


Article

Chloroplast Genome of Rambutan and Comparative Analyses in Sapindaceae

Fei Dong¹, Zhicong Lin², Jing Lin³, Ray Ming^{4,*}  and Wenping Zhang^{3,*}

¹ College of Life Sciences, Fujian Agriculture and Forestry University, Fuzhou 350002, China; dongfei9508@163.com or 1180514010@fafu.edu.cn

² College of Agriculture, Fujian Agriculture and Forestry University, Fuzhou 350002, China; woshi0655@126.com or 2160102007@fafu.edu.cn

³ Center for Genomics and Biotechnology, Fujian Provincial Key Laboratory of Haixia Applied Plant Systems Biology, Key Laboratory of Genetics, Fujian Agriculture and Forestry University, Fuzhou 350002, China; lolyemily@163.com or jingling@fafu.edu.cn

⁴ Department of Plant Biology, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA

* Correspondence: rayming@illinois.edu (R.M.); wenpingzhang@fafu.edu.cn (W.Z.); Tel.: +1-217-333-1221 (R.M.); Tel.: +86-15-8006-2379 (W.Z.)

Abstract: Rambutan (*Nephelium lappaceum* L.) is an important fruit tree that belongs to the family Sapindaceae and is widely cultivated in Southeast Asia. We sequenced its chloroplast genome for the first time and assembled 161,321 bp circular DNA. It is characterized by a typical quadripartite structure composed of a large (86,068 bp) and small (18,153 bp) single-copy region interspersed by two identical inverted repeats (IRs) (28,550 bp). We identified 132 genes including 78 protein-coding genes, 29 tRNA and 4 rRNA genes, with 21 genes duplicated in the IRs. Sixty-three simple sequence repeats (SSRs) and 98 repetitive sequences were detected. Twenty-nine codons showed biased usage and 49 potential RNA editing sites were predicted across 18 protein-coding genes in the rambutan chloroplast genome. In addition, coding gene sequence divergence analysis suggested that *ccsA*, *clpP*, *rpoA*, *rps12*, *psbJ* and *rps19* were under positive selection, which might reflect specific adaptations of *N. lappaceum* to its particular living environment. Comparative chloroplast genome analyses from nine species in Sapindaceae revealed that a higher similarity was conserved in the IR regions than in the large single-copy (LSC) and small single-copy (SSC) regions. The phylogenetic analysis showed that *N. lappaceum* chloroplast genome has the closest relationship with that of *Pometia tomentosa*. The understanding of the chloroplast genomics of rambutan and comparative analysis of Sapindaceae species would provide insight into future research on the breeding of rambutan and Sapindaceae evolutionary studies.



Citation: Dong, F.; Lin, Z.; Lin, J.; Ming, R.; Zhang, W. Chloroplast Genome of Rambutan and Comparative Analyses in Sapindaceae. *Plants* **2021**, *10*, 283. <https://doi.org/10.3390/plants10020283>

Received: 6 January 2021

Accepted: 29 January 2021

Published: 2 February 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: rambutan; *Nephelium lappaceum*; chloroplast genome; Sapindaceae; comparative genomic; RNA editing; phylogeny

1. Introduction

Rambutan (*Nephelium lappaceum* L.) is an important tropical fruit in the family Sapindaceae and originated in Indonesia and the Malay Peninsula [1]. It is widely cultivated in Southeast Asia and the coastal areas of South China. Malaysians refer to it as “rambutan”, because the fruit surface is covered with thick and elongated spines. The fruits of rambutan are popular in the general population due to their rich nutrients, delicate and characteristic flavor and delicious taste. Rambutan peel extract is rich in phenolic content and exhibited antibacterial activity against many pathogenic bacteria, suggesting its antioxidant and/or antimicrobial properties [2]. Rambutan has the potential to be used in natural antioxidants and anti-aging agents in pharmaceutical and food industries to replace synthetic ones [3,4].

The Sapindaceae family contains over 150 genera and 2000 species with several economically important crops, which are widely distributed in tropical and subtropical

regions [5]. However, genomic research on the Sapindaceae family, especially in the *N. lappaceum*, has been relatively scarce. This lack of genetic information makes it difficult to meet the need for improving the quality and agronomic characteristics of rambutan through breeding and gene editing.

Chloroplast (cp) are photosynthetic organelles that provide energy to green plants; they play an important role in the photosynthesis and secondary metabolic activities [6,7]. The chloroplast genomes are maternally inherited in most plants and are highly conserved in terms of their composition and sequence. The typical chloroplast genomes of angiosperms are circular DNA molecules which have a characteristic quadripartite structure with a large single-copy (LSC) region, a small single-copy (SSC) region, and two inverse repeats (IRs) regions [6]. The length of the genome is between 120 and 170 kb and usually encodes 110 to 130 genes, and about 40 genes are specialized, participating in photosynthesis, transcription and translation [8,9]. The first chloroplast genome from tobacco (*Nicotiana tabacum* L.) was sequenced in 1986 [10]. With the rapid development of next-generation sequencing technologies, the cost of whole genome sequencing is dropping rapidly [11]. Complete chloroplast genome sequences now could be easily acquired with a relatively low cost. There has been an explosion in the number of available chloroplast genome sequences. Over 5000 complete chloroplast genome sequences have been submitted in the National Center for Biotechnology Information (NCBI) organelle genome database.

Within the Sapindaceae family, the complete chloroplast genomes of eight plant species have been sequenced and were available from the NCBI database. Nevertheless, no chloroplast genome in the genus *Nephelium* L. has been reported. In this study, we report the first complete chloroplast genome of *N. lappaceum*, exploring its general features, SSRs and long repeats, codon usage and analysis of IR contraction and expansion. In addition, nine chloroplast genome sequences were used for the analysis of molecular evolution in the Sapindaceae family. We constructed a phylogenetic tree to understand the phylogenetic relationships of Sapindaceae plants. The chloroplast genome sequence and the comprehensive chloroplast genomic analysis of *N. lappaceum* would provide a theoretical basis for molecular identification and further understanding of the evolutionary history of the Sapindaceae family.

2. Results

2.1. Chloroplast Genome Features of *N. lappaceum*

The structure of the *N. lappaceum* chloroplast genome was analogous to most chloroplast genomes of plants with a typical quadripartite structure. We assembled a closed circular chloroplast genome with 161,321 bp in *N. lappaceum*. The chloroplast genome contains a pair of inverted repeat regions (IRs) of 28,550 bp, a large single-copy region (LSC) of 86,068 bp and a small single-copy region (SSC) of 18,153 bp (Figure 1). Besides, the overall nucleotide composition of rambutan is: 30.79% A, 31.44% T, 19.27% C, and 18.50% G, with a total GC content of 37.77%. In total, 132 genes were annotated on this chloroplast genome, including 78 protein-coding genes, 29 transfer RNA genes (tRNA) and 4 ribosomal RNA genes (rRNA). Among them, a total of 21 genes were found duplicated in the IR regions, including nine protein-coding genes (*rps3*, *rps7*, *rps12*, *rps19*, *rpl2*, *rpl22*, *rpl23*, *ndhB* and *ycf2*), eight tRNA genes (*trnA-UGC*, *trnI-CAU*, *trnI-GAU*, *trnL-CAA*, *trnM-CAU*, *trnN-GUU*, *trnR-ACG* and *trnV-GAC*) and four rRNA genes (*rrn4.5s*, *rrn5s*, *rrn16s* and *rrn23s*) (Table S2). The gene structure analysis showed that 21 genes contain introns, and 19 of them (11 protein-coding genes and 8 tRNA genes) have one intron, while two genes (*ycf3* and *clpP*) have two introns (Table S3). We characterized the basic features of the chloroplast genome in Sapindaceae and compared them with *N. lappaceum*. The size of the *N. lappaceum* chloroplast genome was slightly larger than that of *Sapindus mukorossi* Gaertn. (160,481 bp), *Pometia tomentosa* (Blume) Teijsm. & Binn. (160,818bp), *Dimocarpus longan* Lour. (160,833bp), *Dodonaea viscosa* (Linn.) Jacq. (159,375bp), and *Eurycorymbus cavaleriei* (H.Lev.) Rehder & Hand.-Mazz. (158,777bp), but shorter than that of *Litchi chinensis* Sonn.

