Heliyon 8 (2022) e12557

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

Comparative phylogenetic analysis of oolong tea (*Phoenix Dancong* tea) using complete chloroplast genome sequences

Yaqun Liu [1], Liyun Lin [1], Dongjuan Yang, Xianghui Zou, Zhenxia Zhang, Mouquan Liu, Min Lin [*], Yuzhong Zheng [**]

School of Life Science and Food Engineering, Hanshan Normal University, Chaozhou, China

ARTICLE INFO

Keywords:
Phoenix dancong tea
Oolong tea
Chloroplast genome
Genomic structure
Repeat analysis
Phylogeny

ABSTRACT

Phoenix Dancong tea, a variety of oolong tea, is produced in Chaozhou, Guangdong Province, China, and is characterized by numerous hybridizations and polyploidization. To assess the genetic diversity and phylogenetic relationships among *Phoenix Dancong* tea and other oolong teas, an integrated circular chloroplast genome was constructed for thirty species of *Phoenix Dancong* tea from Chaozhou. The genome of *Phoenix dancong* tea is a circular molecule of 157,041–157,137 bp, with a pair of inverted repeats (26,072-26,610 bp each) separated by a large single copy (86,615-86,658 bp) and small single copy (18,264-18,284 bp). A total of 135 unique genes were encoded, including 90 protein coding genes, 37 tRNAs and 8 rRNAs. A comparative analysis with the other seven species in the oolong tea family that have been sequenced to date revealed similarities in structural organization, gene content and arrangement. Repeated sequence analysis identified 17–23 tandem repeats, 20–24 forward repeats and 25–27 palindromic repeats. Additionally, a total of 65–70 simple sequence repeats were detected, with mononucleotide repeats being the most common. Phylogenetic analyses showed that *Phoenix Dancong* tea and Fujian oolong tea were clustered with other cultivated *Camellia sinensis* in the genus *Camellia* of the family *Theaceae*, while the two oolong tea species were relatively independently cross-embedded in the genus, *Camellia*. Close genetic relationships were observed between *Phoenix Dancong* tea and other oolong teas, and the overall chloroplast genomes of oolong tea showed patterns with low variations and conserved evolution. The availability of *Phoenix Dancong* tea chloroplast genomes not only elucidated the relationship among oolong teas from different origins in Guangdong and Fujian but also provided valuable genetic resources to assist further molecular studies on the taxonomic and phylogenomic resolution of the genus *Camellia*.

1. Introduction

Oolong tea, a semioxidized tea famous for its elegant fruity and floral aroma, is one of the most important traditional economic beverages and is distributed widely in southern China. It is well known for its health benefits, which can regulate intestinal microbiota and increase digestion, fat transformation and also has other benefits [1]. Traditional oolong tea originated from Beiyuan tribute tea in Fujian Province during the Song Dynasty and has a history of more than 1000 years and is mainly distributed in Fujian, Guangdong and Taiwan [2]. *Phoenix dancong* (*P. Dancong*) tea is an important oolong tea variety with a long history and reputation lasting more than 900 years in Chaozhou, Guangdong Province. It is one of the most tasty and fragrant tea varieties in China. *P. Dancong* tea is an excellent single plant selected from local *Phoenix Narcissus* tea plant varieties through long-term artificial cultivation, each with its own distinctive form and taste [3]. Strikingly, relatively little is known about the similarities and differences of the cultivars from different geographical origins. Most oolong tea processing procedures are performed in the same way, resulting in similar appearance and flavor characteristics that are difficult to distinguish from each other. In addition, the phylogenetic relationships and interspecific boundaries are difficult to determine due to similar species origins, frequent hybridization and polyploidy during the cultivation process [4]. Traditional morphological and biological classification methods based on limited feature selection, such as novel artificial sensing tools [5] and headspace solid-phase microextraction/gas chromatography–mass spectrometry

* Corresponding author.
** Corresponding author.
E-mail addresses: konfutea@hotmail.com (M. Lin), zhengyuzhong@gmail.com (Y. Zheng).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.heliyon.2022.e12557
Received 16 February 2022; Received in revised form 9 August 2022; Accepted 14 December 2022

(HS-SPME/GC–MS) [6], are inherently dynamic and unreliable owing to genetic inheritance and other biological factors. Currently, there is a lack of research on the genetic background and resources of oolong tea. The lack of appropriate sequences and polymorphic genetic markers for phylogenetic analysis has long hindered understanding its phylogeny and increased the controversy over the genetic classification.

Chloroplast genomes are extensively used in plant phylogeny, molecular evolution and population genetics studies [7]. Chloroplasts are one of the most important semiautonomous organelles in green plants. In addition to serving as sites of photosynthesis, chloroplasts also perform carbohydrate biosynthesis and other functions [8]. The chloroplast genome is relatively independent from the nuclear genome and is more conserved in its structure and organization [9]. In general, the chloroplast genome is a circular DNA molecule consisting of 120–160 kb, which encodes 120–130 protein coding genes (CDS), ribosomal RNA (rRNA) and transfer RNA (tRNA) associated with photosynthesis or other gene expressions [10]. It has a typical quadripartite structure comprising four major segments: a large single copy region (LSC, 80–90 kb), a small single copy region (SSC, 16–27 kb) and two inverted repeat regions (IRs, IRa and IRb, 20–28 kb) [11]. The chloroplast genomes of several oolong tea species have recently been published in GenBank (http://www.ncbi.nlm.nih.gov/genbank/), including Baijiguan (Accession number: MT773373), Bantianyao (Accession number: MW046255), Dahongpao (Accession number: MT773374), Rougui (Accession number: MT773375), Shuijingui (Accession number: MT773376), Tieluohan (Accession number: MT773377), and Wuyi Narcissus (Accession number: MT612435), all of which belong to Fujian oolong tea. However, most of these studies have focused on the genome compositions and phylogenetic relationships of other tea species. As a globally famous beverage crop, oolong tea has been studied in great depth in terms of its components and efficacy, but its genetic and evolutionary history deserves further exploration [2]. Unfortunately, to date, studies on the chloroplast genome of oolong tea have only been reported for those varieties cultivated in Fujian Province. Although they demonstrated the power of complete chloroplast genomes, their related species were not included. The genetic relationships among oolong teas from different origins have not been fully elucidated.

To investigate the relationship between the structural characteristics and phylogeny of the chloroplast genome in oolong teas, another important representative of oolong tea cultivated in Guangdong Province (also known as *P. Dancong* tea) was selected for this study. Here, we selected thirty excellent, representative species of *P. Dancong* tea and performed sequencing, assembly and characterization of the chloroplast genomes. Their features and structures were compared with those of Fujian oolong tea to infer the phylogenetic relationships and evolutionary mechanisms arising from the regional and varietal differences in oolong teas. Furthermore, these data will provide sufficient genomic resources for further research and application of tea and contribute to its complex evolutionary history and uncertain phylogeny.

2. Materials and methods

2.1. Materials and DNA extraction

Fresh young leaves of 30 *P. Dancong* tea varieties, namely, Youhuaxiang, Dongfanghong, Laoxianweng, Zongsuoxie, Zhilanxiang, Songzhong, Huangjinye, Jiangmuxiang, Baxian, Potou, Jiangxiang, Zhuye, Rouguixiang, Xiongdizai, Jilongkan, Juduozai, Dawuye, Leikouchai, Nanjiangxiang, Chengmen, Tuofuhou, Wuye, Lanrenxiang, Guihuaxiang, Milanxiang, Yashixiang, Yehuadancong, Shuixian, Hongyin and Baiyin, were used in this study and were collected from Chaozhou City, Guangdong Province, China. All *P. Dancong* teas used in the present study were screened from the *Phoenix narcissus* variety by generations of tea farmers. The research samples were identified as different varieties of *P. Dancong* tea by Dr. Shoujun Guo and Dongjuan Yang from the School of Life Science and Food Engineering, Hanshan Normal University. The other samples outside the experiment were stored in the herbarium of Hanshan Normal University. Total genomic DNA was extracted from young leaves using the cetyltrimethylammonium bromide (CTAB) method [12]. The extracted DNA quality was assessed by 1.0% (wt/v) agarose gel electrophoresis and a NanoDrop spectrophotometer (Thermo Scientific, USA).

2.2. Illumina sequencing, assembly, and annotation

Purified DNA was used to generate an insert size of 300 bp using a TruSeq DNA Sample Prep Kit (Vanzyme, China), and next-generation sequencing was conducted on an Illumina HiSeq 2500 sequencer (Illumina, USA) according to the standard Illumina library preparation procedure. Approximately 8,205,826 (Wuye) to 22,120,726 (Juduozai) raw reads were obtained with paired-end (PE) 150-bp length reads. To guarantee better results for the following mapping and assembly, the raw reads were evaluated using FastQC (version 0.11.8, http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and trimmed using Fastp (version 0.19.5, https://github.com/OpenGene/fastp) to filter low-quality reads and remove adapters with a qualified quality phred cutoff of 20, unqualified percent limit of 20 and required length of 60 [13]. The clean high-quality reads were assembled using MetaSPAdes (version v3.13.0) with default parameters [14]. Here, the published chloroplast genome of Rougui (Accession number: MT773375) was selected as the reference genome. The *P. Dancong* tea chloroplast genome was annotated using CpGAVAS2 (http://47.96.249.172:16019/analyzer/home) [15]. The annotated circular chloroplast genome of *P. Dancong* tea was mapped and visualized using Chloroplot online software (https://irscope.shinyapps.io/chloroplot/) [16]. The codon usage, relative synonymous codon usage (RSCU) values and GC contents were calculated by MEGA7 (http://www.megasoftware.net) [17].

