odorata* are clustered into one category, namely, Subfam. Cedreloideae and compared to Subfam. Melioidea, Toona's branch and Subfam. Swietenioideae have higher support rates, 84% and 100%, respectively, which is similar to that of the Flora of China FRPC[42].

Discussions

3.1 Chloroplast Structure

In this article, we report the complete chloroplast genomes of *Toona sinensis* (OK572965) and *Toona ciliata* (OK572964). Both are extremely conserved in the structure of the chloroplast genome, with minimal divergences in the entire gene sequence as well as in the SC and IR regions. All exhibit traditional quadratic circular structures, which is consistent with the chloroplast genomes of angiosperms. The overall length of the genome was within 1 bp of each other in the four regions; the GC content in the total genome, LSC, SSC, and IR regions were all within 0.1%. Consequently, it is extremely challenging to distinguish the two species by genome sequence length or GC content alone. The GC contents of *Toona sinensis* and *Toona ciliata* were 37.9% and 37.89%, respectively, and these variations were primarily reflected in the SC region. Similar to other angiosperms, the GC content in the IR region is apparently above that in the SC region in *Toona sinensis* and *Toona ciliata*, associated with the location of all rRNA genes in the IR region[43]. The types and numbers of genes encoded at the same time culminated in a high degree of agreement among the four major categories of chloroplast gene composition. Nevertheless, whether there is any distinction in the genes with unknown functions remains to be explored further[44].

Based on the RSCU analysis of the codons, we found that the codons of *Toona sinensis* and *Toona ciliata* have an elevated A/T preference, especially the third part of coding codons having a strong A/T base preference. This finding can provide a theoretical basis for the subsequent design of expression vectors and the improvement of related protein synthesis[45].

3.2 Genetic Comparison

The IR boundary and sequence comparisons of the complete cp genes of the seven Meliaceae species have shown that the variance is dominated by noncoding regions, and these high mutation regions of noncoding regions can be used as DNA barcodes for species identification. In contrast, the protein-coding region shows remarkably conserved consistency with tRNA and rRNA, with a certain degree of agreement over the noncoding region, demonstrating the evolutionary conservatism of plants[46]. In the protein-

coding regions, the coding genes appear to be divergent, and the genes control the traits that are the key to species specificity. However, the specific functionality concerning the genes requires further study, which will also lay the initial foundation for understanding gene regulation and conducting superior breeding. Among them, *Toona ciliata* (OK572964) and its variety *Toona ciliata var. Pubescens* (MZ926838.1) have inconsistent Clp2 genes, thus providing an important breakthrough in the research into related traits between the two species. The variability among chloroplast genomes of plants is small, which is a key characteristic of chloroplast evolution[47]. The analysis of interval boundaries reveals not only the differences between *Toona ciliata* (OK572964) and its variety *Toona ciliata var. Pubescens* (MZ926838.1) along with the differences among the other four relatives, which provide strong data to support the study of molecular markers in *Toona ciliata* congeners and the whole Meliaceae. The function of the relevant genes at the junction and their regulation have yet to be studied, and this will provide useful information for breeding.

3.3 Intrageneric Variation

There are 29 and 30 repeats in *Toona sinensis* (OK572965) and *Toona ciliata* (OK572964), respectively. Most of them are forward and palindromic repeats, again located predominantly in the spacer region, the tRNA region, the intron region, and the protein coding regions. The *yfc2* gene in the protein-coding region is a hotspot for the occurrence of repetitive sequences. The SSRs in the chloroplast genomes of the two plants, 61 and 55, are single nucleotides comprised of A or T elements, except for one, which is a dinucleotide. The base composition has a strong AT preference, mainly in the spacer and intron regions. The *yfc3* gene in the intron region and the *yfc1* gene in the CDs are the hotspot regions where SSRs occur. These SSRs are one of the important tools that will later be applied to the analysis and development of molecular marker[48].

3.4 Systematic Evolution

Chloroplast DNA successions were fundamental in deciding developmental relationshi, for model ,utized the chloroplast genomes of 39 species to assess the phylogenetics and relationship inside the *Toona*. For the analysis of the synergistic cluster of the *Toona*, it was possible to obtain a high level of support for the entire evolutionary system, with little divergence in the system and no significant sequence divergence between species. It is clear from the results that there is high support for the entire evolutionary system. Inside the Meliaceae, the connection between Subfam. melioidea and Subfam. swietenioideae is more positive and preferred upheld over that of Subfam. cedreloideae, which is steady with the aftereffects of the morphological grouping[49]. In the overall, development trees indicate that little systematic differentiation, no significant sequence differentiation between species, and clustering results consistent with the morphological classification, which adequately support the stability of *Toona sinensis* and *Toona ciliata* during the process of evolution. The evolutionary tree revealed that the minor branches clustered by *K. ivorensis* (MZ274124.1), *K. anthotheca* (MZ274122.1), *K. senegalensis* (MZ274125.1) and *K. grandifoliola* (MZ274123.1) in *Khaya* showed a greater divergence from the minor branches clustered by *K. senegalensis* (MZ274125.1) and K. *grandifoliola* (MZ274123.1), with only 66% support.