(162,524 bp), *Koelreuteria paniculata* Laxm. (163,258bp), and *Xanthoceras sorbifolium* Bunge. (161,231bp) chloroplast genome of Sapindaceae (Table 1). The number of chloroplast genes in *N. lappaceum* was 132, the same as that of *D. longan*, *L. chinensis* and *X. sorbifolium* (Table 1). The highest number of genes was found in *E. cavaleriei* (137), followed by *S. mukorossi* (135) and *D. viscosa* (135). The number of protein-coding genes and tRNA genes in these species varied from 85 to 89 and 37 to 40, respectively. Furthermore, there was no significant difference in GC content among the nine analytical genomes in Sapindaceae.

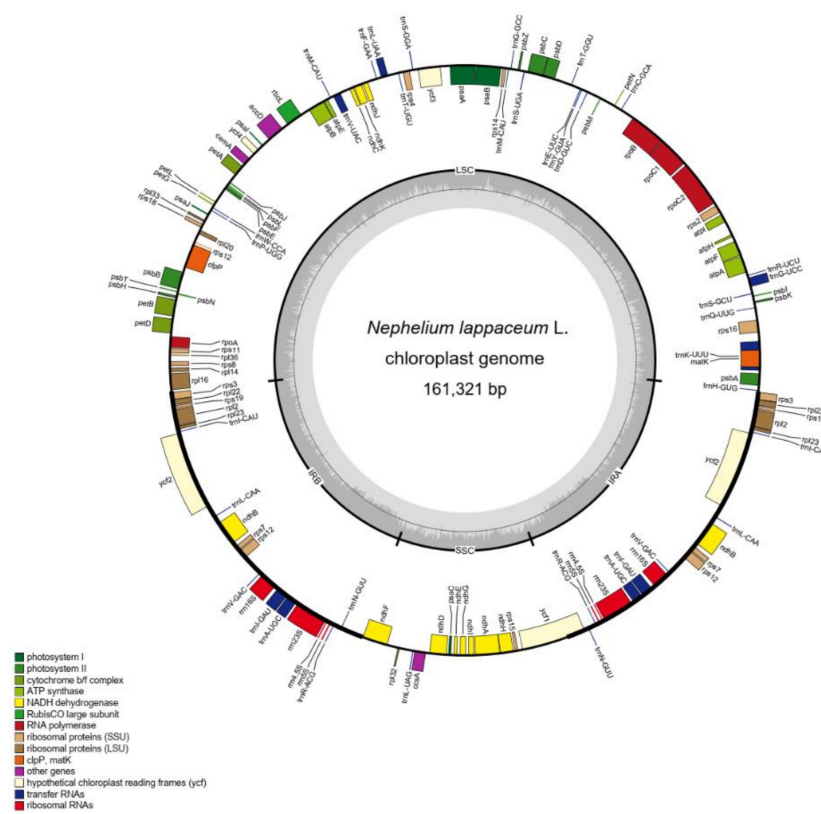


Figure 1. Gene map of *N. lappaceum* chloroplast genome. Genes drawn outside and inside of the circle are transcribed clockwise and counterclockwise, respectively. Genes belonging to different functional groups are color coded. The darker gray in the inner circle corresponds to GC content. Small single-copy (SSC) region, large single-copy (LSC) region, and inverted repeats (IRA and IRB) are indicated.

Table 1. Comparison of the general features of the nine Sapindaceae chloroplast genomes.

Genome Feature	<i>Dimocarpus longan</i>	<i>Litchi chinensis</i>	<i>Pometia tomentosa</i>	<i>Sapindus mukorossi</i>	<i>Nephelium lappaceum</i>	<i>Dodonaea viscosa</i>	<i>Eurycorymbus cavaleriei</i>	<i>Koelreuteria paniculata</i>	<i>Xanthoceras sorbifolium</i>
GenBank	MG214255	KY635881	MN106254	KM454982	MT936934	KM454982	MF155892	MG813997	KY859413
Size (bp)	160,833	162,524	160,818	160,481	161,321	159,375	158,777	163,258	161,231
LSC (bp)	85,707	85,750	85,666	85,650	86,068	872,014	86,940	90,236	85,299
SSC (bp)	18,270	16,568	18,360	18,873	18,153	17,972	17,991	18,268	18,692
IR (bp)	28,428	30,103	28,396	27,979	28,550	27,099	26,923	27,377	28,620
Total genes	132	132	133	135	132	135 (2 Pseudogene)	137	133 (3 Pseudogene)	132
Protein genes	87	87	88	88	87	88	89	85	86
tRNA genes	37	37	37	39	37	37	40	37	38
rRNA genes	8	8	8	8	8	8	8	8	8
GC (%)	37.79%	37.80%	37.87%	37.66%	37.77%	37.86%	37.92%	37.30%	37.69%

Note: IR: Inverted repeats. GC: GC content.

2.2. Characterization of SSRs and Repeat Sequences

A total of 63 SSRs were detected from rambutan chloroplast genome, of which 45 were mononucleotide, 3 dinucleotide, 8 trinucleotide, 5 tetranucleotide and two pentanucleotide (Figure 2B). Moreover, we compared the distribution pattern and number of SSRs with eight other chloroplast genomes in the Sapindaceae family (Table S4, Table S5). The number of mononucleotide repeats was more than the sum of other types (Figure 2A), and the number and types of chloroplast SSRs varied in different species. *S. mukorossi* (91 SSRs) possesses the highest number of SSRs while *E. cavaleriei* (62 SSRs) possesses the lowest. Furthermore, the chloroplast genome of *D. longan*, *L. chinensis*, *P. tomentosa*, *D. viscosa*, *K. paniculata* and *X. sorbifolium* contained 79, 75, 74, 77, 87 and 83 SSRs, respectively. In this study, a total of 98 larger repeats (>10 bp) were identified in *N. lappaceum* chloroplast genome, composed of 42 forward, 11 reverse, 41 palindromic and 4 complement repeats (Table S6) using REPuter [12]. Among them, the largest repeat was a palindromic repeat with a size of 48 bp.

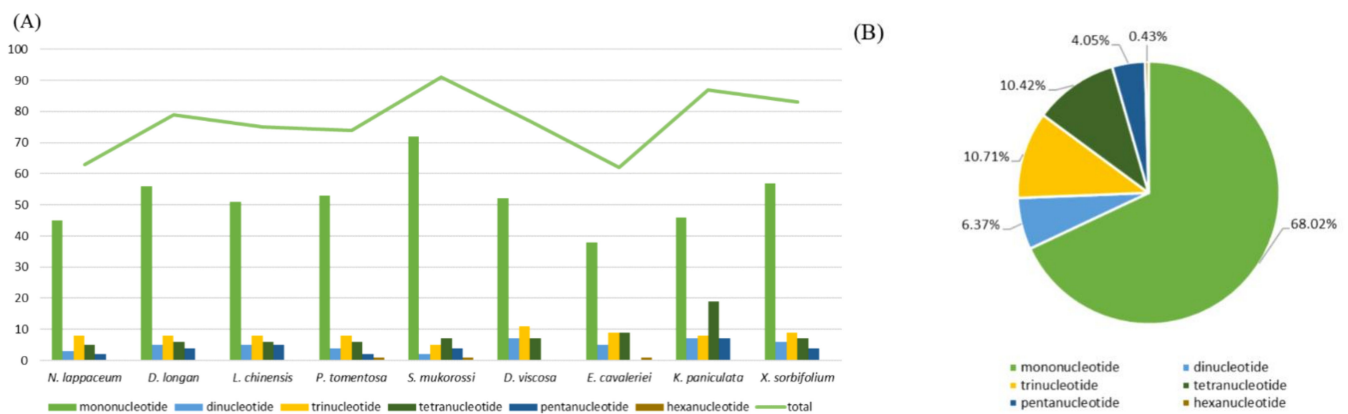


Figure 2. Analysis of simple sequence repeats (SSRs) in nine Sapindaceae (including *N. lappaceum*) chloroplast genomes. **(A)** Number of different SSRs types detected in nine Sapindaceae (including *N. lappaceum*) chloroplast genomes. **(B)** Presence of different SSRs types in all SSRs of nine Sapindaceae (including *N. lappaceum*) chloroplast genomes.