2.3. Genome comparison

The chloroplast genomes of Fujian oolong tea varieties (namely, Baijiguan, Bantianyao, Dahongpao, Rougui, Shuijingui, Tieluohan and Wuyi Narcissus) were used for comparisons with the *P. Dancong* tea genomes obtained in this study, which were downloaded from the GenBank database. Multiple alignments of these 37 oolong tea varieties were carried out by MAFFT Version 7 (https://mafft.cbrc.jp/alignment/software/) using the auto mode option and were adjusted manually where necessary [18]. Sequential comparisons of chloroplast genomes with sequence annotation information were visualized using the mVISTA (http://genome.lbl.gov/vista/mvista/submit.shtml) program in Shuffle-LAGAN mode [19]. Boundaries between the IR and SC regions and gene order rearrangements of the oolong tea species were compared using the online program, IRscope (https://irscope.shinyapps.io/irapp/) [20]. Nucleotide diversity (Pi) was calculated by sliding window analysis in DnaSP v5.0 with a step size of 200 bp and sliding window of 600 bp [21].

2.4. Analysis of long repeats and single sequence repeats

REPuter software (https://bibiserv.cebitec.uni-bielefeld.de/reputer) [22] was used to identify the sizes and locations of forward and palindrome repeats with a Hamming distance of 3, minimal repeat size of 30 bp and maximum computed repeats of 90 bp. Tandem Repeats Finder software (https://tandem.bu.edu/trf/trf.html) [23] was used to identify tandem repeats with default parameters. Simple sequence repeats (SSRs) were detected using MIcroSAtellite software (MISA, https://webblast.ipk-gatersleben.de/misa/) [24] with thresholds of ten repeat units for mononucleotide SSRs, five repeat units for dinucleotide SSRs, four repeat units for trinucleotide SSRs, and three repeat units for tetra-, penta-, and hexanucleotide SSRs. The long repetitive sequences and single sequence repetitive sequences in oolong tea were clustered and analyzed. Samples

containing the gene regions were labeled "Gene", and those without were labeled "No Gene", and these datasets were analyzed by hierarchical clustering using correlation as a similarity measure. A clustering graph was generated using the OmicStudio tools (https://www.omicstudio.cn/) [25].

2.5. Phylogenetic analysis

The chloroplast genome sequences of 37 oolong tea varieties, together with those of 69 *Theaceae* species, were aligned using the MAFFT program version 7. Chloroplast genome sequences were obtained from the GenBank database (Supplementary Table S1). Gaps in alignments were stripped. Phylogenetic trees were constructed using maximum likelihood (ML) methods implemented using RAxML version 8 under the general time reversible (GTR) model of nucleotide substitution and the gamma (Γ) distribution of rate heterogeneity [26, 27]. *Coffea arabica* (Accession number: NC_008535) and *Arabidopsis thaliana* (Accession number: NC_000932) were used as outgroups [28, 29, 30]. The phylogenetic trees were visualized with the Interactive Tree of Life v6 (iTOL, https://itol.embl.de/) [31].

3. Results

3.1. Chloroplast genome features of P. Dancong tea

The complete chloroplast genomes of the thirty *P. Dancong* teas ranged from 157,041 bp (Potou) to 157,137 bp (Zhuye), with differences less than 96 bp. All of these exhibited typical quadripartite structures, including large single copy regions of 86,615-86,658 bp and small single copy regions of 18,264-18,284 bp, which were separated by two equal length inverted repeat regions of 26,072-26,610 bp (Figure 1). The average overall guanine-cytosine (GC) content of these chloroplast genomes was 37.30%, while the GC contents of each region were different. The GC content of the IRs was 43.00%, which was higher than that of LSCs (35.30%) and SSCs (30.50%) (Supplementary Table S2). In the statistical analysis, the comparative differences among the sequences of the 30 species were not obvious, and both the lengths and gene contents of each region were similar.

Each chloroplast genome was found to harbor a total of 135 genes, of which 109 were unique in LSC and SSC, and 26 were duplicated in the IR regions, including 90 protein coding genes, 37 transfer RNA genes and 8

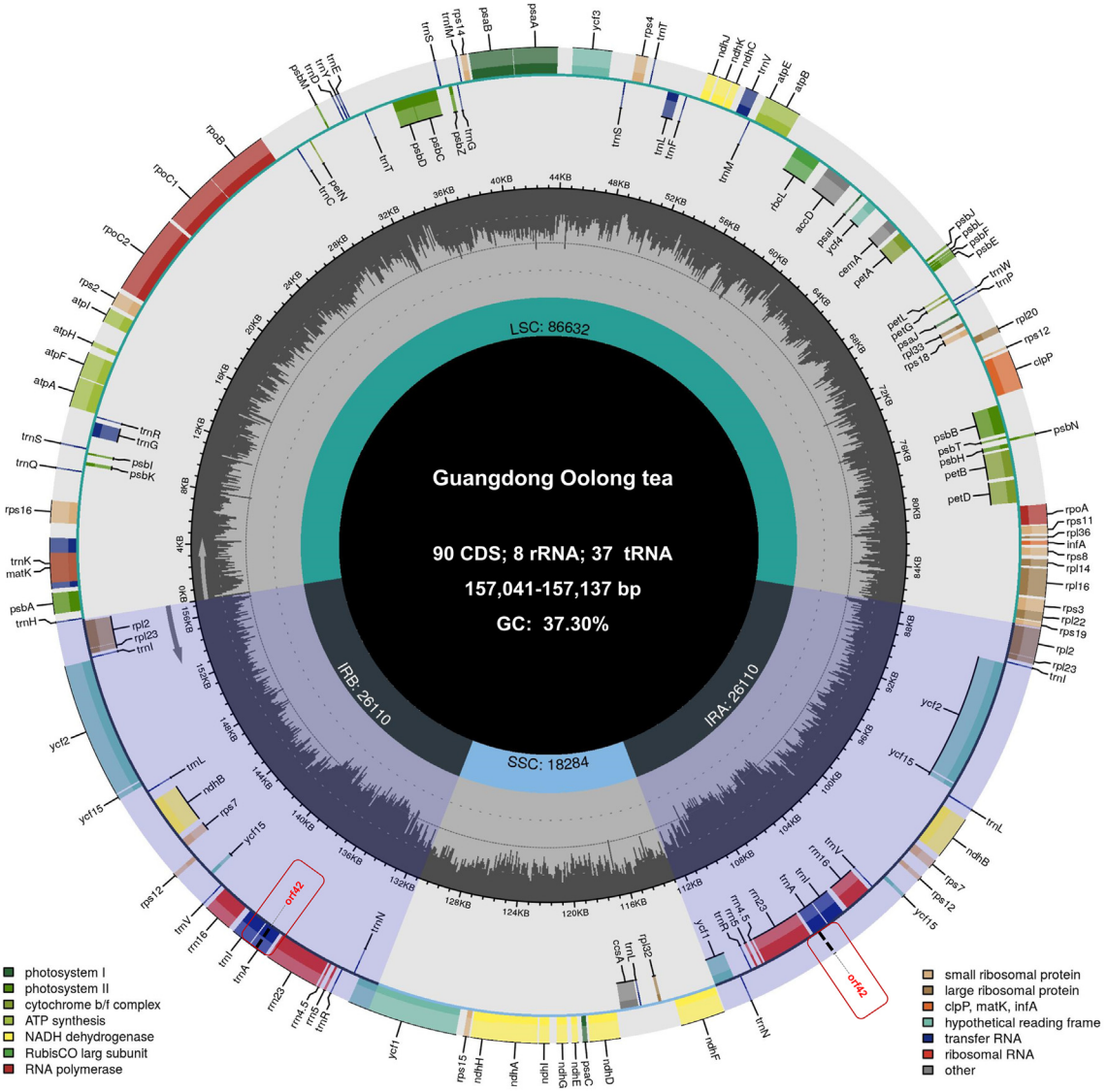

Figure 1. Complete chloroplast genome map of *P. Dancong* tea. The genes shown in the outer circle are transcribed clockwise, and those in the inner circle are transcribed counterclockwise. Genes belonging to different functional groups are color-coded. The darker gray colors in the inner circle indicate the GC content of the chloroplast genome. The red box indicates the missing gene, *orf42*.

ribosomal RNA genes. These genes belong to several categories with different functions and can be divided into self-replication (tRNA, rRNA, ribosome subunits and DNA-dependent RNA polymerase); photosynthesis (NADH oxidoreductase, photosystem subunits, rubisco, cytochrome and ATP synthase); biosynthesis (maturase, envelope membrane protein, acetyl-CoA subunits, translational initiation factor and protease); and conserved open reading frames. We identified 63 protein coding genes and 22 tRNA genes located in LSCs, 16 protein coding genes, 14 tRNA genes and 8 rRNA genes located in IRs, and 11 protein coding genes and one tRNA gene located in SSCs. Among them, three pairs of adjacent genes, *psbD/psbC*, *atpE/atpB* and *rps3/rpl22,* were observed in LSCs, with overlapping regions of 52 bp, 3 bp and 15 bp, respectively. It should be noted that the two open reading frames ORF42 in IRs were absent in *P. Dancong* tea (Figure 1).

A total of 21 genes (13 protein coding genes and 8 tRNAs) contained introns, all of which were single intron genes except for *ycf3* and *clpP* (with two introns). Although intron-containing genes are very conserved, it can be clearly seen that the differences in total gene lengths depend mainly on the differences in intron lengths, which hardly vary across genera. The intron lengths and related gene lengths in IRs were consistent with those in LSC and SSC regions (Figure 2). In addition, the 5′-end and 3′-end exons of the *rps12* gene are located in LSC and IR regions, respectively. *matK*, the gene encoding maturase, is located within the largest intron of *trnK-UUU*.