Evolutionary tree which provides favorable data for the study of speciation and variation in this taxon,moreover, it assists with advancing the investigation of ordered and morphological hereditary variety. However, limited by the number of reported tree species, many affinities are not clear, and the position of many plants is uncertain. In parallel the phylogenetic tree of Toona stays hard to survey because of intricate transformative issues like joined advancement, broad infiltration, hybridization and inadequate genealogical arrangement in Toona[50]. Differential examples of plastid quality locales diverge from atomic information, as recently revealed[51]. Thusly, it is important to develop phylogenetic trees in light of integral data of nDNA and cpDNA to comprehend their phylogenetic connections completely.Further refinement of the study of evolutionary relationships within sections is required to clarify the evolutionary relationships within the families.Afterwards,to take advantage of the power of whole cp genome phylogenies for differentiation at lower taxonomic levels, additional complete cpDNA sequences of Toona members are required.

Conclusion

We further refined the relevant genetic material of the target plants in this series of studies by obtaining the complete chloroplast gene sequences of *Toona sinensis* (OK572965) and *Toona ciliata* (OK572964) by high-throughput sequencing techniques. In addition, comparison of the basic characteristics of the two species, construction of gene maps, partial base sets and comparison of the GC content of each region showed minimal differentiation. Comparative analysis of their IR boundaries, repetitive sequences and SSRs provided variability at the genetic level of the tree species and provided data to support the development of molecular markers. The construction of a phylogenetic tree of the close relatives of *Toona aromatica* aimed to explore the phylogenetic relationships of Meliaceae and Subfam. In contrast, the construction of a phylogenetic tree of relatives of the genus *Toona* aimed to explore the phylogenetic relationships of *Toona* in Subfam. Cedreloideae in which it is located, with the Subfam. Swietenioideae and Subfam. Melioideae in the mahogany family and to provide a basis for understanding the species evolution and taxonomy.

Together, these data and analyses will be of great significance for the future development of the species *Toona sinensis* and *Toona ciliata* and for conservation of the resources and biobank of the endangered species *Toona ciliata*. At present, only a very small proportion of the species has been sequenced, and more investigations of the chloroplast genomes of congeneric species are anticipated to provide a more comprehensive understanding of the basic characteristics and evolutionary patterns of the chloroplast genomes of *Toona sinensis* and *Toona ciliata*, as well as better development and utilization of excellent seed sources.

References

1. Zhou Xiangyu. Study on Chinese Toona [D]. Nanjing Forestry University, 2005.

2. Tang Qile, Wei Dongyang. Toona sinensis germplasm resources conservation and utilization [J]. Green Science and Technology, 2021, 23(05): 81-82. DOI: 10.16663/j.cnki.lskj.2021.05.029.
3. Liu Jun, Chen Yitai, Jiang Jingmin, Hua Chaohui, Zhang Haiyan, Wang Hairong, Cao Yongkang. Toona germplasm resources and their development and utilization[J]. Forestry Practical Technology, 2010(05):56-57.DOI:10.13456/j. cnki.lykt.2010.05.018.
4. J. Robert Sealy,L. H. Bailev. Manual of Cultivated Plants[J]. Kew Bulletin,1950,5(3):
5. E. H. Wilson 1914 Plant Wilson 2:157.
6. Yang Chaochen, Zhang Ting, Zhu Tianran, Cao Lixian, Li Jianan.Establishment of a sterile seedling regeneration system for Chinese toon seeds[J].Economic Forest Research,2018,36(03):114-119.DOI:10.14067/j.cnki.1003- 8981.2018.03.018.zgx
7. Liu Wenjian, Jin Jianer, Li Yanjie, Tan Zifeng, Jiang Jingmin, Liu Jun.Effects of water and nitrogen and phosphorus ratio fertilization on the growth, nutrient distribution and chlorophyll fluorescence characteristics of Toona sinensis seedlings[J].Western Forestry Science,2021,50(06):83- 90:102.
8. Liu Jun, Chen Yitai, He Guiping, etc. Study on the genetic variation of shape at seedling stage of progeny of A. Journal of Jiangxi Agricultural University, 2008, 30(1): 64-67.
9. Liu Jun, Chen Yitai, Luo Yangfu, et al. Research on community structure characteristics of A. chinensis natural forest [J]. Forestry Science Research, 2010, 23(1): 93-97.
10. Zhang Yadong, Zhong Yi, Zhou Guoqing, etc. Preliminary report on seedling breeding experiments of different families of Toona sinensis from Enshi provenance in Hubei [J]. Hubei Forestry Science and Technology, 2013, (181): 17-20.
11. Wu Lili, Wang Mingfeng, Chen Bolin. Biological characteristics and artificial cultivation of red toon tree [J]. Anhui Agricultural Science, 2006,12(7):168-169.
12. Dai Chongwen, Liu Yamin, Liu Yumin, Li Junan, Chen Yinhao, Lv Huilin. Effects of Three Exogenous Substances on Physiological Characteristics of Toona sinensis under drought stress [J]. Journal of Southwest University (Natural Science Edition), 2022, 44(02):48-56.
13. Chen Yitang. 15-year-old Toona sinensis plantation growth response to tending thinning [J]. Fujian Forestry Science and Technology, 2021,48(04):31-34.
14. Zhou Chanyuan, Ruan Jinghua, Huang Jian, etc. Research progress on chemical constituents and biological activities of Toona sinensis J Chinese Patent Medicine, 2020, 42(5): 1279-1291.
15. Guan Peiyan, Liu Meiling, Zhao Mingyu, et al. Extraction and quantification of flavonoids in young and old leaves of Toona sinensis and analysis of their antioxidant functions [J]. China Food Additives, 2020, 31(11) : 50-54.
16. Chen Rong, 1937, Chinese Tree Taxonomy, Chinese Agricultural Society.
17. Hou Kuanzhao, Chen Dezhao, 1955, Chinese Neem Branch, Acta Botany Taxonomy 4(1):39-42.
18. Chen Bangyu, 1995, Geographical distribution of Meliaceae, Journal of Tropical and Subtropical Botany 3(3):12-16.