2.3. Codon Usage Analysis and RNA Editing Sites Prediction

We used 53 protein-coding sequences from rambutan chloroplast genome to calculate codon usage frequency and relative synonymous codon usage (RSCU) frequency (Table S7). All protein-coding sequences contain 21,434 codons. In detail, leucine and cysteine were the highest and lowest number of amino acids; they had 2232 codons (approximately 10.41% of the total) and 236 codons (approximately 1.10% of the total), respectively. While Met (ATG) and Trp (TGG) were encoded by only one codon, showing no biased usage (RSCU = 1). There are 30 codons with RSCU values greater than 1, indicating that they showed biased usage (Figure 3). Among them, excluding the leucine (UUG) codon, which was G-ending, the remaining 29 biased usage codons of *N. lappaceum* were all A/T-ending in the third codon. The usage was generally biased towards A or T(U) with higher RSCU values, including UUA (1.91) in leucine, the stop-codon UAA (1.87), and GCU (1.74) in Alanine in the chloroplast genome of *N. lappaceum*. Besides, there were 49 potential RNA editing sites found across 18 protein-coding genes in the *N. lappaceum* chloroplast genome, and the *ndhB* gene contained the most RNA editing sites (9) (Table S8). We also observed that RNA editing sites all showed C to U conversion, which took place at the first (30.6%) or second (69.4%) positions of the codons, indicating that editing in the third codon position occurred at lower frequency than that in the second or first codon position. Furthermore, serine codons were more frequently edited than codons of other amino acids and the conversion from serine to leucine occurred most frequently.

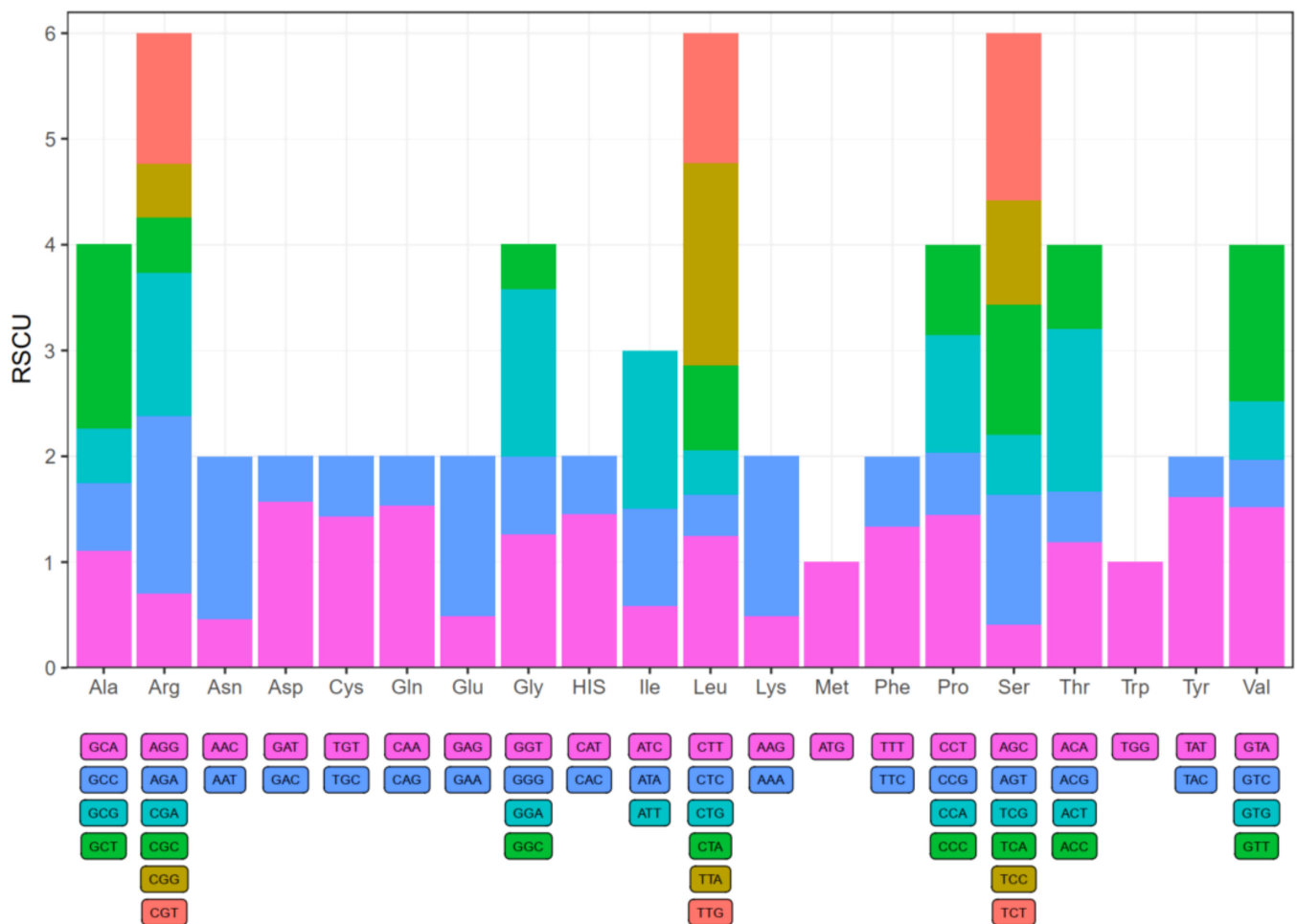


Figure 3. Codon content of 20 amino acids and stop codons in all protein-coding genes of *N. lappaceum* chloroplast genome. The colour of the histogram corresponds to the colour of codons.

2.4. Comparative Genomes Analysis

The comparative analysis based on mVISTA was performed between the chloroplast genomes of rambutan, with the other eight Sapindaceae species with the annotated *D. longan* chloroplast genome as a reference. The nine Sapindaceae family chloroplast genomes' length was between 158,777 and 163,258 bp. The chloroplast genome of *K. paniculata* had the largest size, whereas *E. cavaleriei* had the smallest size. Interestingly, the SSC region (16,568 bp) of *L. chinensis* was the shortest, whereas the SSC region (18,873 bp) of the *S. mukorossi* chloroplast genome was the longest (Figure 4). The IR (A/B) regions exhibited less divergence than the SSC and LSC regions. In addition, the coding regions were more highly conserved than the non-coding regions. Among the nine chloroplast genomes, four rRNA genes (*rrn16S*, *rrn23S*, *rrn5S*, *rrn4.5S*) were the most conserved, while four genes (*ndhF*, *ndhD*, *ndhH* and *ycf1*) showed the most diversity in the coding regions. The highly divergent regions were found in the intergenic spacers and introns, including *trnH-GUG-psbA*, *trnS-GCU-trnG-UCC*, *trnR-UCU-atpA*, *atpF-atpH*, *petN-psbM*, *psbZ-trnG-GCC*, *trnF-GAA-ndhJ*, *ndhC-trnV-UAC*, *psbE-petL*, *ndhF-rpl32*, *rpl16-rps3* and *rpl32-trnL-UAG*.