There are 26796 codons (including stop codons) in the chloroplast genome of *P. Dancong* tea, and 20 amino acids are translated. The most prevalent amino acid is leucine (L, 10.50%), followed by isoleucine (I, 8.70%) and serine (S, 7.90%), while the rarest amino acid is cysteine (C, 1.10%) among the proteins encoded by chloroplast genes. Subsequently, slight differences were found in the RSCU values in the *P. Dancong* tea chloroplast genome. It is well known that codon usage preferences are related to RSCU values [32]. RSCU values greater than 1 were found for 29 codons, of which 28 were A/T-ending codons (except TTG), suggesting that they are biased codons. Thirty codons had RSCU values less than 1, of which 28 were C/G-ending codons (except CTA and ATA). Usage of the start codons, AUG and UGG, with the latter encoding tryptophan, has no bias (RSCU = 1). CGC had the lowest RSCU value (0.335), AGA had the highest RSCU value (1.886), and both encode arginine (Figure 3). To conclude, CGC was negatively biased, and AGA was positively biased.

3.2. Comparison of chloroplast genome sequence variations in Fujian oolong tea

3.2.1. Structural variations in the chloroplast genome of oolong tea

To facilitate the subsequent phylogenetic analyses and smooth plant identification, 7 published Fujian oolong teas were selected for comparison with the 30 *P. Dancong* teas to estimate the sequence similarities of teas from different regions. The overall sequence identities of the 37 oolong tea chloroplast genomes were relatively conserved and similar, including gene order and number. They were more conserved in the IR and SSC regions than in the LSC region and varied more in the noncoding regions than in the coding regions (Figure 4). Furthermore, the nucleotide diversity, Pi, was calculated to determine the sequence divergence levels among the oolong tea chloroplast genomes. All aligned sequences demonstrated surprisingly low divergences, with all regions displaying low variations (Pi < 0.006). The Pi values ranged from 0 to 0.003 with an average of 0.000132, and these small differences were not statistically significant (data not shown). Both results indicate high sequence similarity across the chloroplast genomes, suggesting that the chloroplast genome of oolong tea was highly conserved in gene order and structure.

3.2.2. Contraction and expansion of inverted repeats

The IR region (size from 24,399 bp to 26,110 bp) of 37 oolong tea chloroplast genomes was highly conserved, but structural variations remained in the IR/SC boundary regions. To further observe potential contraction and expansion of the IR regions, we compared the gene variants in the IR/SSC and IR/LSC boundary regions. *rps19-rpl2-trnH* and

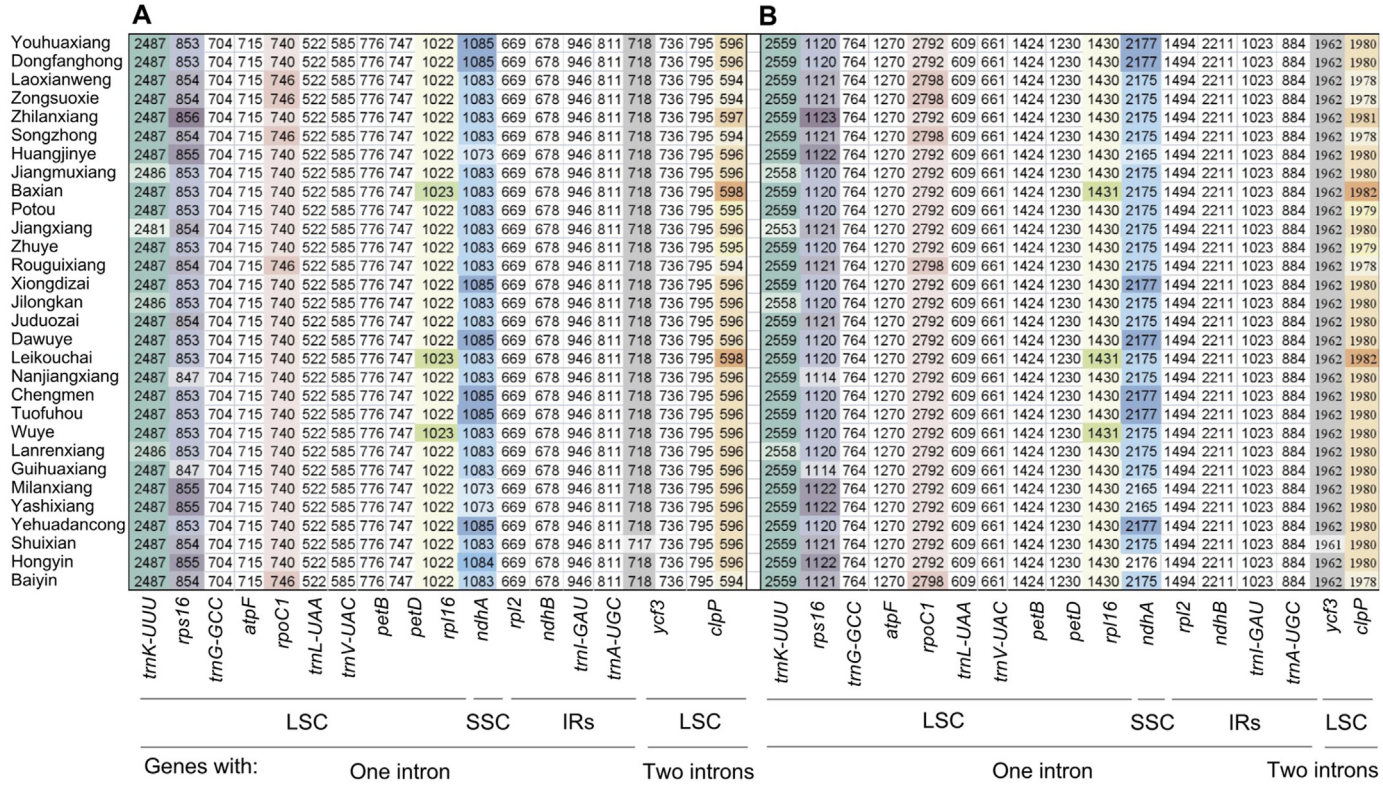

Figure 2. Statistical analysis of intron-containing genes of *P. Dancong* tea. (A) Lengths and numbers of introns. (B) Lengths and numbers of total intron-containing genes.

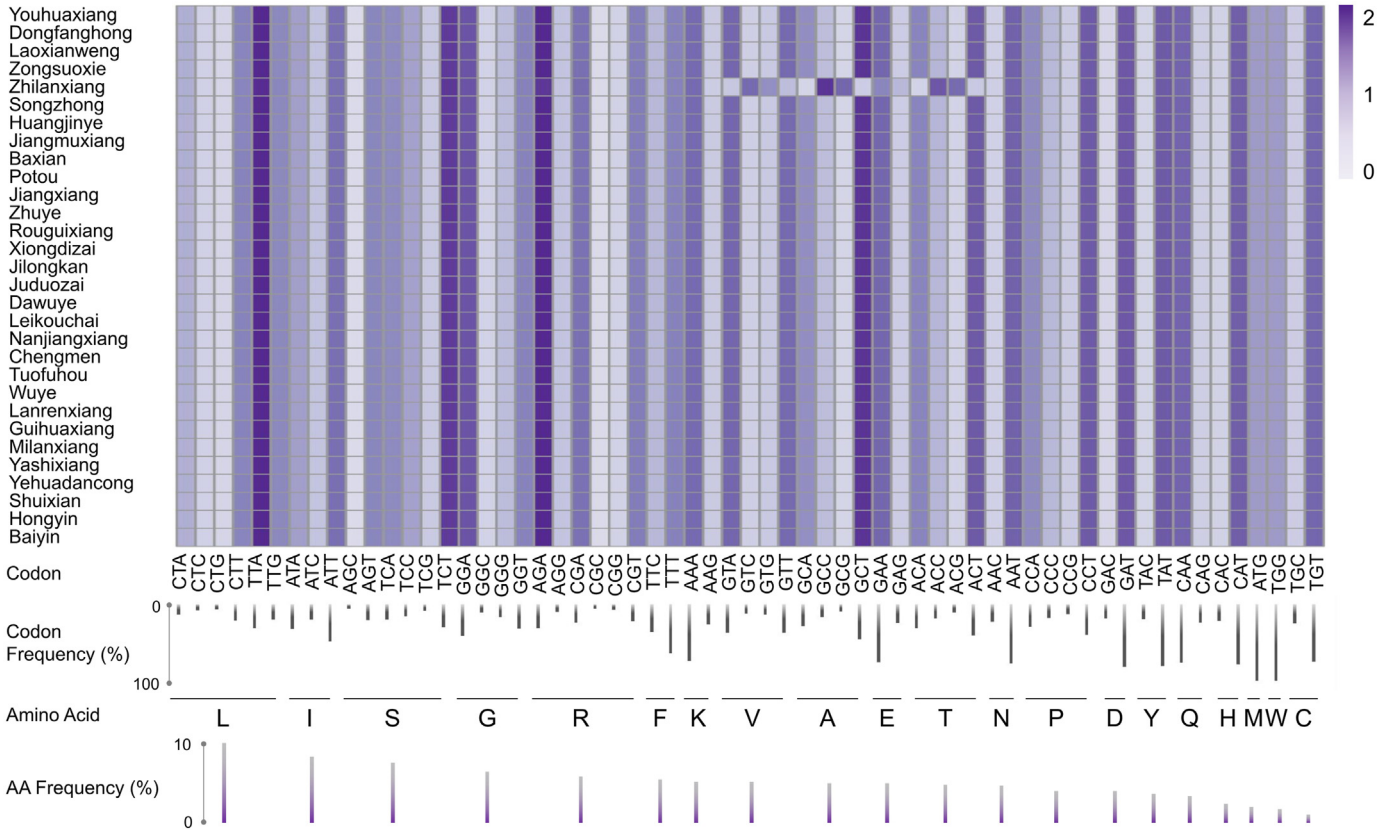

Figure 3. Heatmap of RSCU values and corresponding amino acid analysis of *P. Dancong* tea.

ycf1-ndhF were located in the J_{LB} and J_{LA} regions and J_{SB} and J_{SA} regions, respectively. The *rps19* gene crossed the LSC/IRb region with 46 bp located at the IRb region. Two copies of the *ycf1* gene crossed J_{SB} and J_{SA} with nearly 1 bp in the SSC region and 4553/1069 bp in the IRa region. The *trnH* gene in the LSC region was contracted by 1 bp from the junctional region of the IRa-LSC boundary (Fig. S1). Overall, the contraction/expansion events of the IR region detected in these chloroplast genomes were extremely stable.