19. Zeng Cangjiang, 1962, Correction of scientific names of newly planted plants in Fujian and several plants, Journal of Xiamen University 9(4):303.
20. Editorial Committee of Flora of China, Chinese Academy of Sciences, 1997, Flora of China (Volume 43, Volume III), Science Press.
21. Dyall SD, Brown MT, Johnson PJ. Ancient invasions: from endo-symbionts to organelles J .Science,2004,304:253-257.
22. Li Qiaoli, Yan Na, Song Qiong, et al. Analysis of the chloroplast genome sequence and characteristics of Rusang [J]. Chinese Journal of Botany, 2018, 53(1): 94
23. Yang Jiapeng, Zhu Zile, Fan Yajuan, et al. Comparative analysis of chloroplast genomes of three medicinal plants of the genus Rhododendron and their significance in species identification [J]. Chinese Journal of Pharmacy, 2020(11): 2736–2745.
24. Pfalz J,Bayraktar OA,Prikryl J,et al.Site-specific binding of a PPR protein defines and stabilizes 5'and 3'mRNA termini in chloroplasts J .EMBO J,2009,28:2042-2052.
25. Jansen RK Cairo ZQ Raubeson LA et al. Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns J . Proc Natl Acad Sci U S A,2007,104:19369-19374.
26. Lapidus,A., Antipov,D., Bankevich,A., Gurevich,A., Korobeynikov,A., Nur,S., Prjibelski A., Safonova,Y., Vasilinetc ,I., Pevzner, P. A. New Frontiers of Genome Assembly with SPA des 3.0. (poster), 2014.
27. Liu et al : CpGAVAS, an integrated web server for the annotation, visualization, analysis, and GenBank submission of completely sequenced chloroplast genome sequences. BMC Genomics 2012 13:715.
28. Automatic annotation of organellar genomes with DOGMA. Wyman SK,Jansen RK, Boore JL,Bioinformatics 2004 20 (17) : 3252-3255.
29. GREINER S LEHWARK P K BOORE J L. Automatic annotation of organellarGenomeDRAW (OGDRAW)version1.3.1 expand toolkit for the graphical visualization of organellar genome J .Nucleic Acids Research,2019,47(W1):W59-W64.
30. Rozas, J., Sánchez-DelBarrio, J. C., Messeguer, X., and Rozas, R. (2003).DnaSP, DNA polymorphism analyses by the coalescent and other methods.Bioinformatics 19, 2496–2497. doi: 10.1093/bioinformatics/btg359
31. Katoh, K., and Toh, H. (2010). Parallelization of the MAFFT multiple sequence alignment program. Bioinformatics 26, 1899–1900. doi:10.1093/bioinformatics/btq224
32. Hall, T. (2011). BioEdit: an important software for molecular biology. GERF Bull. Biosci. 2, 60–61.
33. Sharp PM, Li WH. The codon Adaptation Index–a measure of directional synonymous codon usage bias, and its potential applications. Nucleic Acids Res, 1987, 15(3): 1281-1295.
34. Hall T A. Bioedit: A User-Friendly Biological Sequence Alignment Editor and Analysis Program for Windows 95/98/NT[C]//Nucl Acids Symp Ser. 1999:95-98.