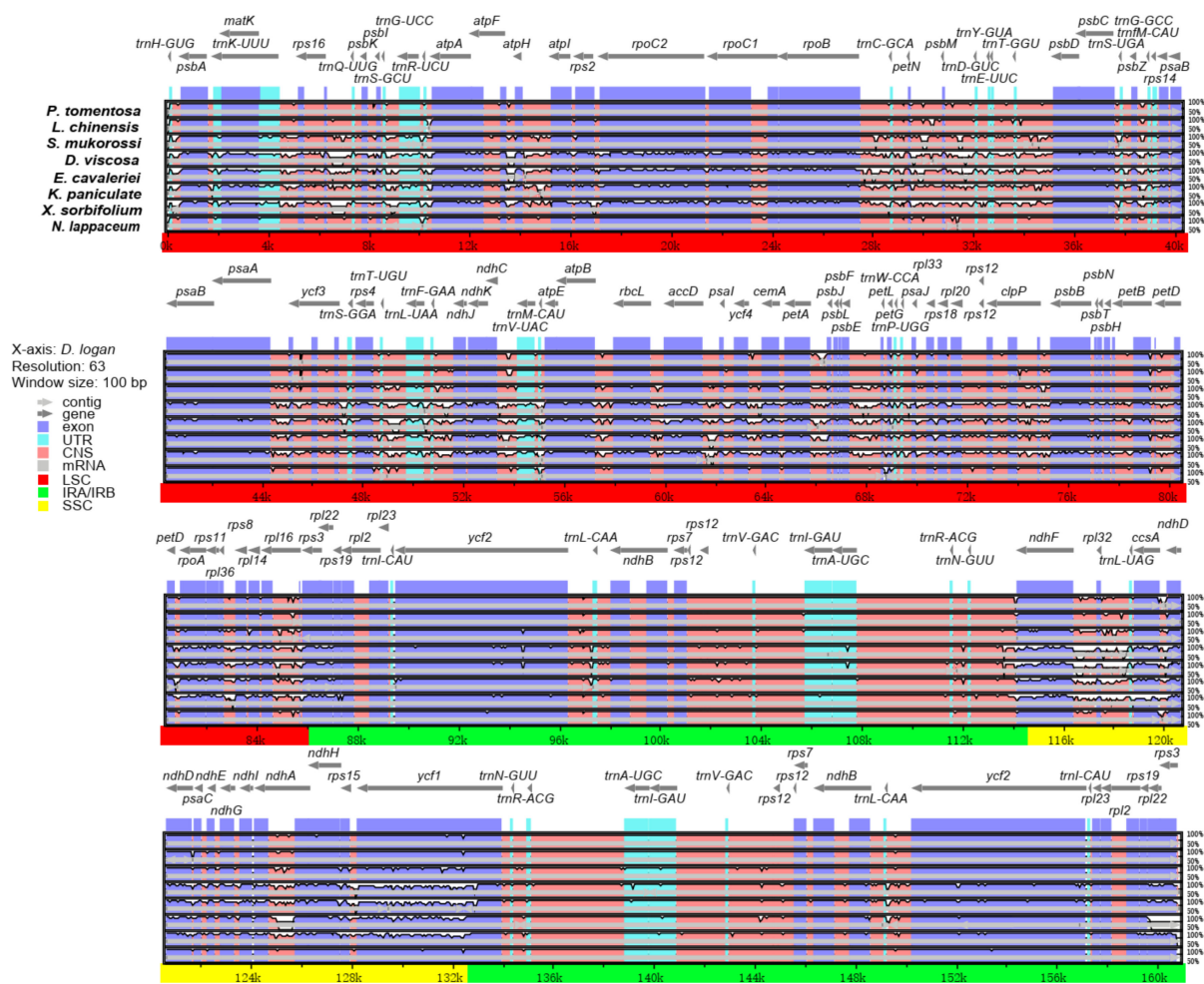


Figure 4. Comparison of nine Sapindaceae chloroplast genomes (including *D. logan*, *L. chinensis*, *P. tomentosa*, *S. mukorossi*, *D. viscosa*, *E. cavaleriei*, *K. paniculata*, *X. sorbifolium* and *N. lappaceum*), with *D. logan* as a reference. Gray arrows and thick black lines above the alignment indicate the direction of the gene. Purple bars represent exons, blue bars represent untranslated regions (UTRs), pink bars represent conserved non-coding sequences (CNS), and gray bars represent mRNA. The y-axis indicates the identity, expressed as a percentage, between 50% and 100%.

2.5. Expansion and Contraction of IR Regions

We compared the IR regions and the junction sites of the LSC and SSC regions of nine Sapindaceae family chloroplast genomes (including *N. lappaceum*) (Figure 5). The IR regions vary in different chloroplast genomes, ranging from 26,923 bp in *E. cavaleriei* to 30,103 bp in *L. chinensis*. In our study, the *ycf1* gene was located at the SSC/IRA junction in all of the nine chloroplast genomes and the fragment located at the IRa region ranged from 962 bp to 3183 bp. Moreover, most junctions between LSC and IRA in this study were located downstream of the *trnH-GUG*, except the *S. mukorossi*. In addition, the LSC/IRb junction of three species *D. viscosa*, *E. cavaleriei* and *K. paniculata* was located within the coding region of *rpl22* and created a location of 110, 40 or 63 bp at the LSC/IRb border. The remaining chloroplast genomes share a similar pattern; the LSC/IRb junction was located in intergenic regions of *rpl16* and *rps3*, and the IRb/SSC junction between IRb and SSC region (JSB) of five species (*S. mukorossi*, *X. sorbifolium*, *D. viscosa*, *E. cavaleriei* and *K. paniculata*) was located between the gene of *ycf1* and *ndhF*. However, the other four chloroplast genomes only have *ndhF* located at or near the JSB.