3.2.3. Long repeat structure analysis

Three categories consisting of long repeats, tandem repeats (TR), forward repeats (FR) and palindromic repeats (PR), were detected in oolong tea chloroplast genomes. The numbers of the three repeated types were consistent as follows: TRs ranged from 17 (Nanjiangxiang, Chengmen, and Guihuaxiang) to 23 (Wuyi Narcissus); FRs ranged from 20 (Youhuaxiang and Dongfanghong) to 24 (Shuixian and Tieluohan); and PRs ranged from 25 (Zongsuoxie, Shuixian, and Tieluohan) to 27 (Zhilanxiang, Jiangmuxiang, and Baxian). The TR lengths were 14–25 bp, and the lengths of the other two categories were 18–56 bp (Figure 5ABC). These repeat loci were distributed in the TR sites, accounting for 17.06% of the total repeats (one site was counted), 38.26% in the FR and 44.78% in the PR sites (two sites were counted) (Figure 5D). The IR regions had the greatest number repeat sites (53.20%), followed by LSC (34.04%) and SSC (12.76%). The maximum number of repeat sites was located in intergenic spacer (IGS) regions (42.73%), and a minority were located in introns (35.88%), which were *petD* and *ndhA*. Only a few types of genes (e.g., *petG*, *rpoB*, *clpP*, *rps19*, *rpl2*, *ycf1*, and *ycf2*) possessed repetitive elements, with the *ycf2* gene containing the highest number of repeat sites (27.71%) (Figure 5EFG).

Oolong tea chloroplast genomes have a total of 32 identical long repeat sequences (9 TR, 10 FR and 13 PR). Based on the quadripartite structure and gene regions, the results also showed that the divergent regions of chloroplast genomes were associated with various repeat sequences. The distribution of TR was rather conserved, except for Wuyi Narcissus, Shuijingui and Zhilanxiang, which were clustered into three classes due to *rrn23/trnA-UGC* and *ycf15/rps7*. While both FR and PR clustered into 16 classes, only *trnQ-UUG/psbK*, *trnS-GCU/trnG-GCC*, *trnG-GCC*, *atpH/atpI*, *psbM/trnD-GUC*, *trnG-UCC*, *trnP-UGG*, *ycf2*, *ycf1* and *ndhA* (itron) were identical in FR, and the other repeat sites were different. The diversity of PR sites is mainly caused by *atpA/atpF*, *trnA-UGC/rrn23* and *ndhG/ndhI* (Figure 5). The specific data are shown in Supplementary Table S3.

3.2.4. Simple sequence repeat analysis

Simple sequence repeats (SSRs) are tandem iterations of single nucleotides or short oligonucleotides with sizes of 1–6 bp [33]. We found that 65–70 SSRs were distributed in the oolong tea chloroplast genome, with sizes ranging from 10 to 20 bp. The incidence of SSRs was nonuniform to the region size, accounting for 66.45%, 14.94% and 18.61% in LSC, SSC, and IRs, respectively. The SSRs were located in CDSs with an average of 18, IGSs with an average of 49 and introns with an average of 9 (Figure 6A).

Five kinds of SSRs were found in the chloroplast genome of oolong tea, with the number of mononucleotide repeats ranging from 0 to 20 (79.63% in LSC, 7.41% in SSC and 12.96% in IRs), followed by dinucleotide repeats and trinucleotide repeats at 4 and 1, respectively, tetranucleotide repeats ranging in number from 8 to 10 and 2 hexanucleotide repeats. However, no pentanucleotide repeats were found. The longest repeat was an 18 bp stretch of the (AAAAAG/CTTTTT)$_3$ hexanucleotide. It is noteworthy that all of the mononucleotide and dinucleotide repeats consisted of A/T. Among the trinucleotide and tetranucleotide repeats, AT and GAA occurred most frequently (Figure 6C). The distributions and clustering of SSRs of Fujian oolong tea and Chaozhou *P. Dancong* tea were more significant. There were a total of 40 identical loci for SSRs, and their distribution was rather conserved in the oolong tea chloroplast genome (Figure 6D). These

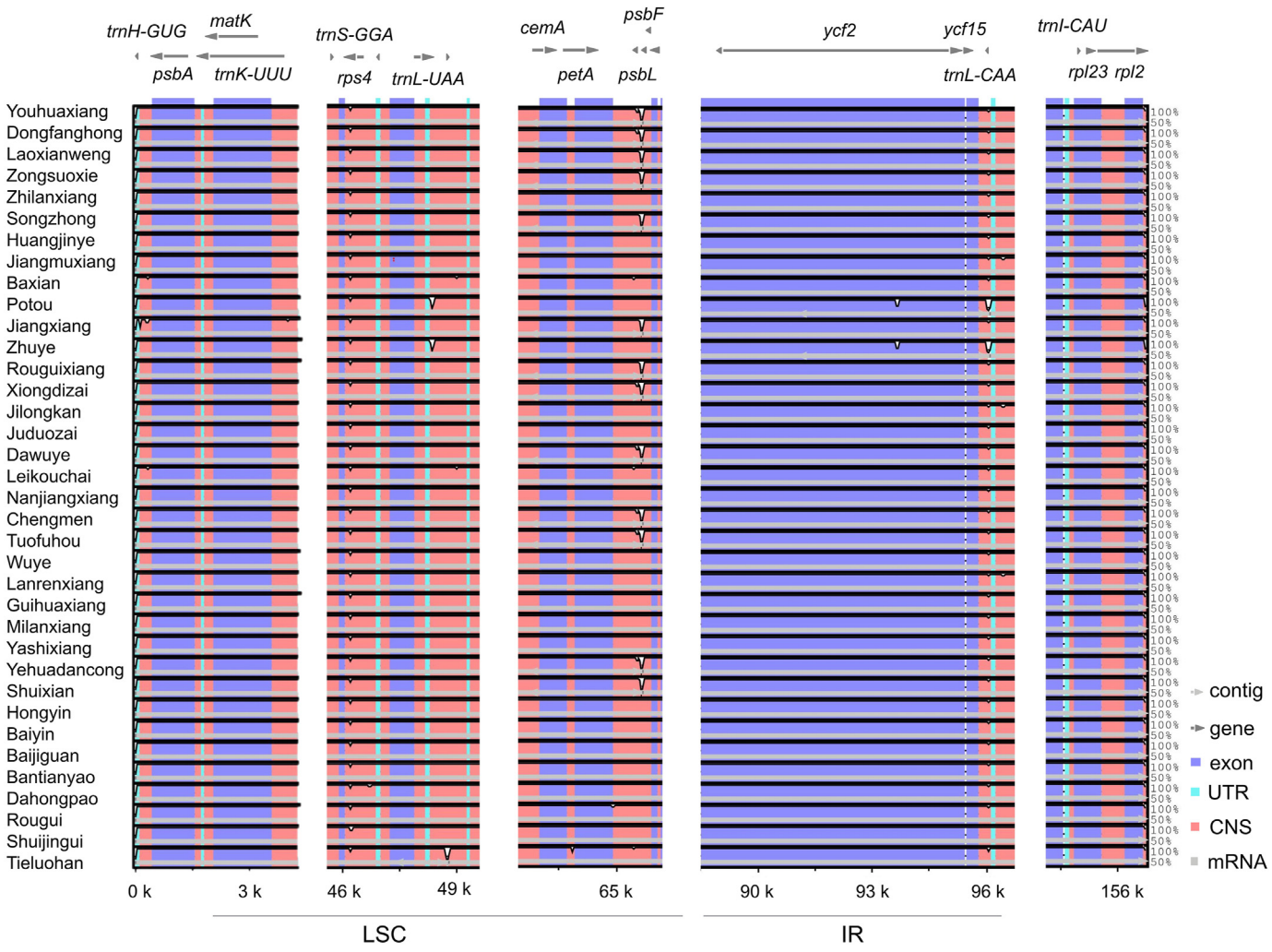

Figure 4. Visualization alignment of oolong tea chloroplast genome sequences. The mVISTA-based identity plots show the sequence identities with Wuyi Narcissus as a reference. Only the variant alignments are shown because of the large number of samples. The *x*-axis represents the coordinates in the chloroplast genome. The *y*-axis indicates the average percentage identity of sequence similarity in the aligned regions, ranging between 50% and 100%. Genome regions are color-coded as protein coding, rRNA coding, tRNA coding or conserved noncoding sequences (CNS).

unique repeats were mainly found in *ndhK/ndhC*, *psaI/ycf4*, *trnP-UGG/psaJ* (IGS), *clpP*, *rps8* (CDS) and *ndhA* (itron), which are located in the LSC and SSC regions (Figure 6B). The specific data are shown in Supplementary Table S4.