35. AMIRYOUSEFI A HYVONEN J POCZAI P. IRscope: an online program to visualize the junction sites of choroplast genomes J .Bioinformatics,2018,34(17):3030-3031.
36. Frazer KA, Pachter L, Poliakov A, Rubin EM, Dubchak I. VISTA: computational tools for comparative genomics. Nucleic Acids Res. 2004 Jul 1;32 (Web Server issue):W273-9.
37. Kurtz, Stefan and Choudhuri, Jomuna V. and Ohlebusch, Enno and Schleiermacher, Chris and Stoye, Jens and Giegerich, RobertREPuter: The Manifold Applications of Repeat Analysis on a Genomic Scale, Nucleic Acids Res., Nucleic Acids Res., 29(22):4633-4642, 2001., 2001.
38. Alexandros Stamatakis; RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models, Bioinformatics, Volume 22, Issue 21, 1 November 2006, Pages 2688–2690.
39. Zhang Min, Yin Yanpeng, Zhou Luojing, et al. A comparative study of the chloroplast genomes of three Magnolia officinalis [J]. Guangxi Plants, 2021, 45(3): 1134.
40. WANG W B YU H,et al.The complete chloroplast genome sequnces of the medicinal plant Forsythia suspensa (Oleaceae)[J].International Journal of Molecular Sciences,2017 18 11 2288.
41. DONG W L,WANG R N,ZHANG N Y,et al. Molecular evolution of chloroplast genome of orchid species insights into phylogenetic relationship and adaptive evolution[J].International Journal of Molecular Sciences,2018,19(3):719.
42. Chinese Academy of Sciences Editorial Committee of Flora of China. Flora of China [M]. Beijing: Science Press, 1993,196 (1): 40. 1940.
43. Zhang Jingwen. Whole genome and phylogeny of the endangered plant lilac pinnae chloroplast[D]. Yangling, Shaanxi: Northwest A&F University, 2019.
44. Li Jia, Zhao Mengyao, Zhang Zhaoxia, et al. Comparison and phylogenetic analysis of chloroplast genomes of Fir [Q]. Biochemical Engineering, 2021, 943(2): 6-7.
45. Liu Hailin. Whole genome sequencing and bioinformatics analysis of Ginkgo biloba [D]. Nanjing: Two Things in Nanjing, 2018.
46. Yang Jiapeng, Zhu Zile, Fan Yajuan, et al. Comparative analysis of chloroplast genomes of three medicinal plants of the genus Rhododendron and its significance in species identification [J]. Chinese Journal of Pharmacy, 2020, 55(11): 2736-2745.
47. Kode V, Mudd E A, Iamtham S, et al. 2005. The tobacco plastid acc D gene is essential and is required for leaf development. The Plant Journal, 44(2): 237-244.
48. Bock D.G., Kane, N.C., Ebert, D.P., Genome skimming reveals the origin of the Jerusalem artichoke tuber crop species: Neither from Jerusalem nor an artichoke J .New Phytologist, 2014,201: 1021-1030.
49. Chen Qianqian, Rong Liyuan, Shao Zijun, Liu Tian, Wei Lei, Song Zhenqiao.Using SRAP and EST-SSR to analyze the genetic diversity of Chinese toonresources [J]. Journal of Horticulture, 2018, 45(05):967-976.DOI:10.16420/j.issn.0513-353x.2018-0044.

50. LIU Minhao, XIAO Xingcui, ZHANG Chunhua, HE Zhiran, HUANG Zhen, LIJiaman. Genetic diversity analysis of the main provenances of Toona sinensisin Sichuan based on SSR[J/OL]. Journal of Central South University of Forestry and Technology: 1-8[2022-05-23].http://kns.cnki.net/kcms/detail/43.1470.S.20220506.1610.004.html.
51. Yu Pengfei, Sun Xiaojian, Li Chenchen, Zhang Xu, Zheng Wei, Liu Changjin.Karyotype analysis of Toona sinensis and development of SSR moleculermarkers [J]. Journal of Horticulture, 2019,46(06):1172-1182.DOI:10.16420/j.issn.0513-353x.2018-0903.