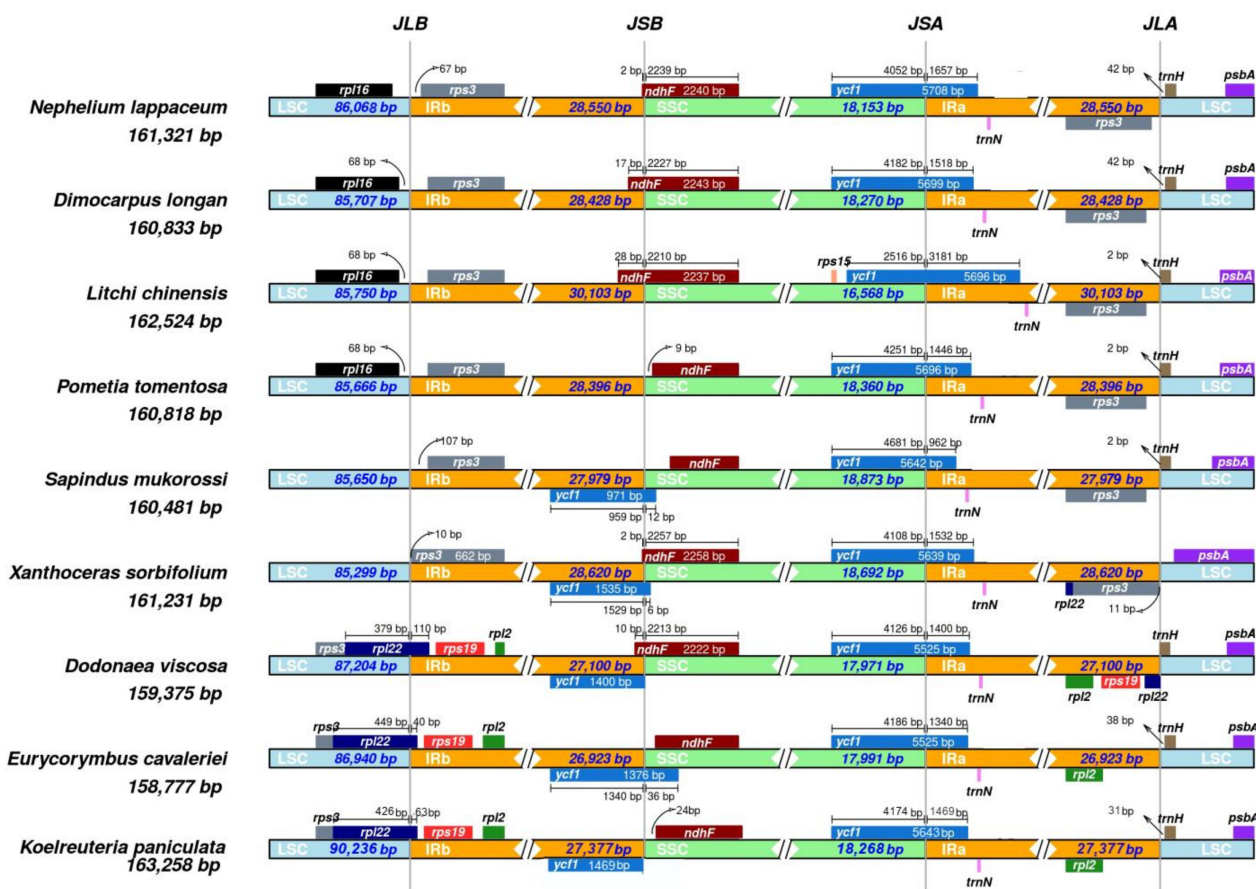


Figure 5. Comparison of the borders of the LSC, SSC and IR regions among nine Sapindaceae chloroplast genomes. For each species, genes transcribed in positive strand are depicted on the top of their corresponding track from right to left direction, while the genes on the negative strand are depicted below from left to right. The numbers at arrows refer to the distance of the start or end position of a given gene from the corresponding junction site. The T bars above or below the genes indicate the extent of their parts with their corresponding values in the base pair. The plotted genes and distances in the vicinity of the junction sites are the scaled projection of the genome. JLB (IRb /LSC), JSB (IRb/SSC), JSA (SSC/IRa) and JLA (IRa/LSC) denote the junction sites between each corresponding two regions of the genome.

2.6. Synonymous (K_s) and Non-Synonymous (K_a) Substitution Rate Analysis

To explore the molecular evolution of orthologous genes shared by nine Sapindaceae species, particularly genes undergoing purifying or positive selection, we calculated the K_a/K_s ratio of 622 orthologous pairs with 78 protein-coding genes (Table S9). Overall, the average K_a/K_s ratio of the nine chloroplast genomes was 0.20. In total, 612 orthologous pairs had a K_a/K_s ratio of less than 1 in the nine comparison groups, out of which 546 orthologs had a K_a/K_s ratio of less than 0.5 (Figure 6), suggesting that most genes were undergoing strong purifying selection pressures. Moreover, 66 orthologs of 31 genes with a K_a/K_s ratio between 0.5 and 1, and 10 orthologous pairs of 6 genes (*ccsA*, *rpoA*, *rps12*, *psbJ*, *clpPc* and *rps19*) with a K_a/K_s ratio greater than 1, were detected in this study, suggesting that these genes might have experienced positive selection in the procedure of evolution. Among them, the K_a/K_s ratio of the *ycf1* gene was greater than 0.5 in eight comparison groups; the *rpoA* and *ycf2* gene with a K_a/K_s ratio greater than 0.5 was also observed in the comparison of seven and six groups, respectively. Besides, the *clpP*, *matK* and *rps15* genes, with K_a/K_s ratios >0.5, were found in four out of the eight comparison groups.

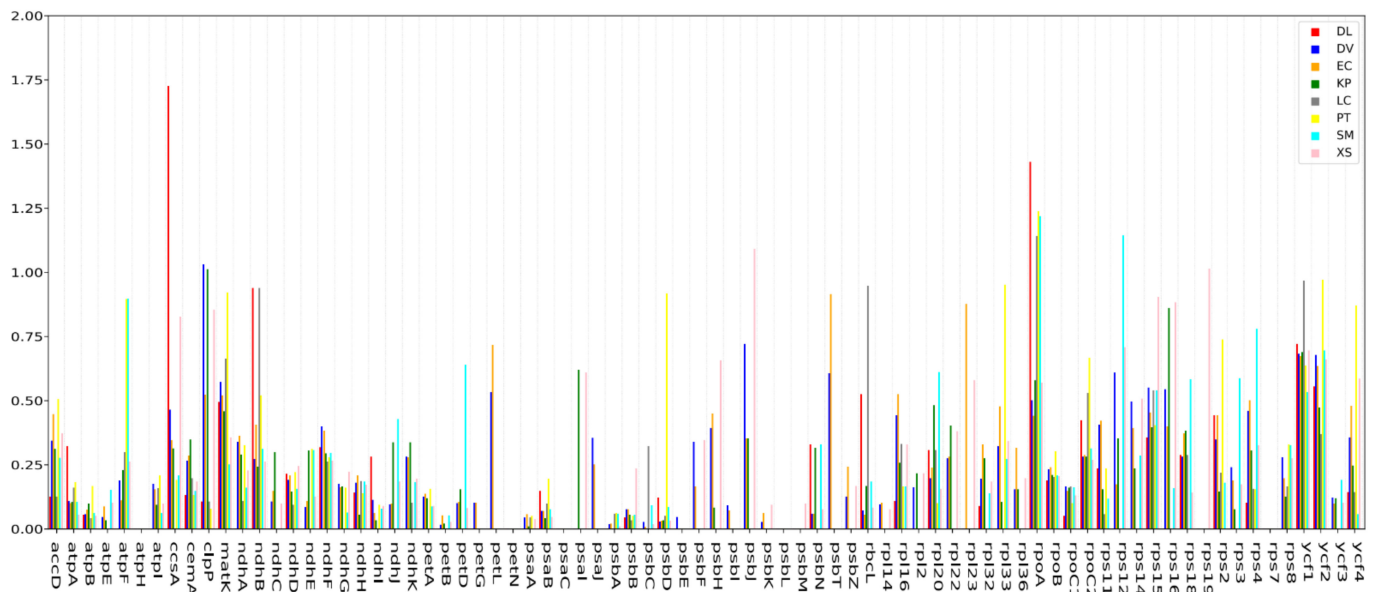


Figure 6. The Ka/Ks ratios of 78 protein-coding genes of the *N. lappaceum* chloroplast genome versus eight closely related species of Sapindaceae.

2.7. Phylogenetic Analysis

We performed multiple sequence alignments using the whole chloroplast genome sequences of nine Sapindaceae species and two Anacardiaceae species as outgroups (Figure 7). All nodes in the ML trees have 100% bootstrap support values, and these 11 chloroplast genome sequences were clustered into three groups. In detail, the five species (*D. longan*, *L. chinensis*, *P. tomentosa*, *N. lappaceum* and *S. mukorossi*) from Sapindoideae clustered into one group, four species (*K. paniculata*, *D. viscosa*, *E. cavaleriei* and *X. sorbifolium*) from Dodonaeoideae were in one group, and the two species (*A. occidentale* and *M. indica*) in Anacardiaceae were clustered into one group. In the Sapindoideae group, the *N. lappaceum* chloroplast genome sequence showed the closest relationship with *P. tomentosa*, followed by *D. longan* and *L. chinensis*, as far as *S. mukorossi*. The three groups of this phylogenetic tree of the 11 chloroplast genome sequences were consistent with traditional taxonomy, suggesting that the chloroplast genome could effectively resolve the phylogenetic positions and relationships of species.