3.3. Phylogenomic analysis

Chloroplast genome sequences are valuable genomic resources for elucidating evolutionary histories and have been widely applied in phylogenetic studies [25]. To determine the phylogenetic position of oolong tea, 76 complete chloroplast genome sequences of *Theaceae* species were obtained from the GenBank database (Supplementary Table S1). A ML phylogenetic tree was constructed based on these chloroplast genome data using *Coffea arabica* and *Arabidopsis thaliana* as outgroups. The phylogenetic analysis indicated that *P. Dancong* tea and Fujian oolong tea clustered together with other cultivated *Camellia sinensis* in the genus *Camellia* of the family *Theaceae*. Apparently, the two oolong tea species clustered in the same clade and were embedded in the family, *Camellia*, which corroborated the close relationship between these two families (Figure 7A). We found that five Fujian oolong tea species formed a monophyletic clade along with *P. Dancong* tea. Wuyi Narcissus and Shuijingui are closely related and relatively independent from the others. Both Baijiguan-Bantianyao and Dahongpao-Rougui reside on a separate, highly supported branch within the Tieluohan clade. The results also revealed that the *P. Dancong* teas, Hongyin and Baiyin, are sisters to Fujian oolong teas, Bantianyao and Rougui, respectively. In addition, we note that Leikouchai-Baxian, Zhuye-Potou, Dongfanghong-Youhuaxiang, Dawuye--Tuofuhou, Lanrenxiang-Jilongkan, Nanjiangxiang-Guihuaxiang, Juduoza i-Wuye, Huangjinye-Milanxiang and Rouguixiang-Laoxianweng are highly conserved (Figure 7B).

4. Discussion

Oolong tea, a unique kind of tea in southern China, has long been classified and identified according to its phenotypic characteristics [2]. In recent years, with the development of high-throughput sequencing technology and efficient large-scale data analysis, an increasing number of studies on plant evolutionary relationships have been carried out. Chloroplasts are specialized photosynthetic organelles in plants that are highly conserved and maintain relatively small copy numbers, making them easy to sequence for further evolutionary studies [7].

In this study, we sequenced the complete chloroplast genome of *P. Dancong* tea by using Illumina high-throughput sequencing technology. The genomes ranged from 157,041 to 157,137 bp, which is consistent with the known sizes of other *Camellia* species, such as *C. japonica* (157,001 bp, Accession number: MW602996.1) [34] and

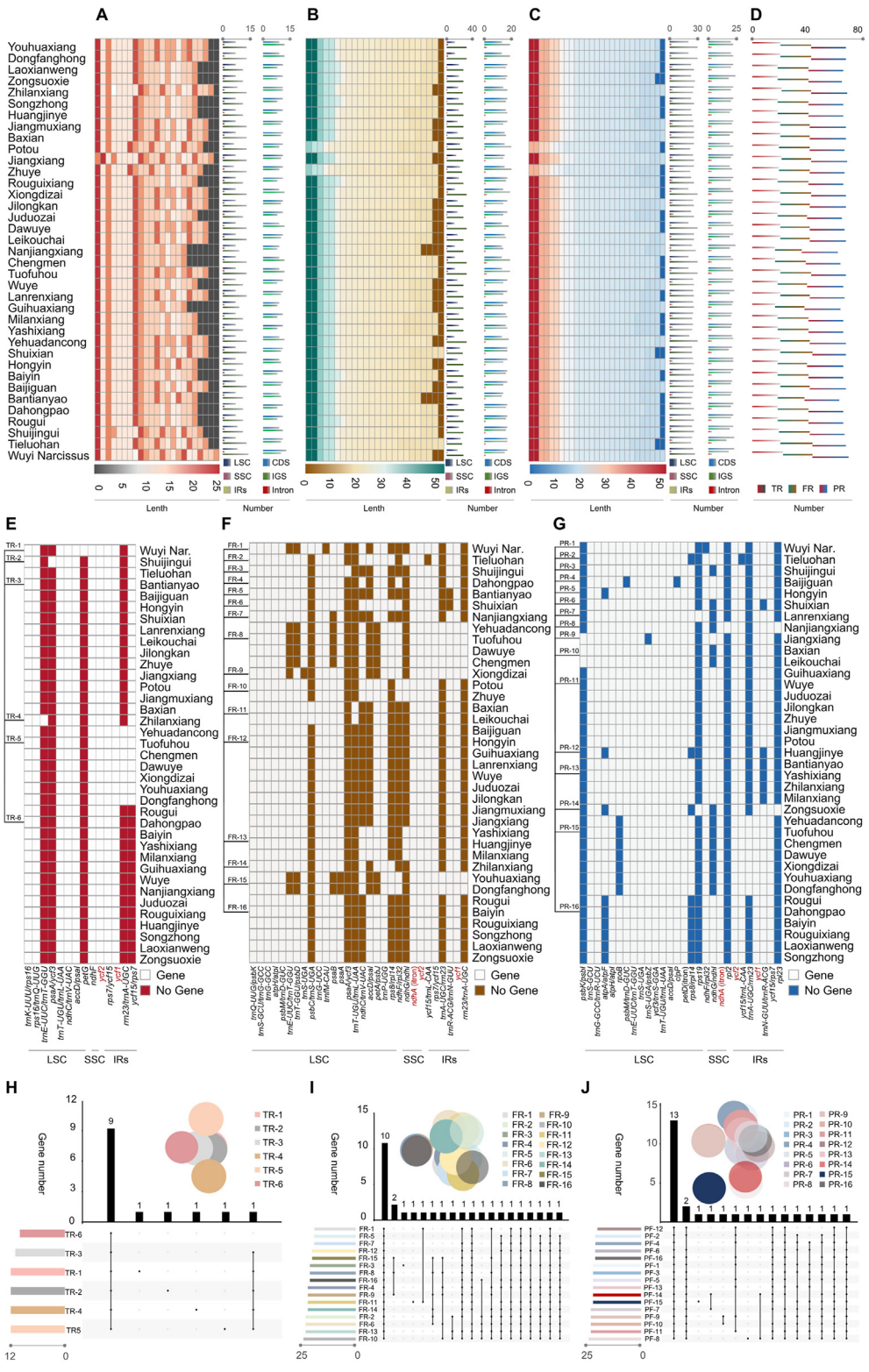

Figure 5. Analysis of long repeated sequences in oolong tea chloroplast genomes. (A) Numbers and lengths of TRs. (B) Numbers and lengths of FRs. (C) Numbers and lengths of PRs. (D) Total lengths of the three repeat types. (E) Identification, distribution and clustering of TR loci. (F) Identification, distribution and clustering of FR loci. (G) Identification, distribution and clustering of PR loci. (H) Summary of unique and shared TR loci. (I) Summary of unique and shared FR loci. (J) Summary of unique and shared PR loci. The plot on the left shows the total number of related genes, and the plot on the right shows the number of unique repeat loci, followed by the intersections/overlaps (connected dots) between branching pairs.

C. rostrata (156,547 bp, Accession number: MW755303.1) [35]. Its quadripartite organization consists of two copies of IRs separating LSC and SSC, which are almost identical to previously reported angiosperm chloroplast genomes [11]. The sizes of the LSC, SSC and IR regions of *P. Dancong* tea were also well within the size range of other sequenced *Camellia* species [26]. The structures and sizes of intraspecific and interspecific genomes are extremely conserved, which may be related to the slow molecular evolution of plant chloroplasts [36]. ORF42 was absent in the IRs of *P. Dancong* tea, which could be associated with sequencing/assembly errors or the occurrence of gene loss. However, ORF42 was absent in IRa and IRb in all 30 *P. Dancong* teas. Loss of ORF42 has been reported in another other cultivated tea, *Camellia sinensis* var. *assamica* cv. Yunkang 10 [29]. It is speculated that the gene was lost in *P. Dancong* tea with a high probability.

The overall GC contents of the chloroplast genomes ranged from 37.29% to 37.31%, with the IR region being the highest (43.00%),