Figures

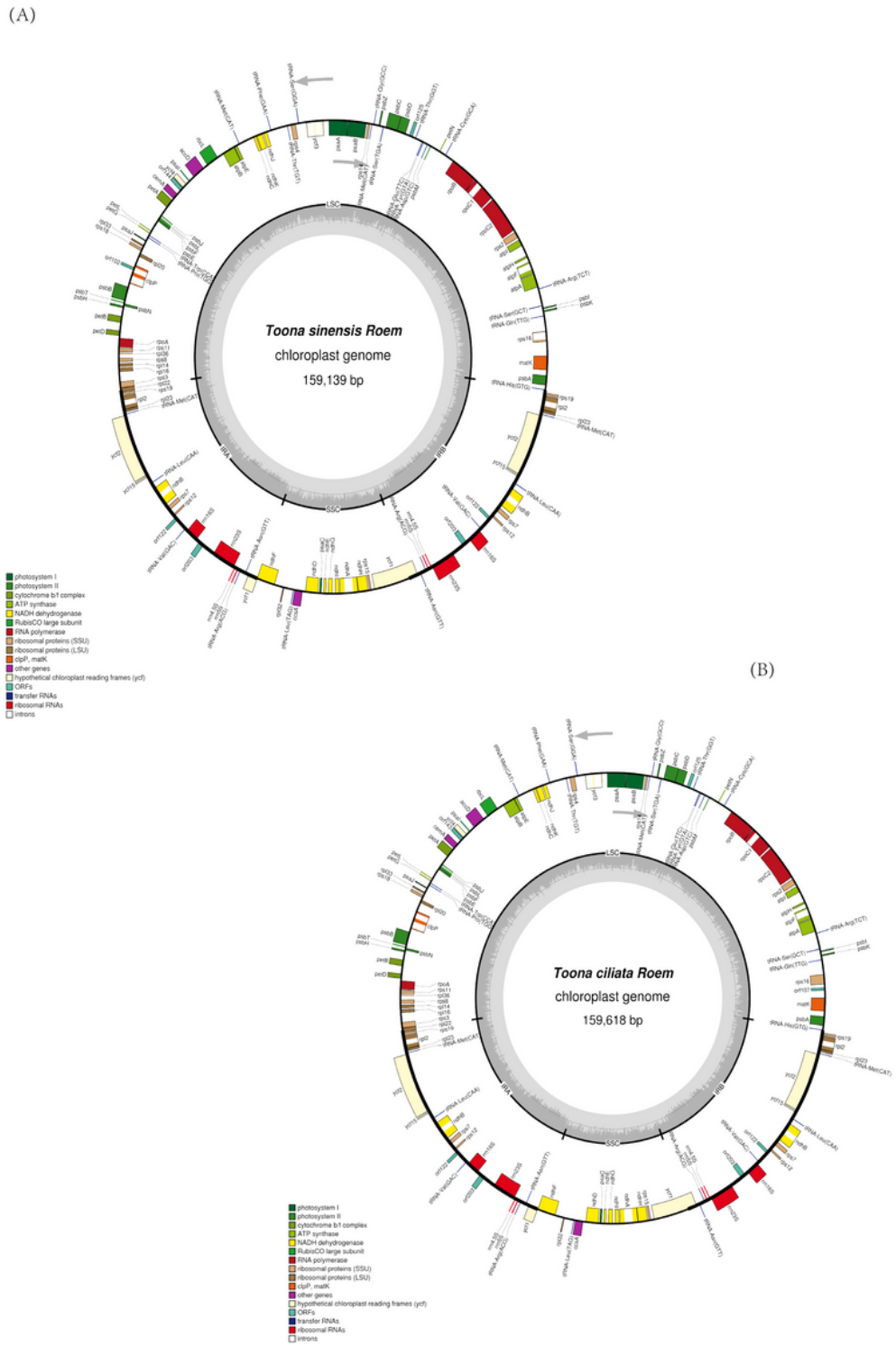

Figure 1

A map of the chloroplast genome of the Toona. (A) *Toona sinensis*,(B) *Toona ciliate*.

Genes on the outside of the circles indicate reverse transcription and those on the inside of the circles are transcribed clockwise. The thick black line on the outer circle represents the two IR regions. The dark grey graph in the inner nucleus represents GC content.

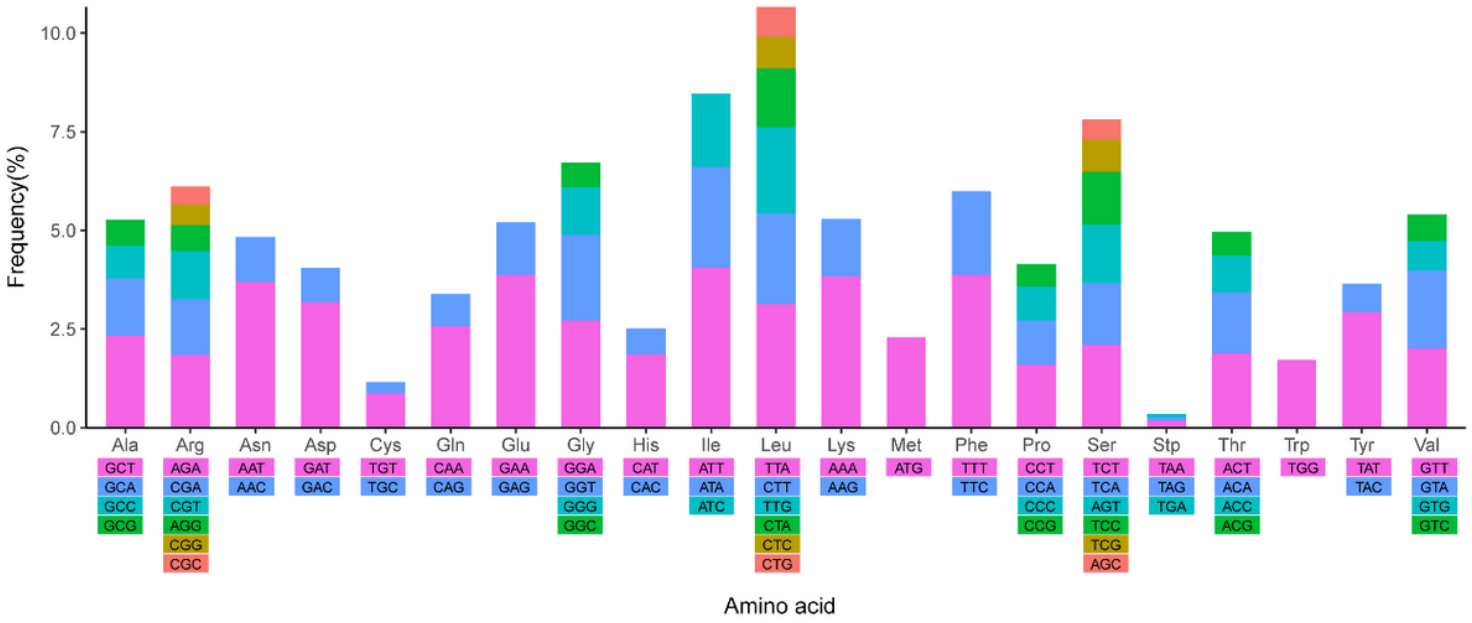

Figure 2

Histogram of codon usage frequency in the chloroplast genome of *Toona sinensis*.