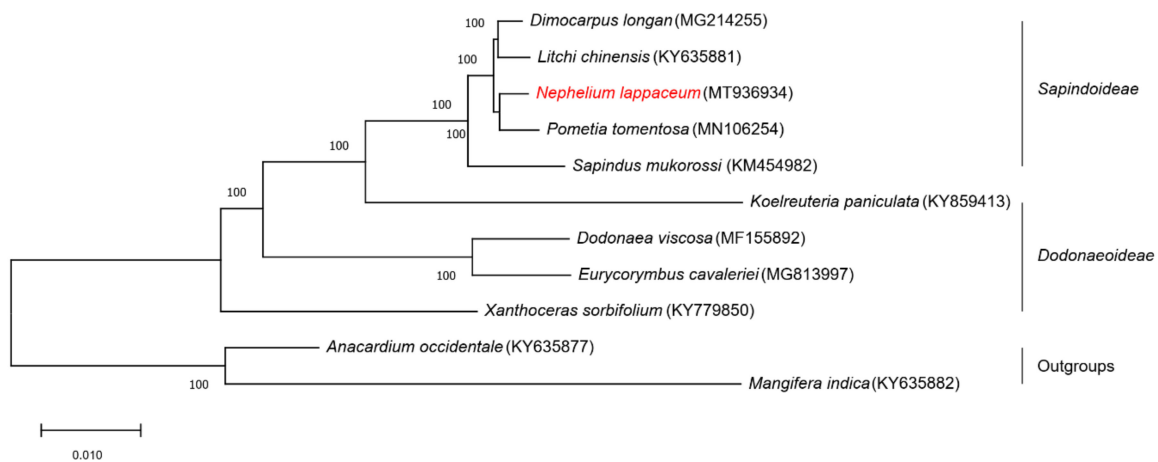


Figure 7. The maximum likelihood (ML) phylogenetic tree of the Sapindaceae family based on chloroplast genome sequences. The numbers in each node were tested by bootstrap analysis with 1000 replicates. *Anacardium occidentale* and *Mangifera indica* were set as the outgroups. The position of *N. lappaceum* is indicated in red text.

3. Discussion

We assembled the complete *N. lappaceum* chloroplast genome sequence and deposited it to GenBank under accession number: MT936934. The *N. lappaceum* chloroplast genome is consistent with the characteristics of most angiosperm species in structure and gene content. Although there are some differences in the sizes of the overall genome, LSC, SSC and IR regions, the numbers of genes and GC content are similar among the nine Sapindaceae chloroplast genomes, which, to some extent, reflects the high conservation of angiosperm chloroplast genomes [6]. Notably, the number of tRNA genes in *E. cavaleriei* and *S. mukorossi* are quite different from other species in Sapindaceae since some of the tRNA genes types and copy numbers are different. The copy number of tRNA genes may be affected by differences in gene codon composition and amino acid usage [13]. Intron plays an important role in RNA stability, regulation of gene expression and alternative splicing, which has been reported in many other species [14,15]. There were two genes (*ycf3* and *clpP*), in the *N. lappaceum* chloroplast genome, that included two introns. It has been reported that the *ycf3* gene was essential for the accumulation of the photosystem I (PSI) complex and acts as a chaperone that interacts with the PSI subunits at a post-translational level [16,17]. Besides, the *clpP* gene functions as the proteolytic subunit of the ATP-dependent Clp protease in plant chloroplasts and has been shown to be essential for the development and/or function of plastids with active gene expression in previous studies [17–19]. Thus, the study of the *ycf3* and *clpP* genes will contribute to further investigation of chloroplast in *N. lappaceum*.

Simple sequence repeats (SSRs), also known as microsatellites, are tandem repeats distributed across the entire genome which have been widely applied as molecular markers for determining genetic variations across species in evolutionary studies because of their unique uniparental inheritance [20–22]. The mononucleotide SSRs, A and T, were identified most frequently (68% on average) among the nine analyzed chloroplast genomes of Sapindaceae species. This result is consistent with the previous report that poly A and T are the most abundant repeats in most angiosperm chloroplast genomes [23–25]. Moreover, repetitive sequences are helpful in phylogenetic study and play a vital role in genome rearrangement [25]. Most of the repetitive sequences in the *N. lappaceum* chloroplast genome are distributed in IGS regions, whereas few are located in the region of the protein-coding genes. These results can provide chloroplast molecular markers of family Sapindaceae that can be used to quickly identify species and confirm hybrid progeny when breeding.

Codon usage biases are found in all eukaryotic and prokaryotic genomes and have been proposed to regulate different aspects of the translation process [26]. High RSCU values of the codons are probably attributed to amino acid functions or peptide structures that avoid transcriptional errors in chloroplast genomes [27,28]. The phenomenon of codons ending in A/T in our study is similar to the pattern reported for other chloroplast genomes, which may be caused by a composition bias for a high A/T ratio [23]. High codon preference is prevalent in other land plant chloroplast genomes and the results of our study are similar to those of other species with chloroplast genome codon usage biases. The research on codon preferences can help us to better understand the gene expression and molecular evolution mechanisms of *N. lappaceum*.

We observed that the *ndhB* gene contained the most RNA editing sites within the 49 potential RNA editing sites, and 16 editing sites were U_A type, indicating there was a U_A bias for the distribution of RNA editing sites that were in accordance with previous reports in other species [24,29]. RNA editing is a post-transcriptional regulation pattern involved in the insertion, deletion, or modification of nucleotides that widely exists in land plants [30]. The first chloroplast RNA editing event of a land plant was discovered in the mRNA transcript of the *rpl2* gene in the maize chloroplast genome in 1991 [31]. The most frequent editing events in plants are C-to-U changes; however, U-to-C editing has also been observed [32,33]. Additionally, RNA editing usually occurs in the first or second base of codons, resulting in the conversion of hydrophilic amino acid to hydrophobic [34].

Comparative analysis of chloroplast genomes is an essential step in genomics that can provide insight into complex evolutionary relationships. The mVISTA analysis showed that nine Sapindaceae chloroplast genomes were conserved, with a high degree of similarity and gene order conservation, and the coding region was more conserved than the non-coding region, which is consistent with reports on other angiosperms [35], suggesting an evolutionary conservation of these genomes at the genome-scale level. Notably, the five species in the Sapindoideae subfamily presented the same divergence pattern, but the divergence of the four species in the Dodonaeoideae subfamily was greater. In addition, the *ycf1* gene showed the greatest degree of differentiation. Previous studies reported that *ycf1* is helpful to provide phylogenetic information at the species level and more variable than *matK* in Orchidaceae [36]. Furthermore, *ycf1* performed better in identifying DNA barcodes of high resolution at the species level than either *matK*, *rbcL* or *trnH-psbA* [35]. These variable genic regions found in our study can be regarded as molecular markers for DNA barcoding and phylogenetic studies in Sapindaceae.

Although most land plants have relatively conserved cp genomes, the end of the inverted repeats (IRa and IRb) regions differs among various plant species. The expansion or contraction of the IR regions represent important evolutionary events that often influence the size variation of different chloroplast genomes, and it is thus helpful to study the chloroplast genome evolution history [37,38]. In this study, our results suggested that the boundary of IR/LSC and IR/SSC might be conserved among chloroplast genomes of closely related family species but some differences also occur between relatively distantly related family species, such as gene overlap length, a duplicate of the *ycf1* and *rps3* genes, even the distance of *trnH-GUG* from the border near the LSC/IRB junctions, indicating that the expansion and contraction of the IR regions led to length and structure changes in chloroplast genomes.