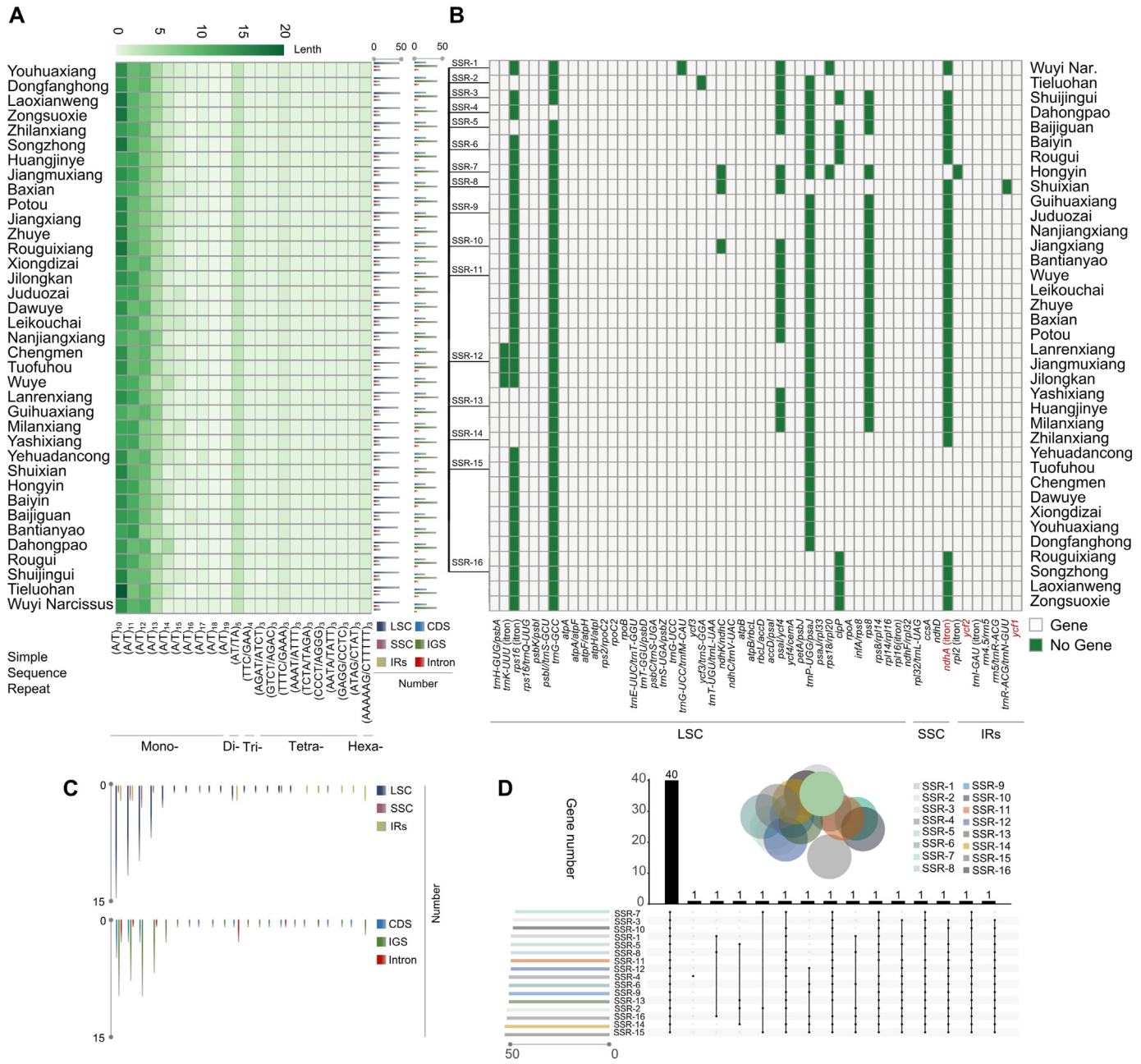

Figure 6. Comparison of SSR distributions in oolong tea chloroplast genomes. (A) Numbers and lengths of different SSR types. (B) Identification, distribution and clustering of SSR loci. (C) Numbers of mono-, di-, tri-, tetra- and hexa-nucleotides. (D) Summary of unique and shared SSR loci. The plot on the left shows the total numbers of related genes, and the plot on the right shows the numbers of unique repeat loci, followed by intersections/overlaps (connected dots) between branching pairs.

followed by LSC (35.30%) and SSC (30.50%), which may be due to the presence of rRNAs (e.g., rrna4.5, rrna5, rrna23 and rrna16), which are GC-rich in the IR region. It should be noted that although the chloroplast genomes of most angiosperms generally encode 74 CDSs [37], *P. Dancong* tea, like other teas, harbors 90 CDSs (including *accD*, *ycf1*, *ycf2*, *rpl23* and *infA*), and other plants also have the same numbers, such as *Rosa filipes* [38] and *Cerasus humilis* [39]. Relative synonymous codon usage analysis showed that the starting codons of *rps19* and *ndhD* had been replaced by GTG and ended with TAA. A similar situation occurs for the chloroplast genomes of other plants, such as *Neotropical Bulbophyllum* [40] and *Nelumbo nucifera* [41]. Leucine and isoleucine are the most common essential amino acids of plants and are used with the highest frequency in *P. Dancong* tea [42]. Furthermore, almost all of the A/T-ending codons showed RSCU values greater than 1, whereas the C/G-ending codons showed RSCU values less than 1, indicating that A/T ending codons are used more frequently.

The gene content and order of arrangement of the chloroplast genome of *P. Dancong* tea are basically the same as those of Fujian oolong tea and other unorganized *Camellia* [26]. Therefore, the exact IR boundary locations and adjacent genes of the oolong tea chloroplast genomes from two different origins were compared in this study. Expansion and contraction of the IR region are critical for the size variations of the chloroplast genome and play an important role in its stability and evolution [43, 44]. The IR region of oolong tea was not only highly conserved in terms of length, but the locations and sizes of the boundary genes, *rps19*, *ycf1* and *trnH* were also relatively stable. Although the positions of these boundary genes are slightly different from those of other plants, the gene types are almost the same, which also exists in

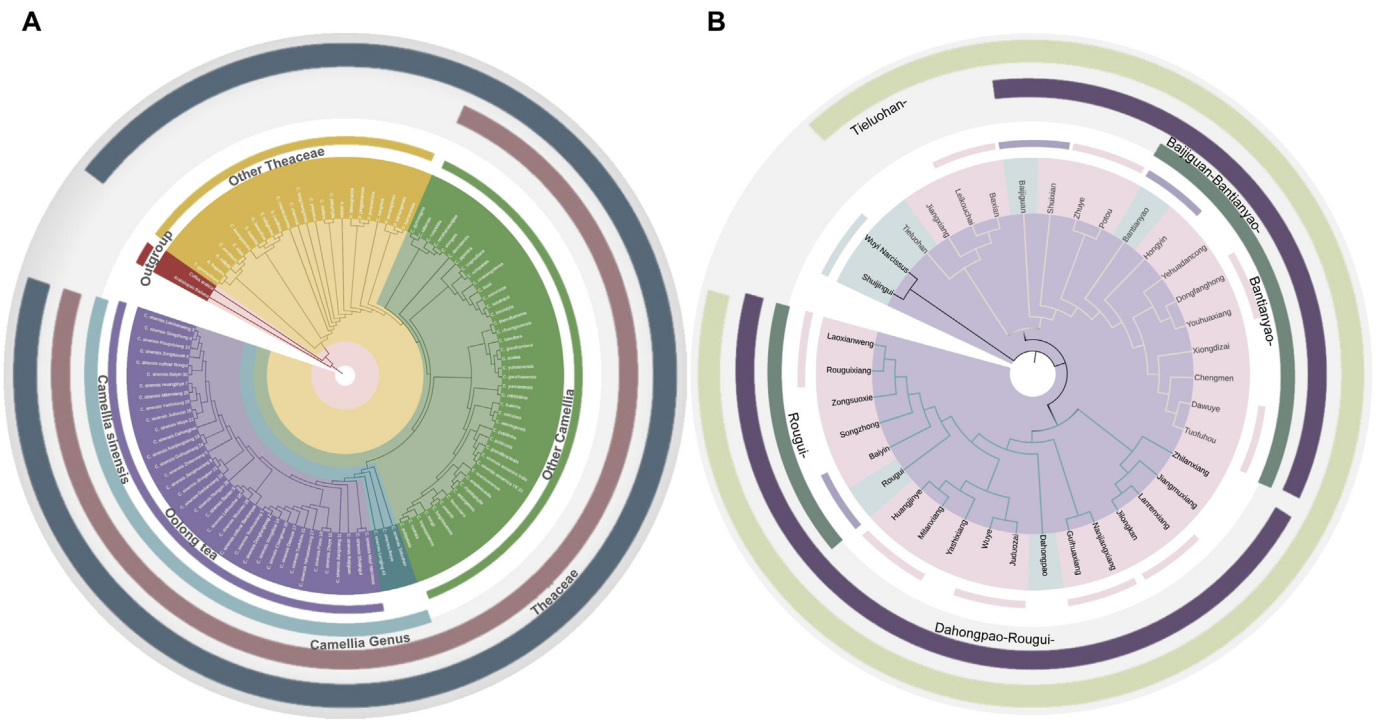

Figure 7. Phylogenetic tree constructed based on the complete chloroplast genomes of *P. Dancong* teas, oolong teas and other *Theaceae* species. (A) Phylogenetic tree of oolong tea in the family *Theaceae*. Circles from outside to inside: outermost circle-dark blue: *Theaceae* family; second circle-light red: *Camellia* family; third circle-light green: *Camellia sinensis*; innermost circle-purple: oolong tea, -green: other *Camellia*, -yellow: other *Theaceae*, -red: outgroup. (B) Phylogenetic tree based on the complete chloroplast genomes of *P. Dancong* teas and Fujian oolong teas. Circles from outside to inside: outermost circle-light green: *P. Dancong* tea that is on the same branch as Tieluohan; second circle-purple: *P. Dancong* tea that is on the same branch as Baijiguan-Bantianyao and Dahongpao-Rougui; third circle-dark green: *P. Dancong* tea that is on the same branch as Bantianyao and Rougui; innermost circle-pink: *P. Dancong* tea Leikouchai-Baxian, Zhuye-Potou, Dongfanghong-Youhuaxiang, Dawuye-Tuofuhou, Lanrenxiang-Jilongkan, Nanjiangxiang-Guihuaxiang, Juduozai-Wuye, Huangjinye-Milanxiang and Rouguixiang-Laoxianweng are on the same branch.

Stemona sessilifolia [45], *Myristicaceae* [46], and *Brassicaceae* [47]. Accordingly, we infer that there should be no simple phylogenetic signals at the IR boundary.

It is obvious that the chloroplast genome of oolong tea contains repeats other than IRs, which can be of great help in further research on genome rearrangements and phylogenetic analyses. There were three long repeats (TR, FR and PR) in the chloroplast genome of oolong tea, of which the number in PR were the largest and were mainly distributed in the IGS. These corresponding regions are potential hotspots of gene reconstruction and can provide information sources for genetic markers. The genes, *ycf1* and *ycf2,* appeared simultaneously in three long repeats in all samples, suggesting that these long repeats may be shared among oolong teas. Although a large number of loci with long repeated sequences have been discovered, as in other plants, the origin mechanism and function remain unclear [26].

Five types of SSRs (e.g., mononucleotides, dinucleotides, trinucleotides, tetranucleotides and hexanucleotides) were detected, ranging in number from 65 to 70 and in size from 10 bp to 20 bp. They are mostly located in noncoding regions, especially in intergenic spacer regions, which usually consist of A/T. The SSR loci identified in this study may be useful in population and evolutionary research. Many of these loci, such as *rps16/trnQ-UUG*, *psbK/psbI*, *atpH/atpI*, *rps2/rpoC2*, and *atpB/rbcL*, have been applied in research on the interspecific phylogenetic relationships and intraspecific genetic variations of *Oryza* [48], *Bamboo* [49], *Brassica* [50], and *Impatiens* [51]. Introns of *trnK-UUU*, *rps16* and *rpl16* were also the focus of chloroplast gene sequence variation analyses [52]. In addition, there is a rare SSR locus, *rps8/rpl14,* in the chloroplast genome of oolong tea, which may contribute to its biochemical genetic identification and population genetic studies. Little research has been reported on the development and application of SSR molecular markers for the chloroplast genome associated with oolong tea. Based on the available data, we can develop SSR molecular markers to obtain a more comprehensive understanding of the evolutionary history and genetic polymorphisms of oolong tea, which is the direction of our future research.