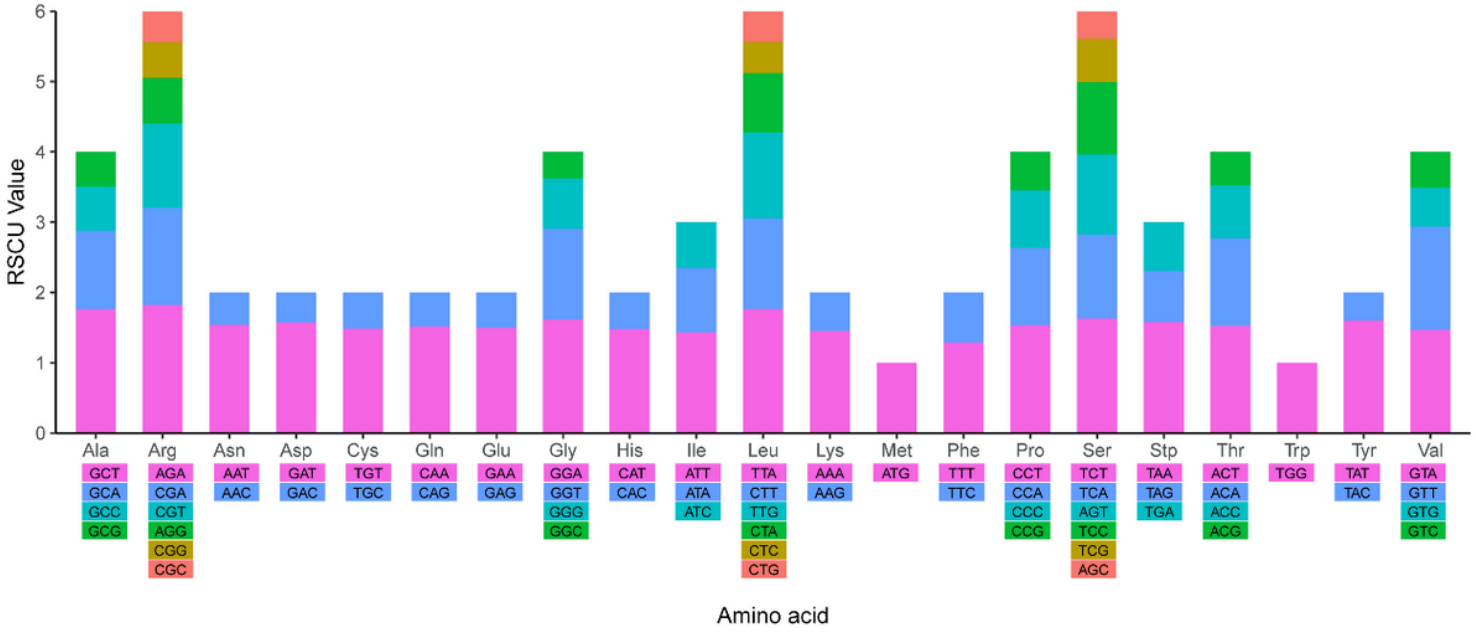

Figure 3

Histogram of codon usage frequency in the chloroplast genome of *Toona ciliata*.