The ratio between nonsynonymous (Ka) and synonymous (Ks) nucleotide substitution has been widely used as an important marker in genome or gene evolution studies [8]. $Ka/Ks = 1$ signifies neutral evolution, $Ka/Ks > 1$ indicates that the gene is affected by positive selection, whereas $Ka/Ks < 1$ indicates that the gene is affected by purifying selection [39]. Additionally, a Ka/Ks ratio of 0.5 was considered as a useful cut-off value to identify genes under positive selection in previous studies [40]. In our study, the *ccsA*, *rpoA*, *rps12*, *psbJ*, *clpP* and *rps19* gene with $Ka/Ks > 1$. It is noteworthy that the *ycf1* gene also exhibited high Ka/Ks ratios with $Ka/Ks > 0.5$ in eight comparison groups. This result is in keeping with the previous observations that the *ycf1* gene was more variable than the *matK* and *rbcL* genes in most plants, and could be used as an effective biological tool for plant phylogeny study [35]. The positive selection of genes in *N. lappaceum* possibly provided help for adaptations to its particular living environment.

Numerous studies have shown that chloroplast genome sequences have been successfully used in taxonomic and phylogenetic studies [41], and contribute to describing the evolutionary relationships between species [42]. In this study, the topology of the trees consists of two main branches: Dodonaeoideae and evolutionary younger Sapindoideae. Furthermore, generic relationships of the two subfamilies are basically congruent with the taxonomy of these families. The availability of the completed *N. lappaceum* chloroplast genome provided us with sequence information that can be used to confirm the phylogenetic position of *N. lappaceum* and understand the phylogenetic relationships among Sapindaceae. However, as we used only a small number of species in Sapindaceae, further research on other chloroplast genomes as well as nuclear genome sequences of Sapindaceae should be conducted to provide more sufficient evidence to accurately illustrate the evolution of the family Sapindaceae.

4. Materials and Methods

4.1. Plant Material, DNA Extraction, and Sequencing

Young, healthy leaves of the major cultivar of rambutan, Baoyan7, were collected from Baoting (N18°23', E109°21') in Hainan Province, China. The leaves were frozen

in liquid nitrogen and maintained at $-80\text{ }^{\circ}\text{C}$. The total genomic DNA was extracted by 2X cetyltrimethylammonium bromide (CTAB) method [43]. In addition, a library with insert sizes of 300–500 bp was constructed and then sequenced on an Illumina HiSeq2500 platform (Illumina, San Diego, CA, USA) using the double terminal sequencing method (150 pair-ends).

4.2. Chloroplast Genome Assembly and Annotation

First, FastQC software was performed to evaluate the quality of Illumina paired-end raw reads [44], and low-quality reads were filtered. The remaining clean reads were used for assembly via NOVOPlasty [45] using *D. longan* chloroplast genome (GenBank: MG214255) [46] as the reference genome to generate the first version of rambutan genome. Next, all clean reads were mapped onto the first version genome and the mapped reads were assembled using SPAdes3.14.1 [47] and assembled contigs were corrected using the pair-end short reads from HiSeq2500 in Pilon version 1.23 (<https://github.com/broadinstitute/pilon>) [48] to generate the second version of rambutan chloroplast genome. These two versions were compared and mutually corrected to get the final complete rambutan chloroplast genome.

The chloroplast genome was annotated by the online program GeSeq (<https://chlorobox.mpimp-golm.mpg.de/geseq.html>) [49] and CPGAVAS2 [50]. Genome features like start/stop codons and intron/exon borders were manually corrected through the comparison of other reported Sapindaceae family chloroplast genomes. In addition, tRNA genes were identified by tRNAscan-SE 2.0 (<http://lowelab.ucsc.edu/tRNAscan-SE/>) [51]. A circular map of the revised annotated rambutan chloroplast genome was illustrated by using Organellar Genome DRAW (OGDRAW) (<https://chlorobox.mpimpgolm.mpg.de/OGDraw.html>) [52].

4.3. Chloroplast Genome Analysis

The simple sequence repeats (SSR) in nine chloroplast genome sequences of Sapindaceae (including *N. lappaceum*) (Table S1) were identified using the MISA online tool (<https://webblast.ipk-gatersleben.de/misa/>) [53] and the threshold settings were as follows: ten was applied to mononucleotide repeats, five to dinucleotide repeats and four to trinucleotide repeats, three for tetra-, penta-, and hexanucleotide repeats [54]. Repetitive sequences including forward, reverse, palindrome, and complement sequences were analyzed by REPuter program [12], and the parameter was set with the minimum length of repeat region set to 10bp and the minimum sequence identity set to 90%.

The expansion and contraction of the inverted repeat (IR) regions at junction sites from eight Sapindaceae family chloroplast genome sequences, including *Dimocarpus longan* (MG214255), *Litchi chinensis* (KY635881), *Pometia tomentosa* (MN106254), *Sapindus mukorossi* (KM454982), *Dodonaea viscosa* (MF155892), *Eurycorymbus cavaleriei* (MG813997), *Koelreuteria paniculata* (KY859413) and *Xanthoceras sorbifolium* (KY779850), were examined and plotted using IRscope online program (<https://irscope.shinyapps.io/irapp/>) [55]. Codon usage of the *N. lappaceum* chloroplast genome was analyzed via the GALAXY platform (<https://galaxy.pasteur.fr>) [56] with the CodonW online tool. Protein-coding genes of less than 300 nucleotides in length, and the repetitive gene sequences, were removed to reduce deviation of the results [57]. Finally, 53 CDS in *N. lappaceum* were selected for further codon usage analysis. Besides, putative RNA editing sites were predicted using the PREP-Cp web server (<http://prep.unl.edu/cgi-bin/cp-input.pl>) [58] with a cutoff value of 0.8.

4.4. Genome Comparison

We downloaded four whole chloroplast genome sequences of Sapindaceae family from the National Center for Biotechnology Information (NCBI) Organelle Genome and Nucleotide Resources database, including *D. longan* [46], *L. chinensis*, *P. tomentosa* [59], *S. mukorossi* [60], *D. viscosa* [61,62], *E. cavaleriei* [62], *K. paniculata* [63] and *X. sorbifolium* [64]. The mVISTA online program (Shuffle-LAGAN mode) [65,66] was used to compare chloro-

plast genome sequence of rambutan with the other species from Sapindaceae, in which the annotation of *D. longan* was the reference.

4.5. Positive Selection Analysis of Protein Sequence

We analyzed synonymous (Ks) and non-synonymous (Ka) substitution rates to investigate the molecular evolutionary process of the Sapindaceae family. The protein-coding genes of *N. lappaceum* were separately compared with eight closely related species in the Sapindaceae family: *D. longan*, *L. chinensis*, *P. tomentosa*, *S. mukorossi*, *D. viscosa*, *E. cavaleriei*, *K. paniculata* and *X. sorbifolium*, using ParaAT 2.0 [67], then the Ka/Ks value was calculated using KaKs_calculator 2.0 [68] with the NG method [69].

4.6. Phylogenetic Analysis

In order to deeply detect the evolutionary relationships of Sapindaceae family, we aligned 9 complete chloroplast genomes (including *N. lappaceum*) with MAFFT version 7 [70]. The best fitting nucleotide substitution model (TVM + I + G) was chosen by jModelTest v2.1.7 [71]. Phylogenetic analysis was then inferred by ML (maximum-likelihood) method based on the TVM + I + G substitution model in PAUP* 4.0 [72] with 1000 bootstrap replicates. *Anacardium occidentale* (KY635877) and *Mangifera indica* (KY635882) [73] in Anacardiaceae family were set as the outgroup.