The *Theaceae* family is taxonomically and phylogenetically challenging due to the similarity of its morphological characteristics. A comparative analysis of the chloroplast genome of *P. Dancong* tea with the previously sequenced family, *Theaceae,* revealed the characteristics of chloroplasts in terms of gene contents and structural organizations. The complete chloroplast genomes of *P. Dancong* tea and other tea species all displayed conservative quadripartite circular structures and stable structural arrangements. The chloroplast genomes are accompanied by sequence repeats and indels among different tea species, which are mainly reflected in the variations in lengths and orientations of the IR region. This result coincided with the previously proposed view that sequence variations were the most important evolutionary dynamic contributing to the diversification of the tea chloroplast genome [53]. Molecular phylogenetic analysis showed that *P. Dancong* tea and Fujian oolong tea were derived from *Camellia sinensis*, a genus *Camellia* of the family *Theaceae*, and formed a closely intersecting branch. This suggested that the chloroplast genome of *P. Dancong* tea is most similar to that of Fujian oolong tea, which may be caused by closely homologous relatives sharing their genome contents. From the perspective of oolong tea, Fujian oolong teas, Wuyi Narcissus, Shuijingui and Tieluohan, are ancient members of an early differentiation, among which Tieluohan forms a sister relationship with *P. Dancong* tea. Baijiguan, Bantianyao and Rougui are closer to Shuixian, Hongyin and Baiyin, respectively. On this basis, it is considered that a single chloroplast phylogenetic lineage has developed over a long period of evolution. This phenomenon also explains the fact that the chloroplast genome is more conserved and evolves much more slowly than the nuclear genome, which is more susceptible to environmental influences. Therefore, molecular data from the nuclear

genome and population-level genetic analyses are necessary to thoroughly explore the evolutionary mechanisms. Our results suggested the phylogenetic relationships of some *P. Dancong* tea species according to the available chloroplast genomes, but more comprehensive genome sequences are required to resolve the overall phylogeny of this species. In future research, it is necessary to collect more species and use more molecular data to verify the interspecific relationships and interspecific differentiation of *P. Dancong* teas.

5. Conclusions

In this work, we present for the first time the complete chloroplast genome of *P. Dancong* tea using Illumina sequencing technology. Its size and structure were fully characterized and compared with the chloroplast genomes of related species. A comparative analysis indicated that the chloroplast genomes of *P. Dancong* teas and Fujian oolong teas exhibited similar gene contents and organizations. By analyzing the chloroplast genome characteristics, alignments, codon usage biases, repeated sequences and phylogenies, the hypothesis that the chloroplast genome of oolong tea has a low variation rate and conservative evolutionary history was proposed, which means that the known data can be used as a reference for annotating the chloroplast genomes of other oolong tea species. Furthermore, *P. Dancong* tea and Fujian oolong tea clustered together in a monophyletic group, suggesting a clear family relationship. These genomic information and analysis results will provide a theoretical basis for subsequent research on the phylogeny, taxonomy and biogeography of oolong tea.

Declarations

Author contribution statement

Yaqun Liu: Performed the experiments; Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data; Wrote the paper.

Liyun Lin; Dongjuan Yang; Xianghui Zou: Performed the experiments; Contributed reagents, materials, analysis tools or data.

Zhenxia Zhang; Mouquan Liu: Contributed reagents, materials, analysis tools or data.

Min Lin; Yuzhong Zheng: Conceived and designed the experiments; Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data; Wrote the paper.

Funding statement

Yuzhong Zheng was supported by Guangdong Key Laboratory of Functional Substances in Medicinal Edible Resources and Healthcare Products [No. 2021B1212040015].

PhD Yaqun Liu was supported by Doctor Initiating Project of the Hanshan Normal University [No. QD202125].

Min Lin was supported by Doctor Initiating Project of the Hanshan Normal University [No. QD20190527], Guangxi Provincial Natural Science Foundation [No. 2019JJD140052 & No. 2020JJA140656], Education Department Project of Guangdong Province [No. 2019-GDXK-0032 & No. 2020KZDZX1146], Eastern Guangdong Technological Engineering Research Center [No. P19004].

Data availability statement

Data associated with this study has been deposited at GenBank (Accession number: MZ379786 for Youhuaxiang; OL690346-OL690374 for Dongfanghong, Laoxianweng, Zongsuoxie, Zhilanxiang, Songzhong, Huangjinye, Jiangmuxiang, Baxian, Potou, Jiangxiang, Zhuye, Rouguixiang, Xiongdizai, Jilongkan, Juduozai, Dawuye, Leikouchai, Nanjiangxiang, Chengmen, Tuofuhou, Wuye, Lanrenxiang, Guihuaxiang, Milanxiang, Yashixiang, Yehuadancong, Shuixian, Hongyin and Baiyin).

Declaration of interest's statement

The authors declare no competing interests.

Additional information

Supplementary content related to this article has been published online at https://doi.org/10.1016/j.heliyon.2022.e08895.

DNA sequences and GenBank accession numbers

All data generated or analyzed during this study are included in this published article. The chloroplast genome sequences of *P. Dancong* tea were submitted to GenBank (Accession number: MZ379786 for Youhuaxiang; OL690346-OL690374 for Dongfanghong, Laoxianweng, Zongsuoxie, Zhilanxiang, Songzhong, Huangjinye, Jiangmuxiang, Baxian, Potou, Jiangxiang, Zhuye, Rouguixiang, Xiongdizai, Jilongkan, Juduozai, Dawuye, Leikouchai, Nanjiangxiang, Chengmen, Tuofuhou, Wuye, Lanrenxiang, Guihuaxiang, Milanxiang, Yashixiang, Yehuadancong, Shuixian, Hongyin and Baiyin). The accession numbers corresponding to the additional datasets used and analyzed in this study can be found in the Supplementary Table 1.

Acknowledgements

We thank technician Youcai Li and other laboratory members at Genesky Biotechnology Inc. (Shanghai, China) for their technical help and collaboration. We also thank the Weiye Tea Plantation for providing the raw materials used in this research and for helpful discussions. We are especially grateful to our team members for their important ideas and active participation in this study.

References

[1] P.O. Owuor, M. Obanda, H.E. Nyirenda, W.L. Mandala, Influence of region of production on clonal black tea chemical characteristics, Food Chem. 108 (1) (2008) 263–271.

[2] K.W. Ng, Z.J. Cao, H.B. Chen, Z.Z. Zhao, L. Zhu, T. Yi, Oolong tea: a critical review of processing methods, chemical composition, health effects, and risk, Crit. Rev. Food Sci. Nutr. 58 (17) (2018) 2957–2980.

[3] X. Zhang, Z. Song, Y. You, X. Li, T. Chen, *Phoenix Dan Cong* Tea: an Oolong Tea variety with promising antioxidant and in vitro anticancer activity, Food Nutr. Res. (2018 Nov 13) 62.

[4] L.Q. Liu, Z.J. Gu, Genomic in situ hybridization identifies genome donors of *Camellia reticulata (Theaceae)*, Plant Sci. 180 (3) (2011 Mar) 554–559.

[5] Q.S. Chen, C.C. Sun, Q. Ouyang, Y.X. Wang, A.P. Liu, H.H. Li, J.W. Zhao, Classification of different varieties of Oolong tea using novel artificial sensing tools and data fusion, Lwt Food Science & Technology 60 (2) (2015) 781–787.

[6] J. Lin, P. Zhang, Z. Pan, H. Xu, Y. Luo, X. Wang, Discrimination of oolong tea (*Camellia sinensis*) varieties based on feature extraction and selection from aromatic profiles analysed by HS-SPME/GC-MS, Food Chem. 141 (1) (2013) 259–265.

[7] Q.J. Li, N. Su, L. Zhang, R.C. Tong, X.H. Zhang, J.R. Wang, Z.Y. Chang, L. Zhao, D. Potter, Chloroplast genomes elucidate diversity, phylogeny, and taxonomy of *Pulsatilla (Ranunculaceae)*, Sci. Rep. 10 (1) (2020), 19781.

[8] G. Hölzl, P. Dörmann, Chloroplast lipids and their biosynthesis, Annu. Rev. Plant Biol. 70 (2019) 51–81.

[9] V. Ravi, J.P. Khurana, A.K. Tyagi, P. Khurana, An update on chloroplast genomes, Plant Systemat. Evol. 271 (2008) 101–122.

[10] A.V. Williams, L.M. Boykin, K.A. Howell, P.G. Nevill, I. Small, The complete sequence of the *Acacia ligulata* chloroplast genome reveals a highly divergent *clpP1* gene, PLoS One 10 (5) (2015), e0125768.

[11] J.D. Palmer, Comparative organization of chloroplast genomes, Annu. Rev. Genet. 19 (1985) 325–354.

[12] S.J. Wei, Y.B. Lu, Q.Q. Ye, S.Q. Tang, Population genetic structure and phylogeography of *Camellia flavida* (*Theaceae*) based on chloroplast and nuclear DNA sequences, Front. Plant Sci. 8 (2017 May 19) 718.

[13] S. Chen, Y. Zhou, Y. Chen, J. Gu, fastp: an ultra-fast all-in-one FASTQ preprocessor, Bioinformatics 34 (17) (2018) i884–i890.

[14] S. Nurk, D. Meleshko, A. Korobeynikov, P.A. Pevzner, metaSPAdes: a new versatile metagenomic assembler, Genome Res. 27 (5) (2017) 824–834.

[15] L. Shi, H. Chen, M. Jiang, L. Wang, X. Wu, L. Huang, C. Liu, CPGAVAS2, an integrated plastome sequence annotator and analyzer, Nucleic Acids Res. 47 (W1) (2019) W65–W73.