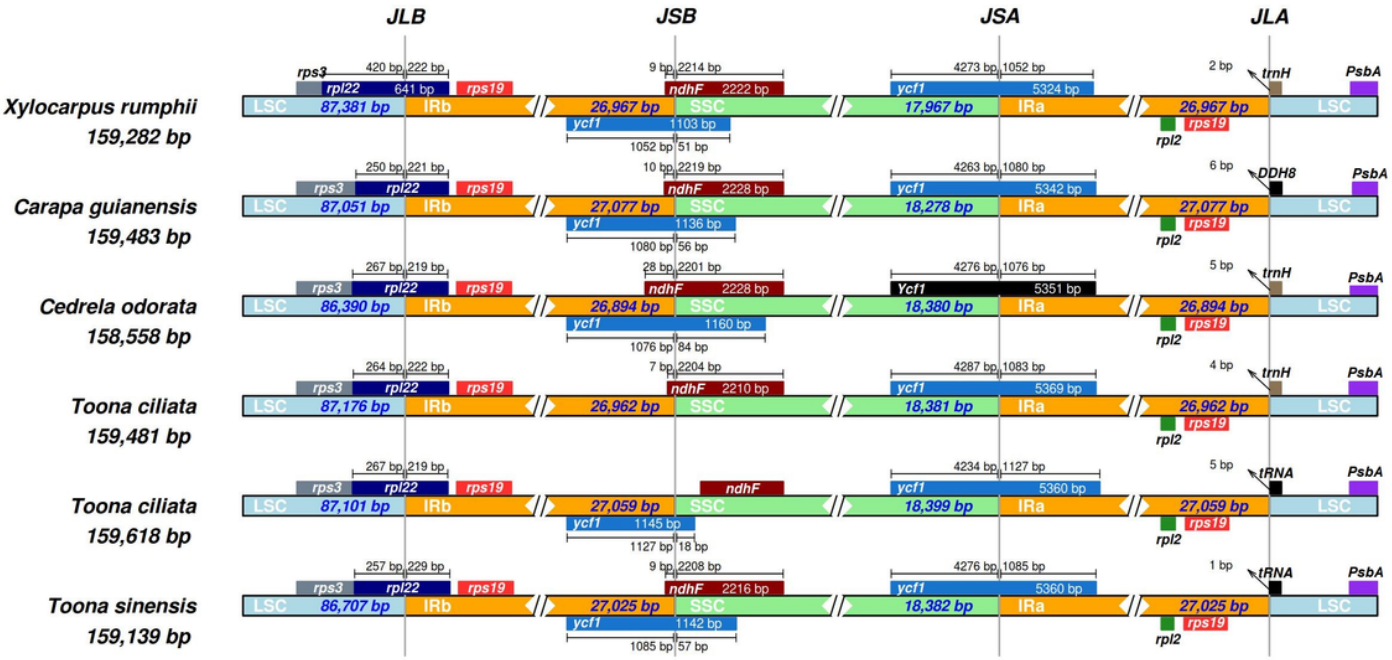

Figure 4

Boubdary comparison of IR and SC region of chloroplast genomes in 9 species

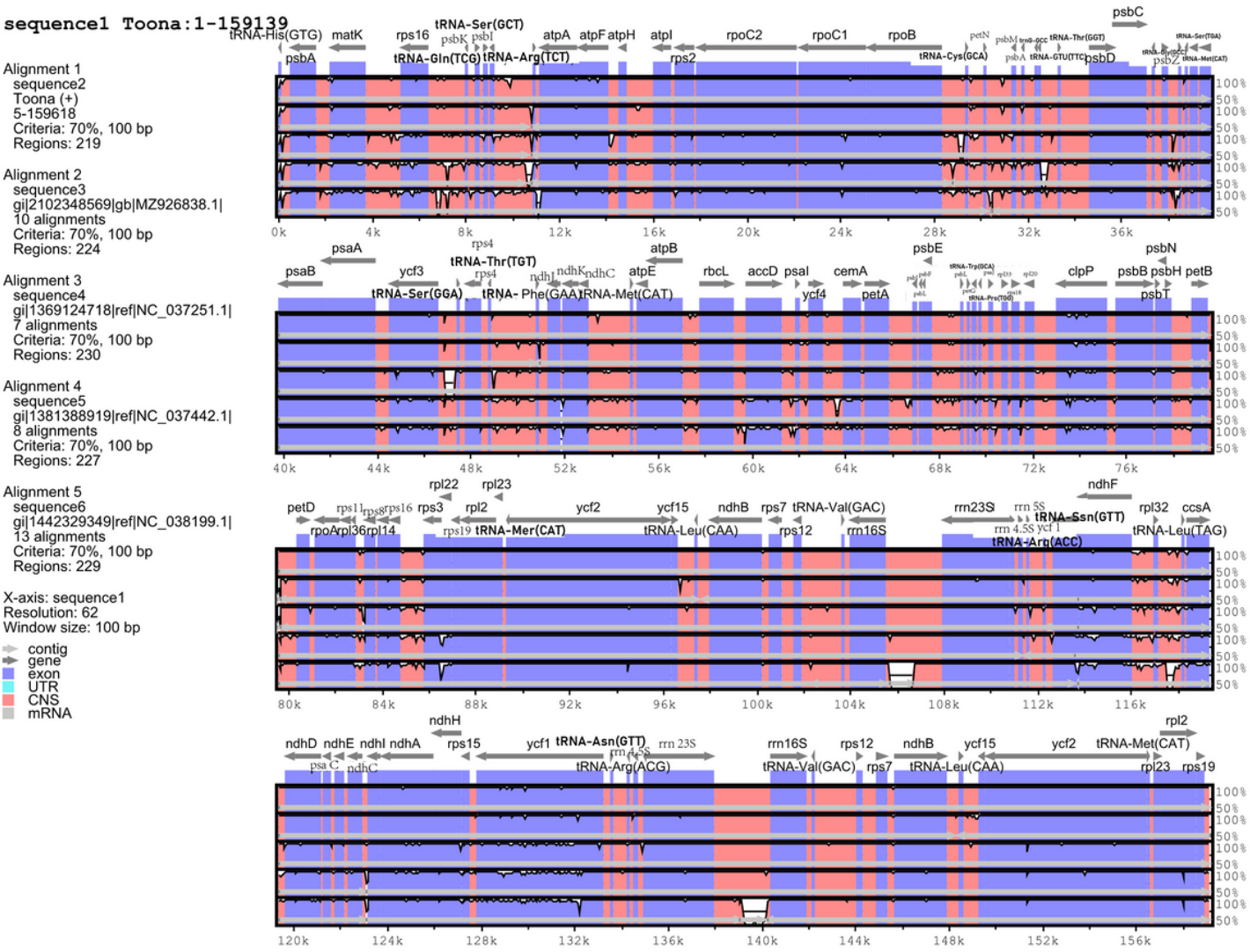

Figure 5

Alignment of chloroplast genome sequences in 6 species of Meliaceae.