[16] S. Zheng, P. Poczai, J. Hyvönen, J. Tang, A. Amiryousefi, Chloroplot: an online program for the versatile plotting of organelle genomes, Front. Genet. 11 (2020), 576124.

[17] S. Kumar, G. Stecher, K. Tamura, MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets, Mol. Biol. Evol. 33 (7) (2016) 1870–1874. Epub 2016 Mar 22.
[18] K. Katoh, D.M. Standley, MAFFT multiple sequence alignment software version 7: improvements in performance and usability, Mol. Biol. Evol. 30 (4) (2013) 772–780.
[19] K.A. Frazer, L. Pachter, A. Poliakov, E.M. Rubin, I. Dubchak, VISTA: computational tools for comparative genomics, Nucleic Acids Res. 32 (2004) W273–W279. Web Server issue).
[20] A. Amiryousefi, J. Hyvönen, P. Poczai, IRscope: an online program to visualize the junction sites of chloroplast genomes, Bioinformatics 34 (17) (2018) 3030–3031.
[21] P. Librado, J. Rozas, DnaSP v5: a software for comprehensive analysis of DNA polymorphism data, Bioinformatics 25 (11) (2009) 1451–1452.
[22] S. Kurtz, C. Schleiermacher, REPuter: fast computation of maximal repeats in complete genomes, Bioinformatics 15 (5) (1999 May) 426–427.
[23] G. Benson, Tandem repeats finder: a program to analyze DNA sequences, Nucleic Acids Res. 27 (2) (1999) 573–580.
[24] T. Thiel, W. Michalek, R.K. Varshney, A. Graner, Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.), Theor. Appl. Genet. 106 (3) (2003) 411–422.
[25] W. Luo, C. Brouwer, Pathview: an R/Bioconductor package for pathway-based data integration and visualization, Bioinformatics 29 (14) (2013 Jul 15) 1830–1831.
[26] H. Huang, C. Shi, Y. Liu, S.Y. Mao, L.Z. Gao, Thirteen *Camellia* chloroplast genome sequences determined by high-throughput sequencing: genome structure and phylogenetic relationships, BMC Evol. Biol. 14 (2014 Jul 7) 151.
[27] A. Stamatakis, RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies, Bioinformatics 30 (9) (2014) 1312–1313.
[28] H.C. Rawal, S. Borchetia, B. Bera, S. Soundararajan, R.V.J. Ilango, A.K. Barooah, T.R. Sharma, N.K. Singh, T.K. Mondal, Comparative analysis of chloroplast genomes indicated different origin for Indian tea (*Camellia assamica* cv TV1) as compared to Chinese tea, Sci. Rep. 11 (1) (2021 Jan 8) 110.
[29] J. Peng, Y. Zhao, M. Dong, S. Liu, Z. Hu, X. Zhong, Z. Xu, Exploring the evolutionary characteristics between cultivated tea and its wild relatives using complete chloroplast genomes, BMC Ecol Evol 21 (1) (2021 Apr 30) 71.
[30] S. Chen, R. Li, Y. Ma, S. Lei, R. Ming, X. Zhang, The complete chloroplast genome sequence of *Camellia sinensis* var. sinensis cultivar Tieguanyin (*Theaceae*), Mitochondrial DNA B Resour 6 (2) (2021 Feb 8) 395–396.
[31] I. Letunic, P. Bork, Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees, Nucleic Acids Res. 44 (W1) (2016) W242–W245.
[32] M. Oresic, D. Shalloway, Specific correlations between relative synonymous codon usage and protein secondary structure, J. Mol. Biol. 281 (1) (1998 Aug 7) 31–48.
[33] W. Powell, M. Morgante, R. McDevitt, G.G. Vendramin, J.A. Rafalski, Polymorphic simple sequence repeat regions in chloroplast genomes: applications to the population genetics of pines, Proc. Natl. Acad. Sci. U. S. A. 92 (17) (1995) 7759–7763.
[34] W. Li, C. Zhang, X. Guo, Q. Liu, K. Wang, Complete chloroplast genome of *Camellia japonica* genome structures, comparative and phylogenetic analysis, PLoS One 14 (5) (2019), e0216645.
[35] Z.Y. Zhang, Y.K. Yang, P.M. Ye, J.L. Ma, H. Ye, The complete chloroplast genome sequence of *Camellia rostrata* S. X. Yang & S. F. Chai (*Theaceae*), a critically endangered yellow *Camellia* from southwest China, Mitochondrial DNA B Resour 6 (8) (2021) 2427–2429.
[36] B. Ghada, B.A. Ahmed, C. Khaled, S. Olfa, M. Messaoud, T. Mokhtar, S.H. Amel, Molecular evolution of chloroplast DNA in fig (*Ficus carica* L.): footprints of sweep selection and recent expansion, Biochem syst ecol 38 (4) (2010) 563–575.
[37] R.S. Millen, R.G. Olmstead, K.L. Adams, J.D. Palmer, N.T. Lao, L. Heggie, T.A. Kavanagh, J.M. Hibberd, J.C. Gray, C.W. Morden, P.J. Calie, L.S. Jermiin, K.H. Wolfe, Many parallel losses of *infA* from chloroplast DNA during angiosperm evolution with multiple independent transfers to the nucleus, Plant Cell 13 (3) (2001) 645–658.
[38] S.Q. Wang, J.F. Chen, Z.M. Zhu, The complete chloroplast genome sequence of *Rosa filipes* (Rosaceae)[J], Mitochondrial DNA Part B 5 (2) (2020) 1376, 137.
[39] X. Mu, P. Wang, J. Du, Y.G. Gao, J. Zhang, The chloroplast genome of *Cerasus humilis*: genomic characterization and phylogenetic analysis, PLoS One 13 (4) (2018 Apr 25), e0196473.
[40] M. Zavala-Páez, L.D.N. Vieira, V.A. de Baura, E. Balsanelli, E.M. de Souza, M.C. Cevallos, M.W. Chase, E.C. Smidt, Comparative plastid genomics of neotropical *Bulbophyllum* (orchidaceae; epidendroideae), Front. Plant Sci. 11 (2020) 799.
[41] Z. Wu, S. Gui, Z. Quan, L. Pan, S. Wang, W. Ke, D. Liang, Y. Ding, A precise chloroplast genome of *Nelumbo nucifera* (Nelumbonaceae) evaluated with Sanger, Illumina MiSeq, and PacBio RS II sequencing platforms: insight into the plastid evolution of basal eudicots, BMC Plant Biol. 14 (2014) 289.
[42] X. Liu, Y. Li, H. Yang, B. Zhou, Chloroplast genome of the folk medicine and vegetable plant *Talinum paniculatum* (jacq.) gaertn.: gene organization, comparative and phylogenetic analysis, Molecules 23 (4) (2018) 857.
[43] R. Kolodner, K.K. Tewari, Inverted repeats in chloroplast DNA from higher plants, Proc. Natl. Acad. Sci. U. S. A. 76 (1) (1979 Jan) 41–45.
[44] F.M. Xie, D.F. Xie, C. Xie, Y. Yu, S.D. Zhou, X.J. He, Adaptation evolution and phylogenetic analyses of species in Chinese *allium* section *pallasia* and related species based on complete chloroplast genome sequences, BioMed Res. Int. (2020 Jun 11), 8542797, 2020.
[45] J. Liu, M. Jiang, H. Chen, Y. Liu, C. Liu, W. Wu, Comparative genome analysis revealed gene inversions, boundary expansions and contractions, and gene loss in the *Stemona sessilifolia* (Miq.) Miq. chloroplast genome, PLoS One 16 (6) (2021), e0247736.
[46] C.N. Cai, H. Ma, A.Q. Ci, J.G. Conran, J. Li, Comparative phylogenetic analyses of Chinese *Horsfieldia* (Myristicaceae) using complete chloroplast genome sequences, J. Systemat. Evol. 59 (3) (2020).
[47] C. Osuna-Mascaro, R. Casas, J.B. Landis, F. Perfectti, Genomic resources for *Erysimum* spp. (Brassicaceae): transcriptome and chloroplast genomes, Front Ecol Evol 9 (2021), 620601.
[48] Y. Song, Y. Chen, J. Lv, J. Xu, S. Zhu, M. Li, N. Chen, Development of chloroplast genomic resources for Oryza species discrimination, Front. Plant Sci. 8 (2017) 1854.
[49] T.R. Jani, B.P. Sheth, V.S. Thaker, Chloroplast microsatellite based molecular and computational studies in selected Bamboo species[J], Int. J. Plant Res. 26 (2) (2013) 403.
[50] S.G. Gandhi, P. Awasthi, Y.S. Bedi, Analysis of SSR dynamics in chloroplast genomes of Brassicaceae family, Bioinformation 5 (1) (2010) 16–20.
[51] C. Salgado-Salazar, Y. Rivera, D. Veltri, J.A. Crouch, Polymorphic SSR markers for *Plasmopara obducens* (Peronosporaceae), the newly emergent downy mildew pathogen of *Impatiens* (Balsaminaceae), Appl. Plant Sci. 3 (11) (2015), 1500073.
[52] A.O. Simes, M.E. Endress, T. Niet, L.S. Kinoshita, E. Conti, Is *Mandevilla* (apocynaceae, mesechiteae) monophyletic? Evidence from five plastid DNA loci and morphology, Ann. Mo. Bot. Gard. 93 (4) (2006) 565–591. IMAMME]2.0.CO;2.
[53] L. Li, Y. Hu, M. He, B. Zhang, W. Wu, P. Cai, D. Huo, Y. Hong, Comparative chloroplast genomes: insights into the evolution of the chloroplast genome of *Camellia sinensis* and the phylogeny of *Camellia*, BMC Genom. 22 (1) (2021) 138.