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 the genome regions. Blue: Regions coding for proteins; Pink: Regions that are non-coding; Light blue: Regions coding for tRNAs and rRNAs.

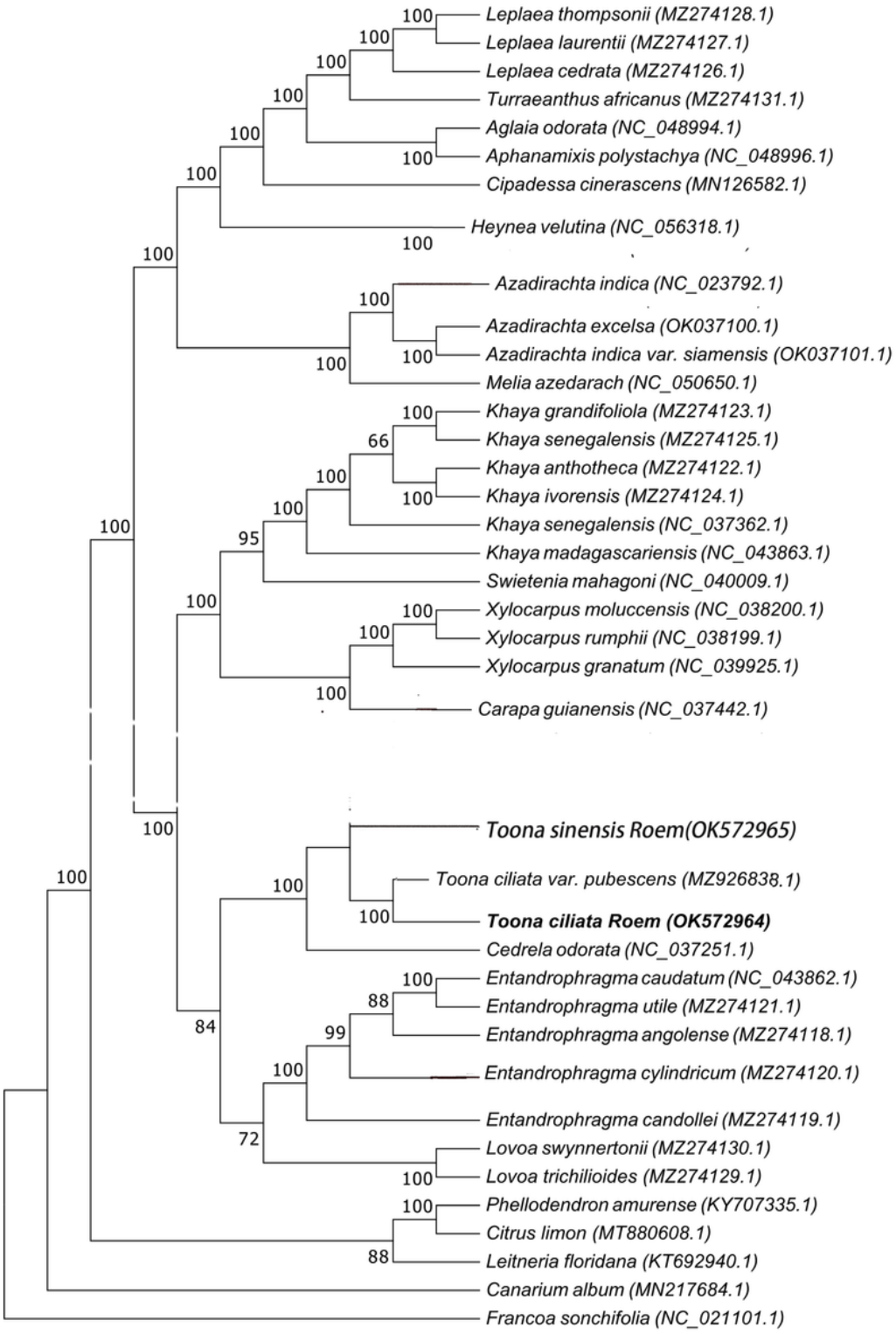

Figure 6

Phylogenetic tree based on 39 species related of the *Toona*

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- AppendicesD.xls
- AppendicesA.xls
- AppendicesB.xlsx
- AppendicesC.xlsx
- AppendicesE.xlsx
- AppendicesF.xls
- AppendicesG.xls
- AppendicesH.xls
- AppendicesI.xls
- AppendicesJ.xls
- featuretablefix.txt
- taxonomy.xls