5. Conclusions

We assembled the first complete chloroplast genome of rambutan using Illumina sequencing technology and compared its structure with other Sapindaceae species. The chloroplast genome of *N. lappaceum* exhibits similar quadripartite structure, gene order, and G + C content, when compared with other Sapindaceae chloroplast genomes. A total of 63 SSRs and 98 repeat sequences were identified in the *N. lappaceum* chloroplast genome. The research on codon usage of *N. lappaceum* shows that some amino acids have obvious codon usage bias and the codon preferences may help us to understand the evolution mechanisms of *N. lappaceum*. With PREP prediction, we detected 49 RNA editing loci in 18 protein-coding genes in *N. lappaceum*. Moreover, the expansion and contraction of the IR regions led to variations in the genome sizes of nine Sapindaceae chloroplasts. There are 6 genes (*ccsA*, *rpoA*, *rps12*, *psbJ*, *clpP* and *rps19*) that were detected with a Ka/Ks ratio >1, suggesting that these genes experienced positive selection in the evolution. Additionally, phylogenetic analysis using nine complete chloroplast genome sequences in Sapindaceae strongly supports the close relationships of *N. lappaceum* and *P. tomentosa* among sequenced chloroplast genomes in Sapindaceae.

Supplementary Materials: The following are available online at <https://www.mdpi.com/2223-7747/10/2/283/s1>, Table S1: The reference chloroplast genomes used in this study, Table S2: Gene composition in *N. lappaceum* chloroplast genome, Table S3: The introns and exons length of intron-containing genes in *N. lappaceum* chloroplast genome, Table S4: Total number of perfect simple sequence repeats (SSRs) identified within the chloroplast genome of *Nephelium lappaceum*, Table S5: The distribution pattern and number of simple sequence repeats (SSRs) identified within the chloroplast genome of Sapindaceae, Table S6: Long repeat sequences in the *N. lappaceum* chloroplast genome, Table S7: Codon usage for *N. lappaceum* chloroplast genome. Total: 21,434 codons, using 53 CDS, Table S8: RNA editing sites in the *Nephelium lappaceum* chloroplast genome, Table S9: Ka/Ks ratio between pairwise of species protein-coding sequences in nine Sapindaceae species.

Author Contributions: Conceptualization, R.M.; Data curation, W.Z.; Formal analysis, F.D.; Funding acquisition, R.M.; Investigation, F.D.; Methodology, F.D. and Z.L.; Project administration, W.Z.; Resources, F.D., Z.L., J.L., R.M. and W.Z.; Software, F.D. and J.L.; Supervision, R.M. and W.Z.; Validation, F.D.; Visualization, F.D. and J.L.; Writing—original draft, F.D.; Writing—review and editing, R.M. and W.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: All data generated or analyzed during this study are included in this published article.

Acknowledgments: This work was supported by a startup fund from Fujian Agriculture and Forestry University.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Lim, T.K. *Edible Medicinal and Non Medicinal Plants 2015*; Springer: Berlin, Germany, 2012.
- Palanisamy, U.D.; Cheng, H.M.; Masilamani, T.; Subramaniam, T.; Ling, L.T.; Radhakrishnan, A.K. Rind of the rambutan, *Nephelium lappaceum*, a potential source of natural antioxidants. *Food Chem.* **2008**, *109*, 54–63. [[CrossRef](#)]
- Zhuang, Y.; Ma, Q.; Guo, Y.; Sun, L. Protective effects of rambutan (*Nephelium lappaceum*) peel phenolics on H₂O₂-induced oxidative damages in HepG2 cells and d-galactose-induced aging mice. *Food Chem. Toxicol.* **2017**, *108*, 554–562. [[CrossRef](#)]
- Phuong NN, M.; Le, T.T.; Van Camp, J.; Raes, K. Evaluation of antimicrobial activity of rambutan (*Nephelium lappaceum* L.) peel extracts. *Int. J. Food Microbiol.* **2020**, *321*, 108539. [[CrossRef](#)]
- Harrington, M.G.; Edwards, K.J.; Johnson, S.A.; Chase, M.W.; Gadek, P.A. Phylogenetic Inference in Sapindaceae sensu lato Using Plastid *matK* and *rbcL* DNA Sequences. *Syst. Bot.* **2005**, *30*, 366–382. [[CrossRef](#)]
- Wicke, S.; Schneeweiss, G.M.; Depamphilis, C.W.; Kai, F.M.; Quandt, D. The evolution of the plastid chromosome in land plants: Gene content, gene order, gene function. *Plant Mol. Biol.* **2011**, *76*, 273–297.
- Bobik, K.; Burch-Smith, T.M. Chloroplast signaling within, between and beyond cells. *Front. Plant Sci.* **2015**, *6*, 781. [[CrossRef](#)]
- Wolfe, K.H.; Li, W.; Sharp, P.M. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *Proc. Natl. Acad. Sci. USA* **1987**, *84*, 9054–9058. [[CrossRef](#)]
- Palmer, J.D. Comparative Organization of Chloroplast Genomes. *Annu. Rev. Genet.* **1985**, *19*, 325–354. [[CrossRef](#)]
- Shinozaki, K.; Ohme, M.; Tanaka, M.; Wakasugi, T.; Sugiura, M. The complete nucleotide sequence of the tobacco chloroplast genome: Its gene organization and expression. *Plant Mol. Biol. Rep.* **1986**, *5*, 2043–2049. [[CrossRef](#)]
- Li, C.; Lin, F.; An, D.; Wang, W.; Huang, R. Genome Sequencing and Assembly by Long Reads in Plants. *Genes* **2017**, *9*, 6. [[CrossRef](#)]
- Kurtz, S.; Choudhuri, J.V.; Ohlebusch, E.; Schleiermacher, C.; Stoye, J.; Giegerich, R. REPuter: The manifold applications of repeat analysis on a genomic scale. *Nucleic Acids Res.* **2001**, *29*, 4633–4642. [[CrossRef](#)]
- Marek, A.; Tomala, K. The contribution of purifying selection, linkage, and mutation bias to the negative correlation between gene expression and polymorphism density in yeast populations. *Genome Biol. Evol.* **2018**, *10*, 2986–2996. [[CrossRef](#)]
- Nguyen Dinh, S.; Sai, T.Z.T.; Nawaz, G.; Lee, K.; Kang, H. Abiotic stresses affect differently the intron splicing and expression of chloroplast genes in coffee plants (*Coffea arabica*) and rice (*Oryza sativa*). *J. Plant Physiol.* **2016**, *201*, 85–94. [[CrossRef](#)]
- Mirzaei, S.; Mansouri, M.; Mohammadi-Nejad, G.; Sablok, G. Comparative assessment of chloroplast transcriptional responses highlights conserved and unique patterns across Triticeae members under salt stress. *Photosynth. Res.* **2017**, *136*, 357–369. [[CrossRef](#)]
- Naver, H.; Boudreau, E.; Rochaix, J.D. Functional studies of *YCF3*: Its role in assembly of photosystem I and Interactions with some of its subunits. *Plant Cell* **2002**, *13*, 2731–2745. [[CrossRef](#)]
- Boudreau, E.; Takahashi, Y.; Lemieux, C.; Turmel, M.; Rochaix, J.D. The chloroplast *ycf3* and *ycf4* open reading frames of *Chlamydomonas reinhardtii* are required for the accumulation of the photosystem I complex. *Embo. J.* **1997**, *16*, 6095–6104. [[CrossRef](#)]
- Clarke, A.K.; Schelin, J.; Porankiewicz, J. Inactivation of the *clpP1* gene for the proteolytic subunit of the ATP-dependent Clp protease in the cyanobacterium *Synechococcus* limits growth and light acclimation. *Plant. Mol. Biol.* **1998**, *37*, 791–801. [[CrossRef](#)]
- Bruce, C.A.; Cunningham, K.A.; Stern, D.B. The plastid *clpP* gene may not be essential for plant cell viability. *Plant. Cell Physiol.* **2003**, *44*, 93–95.
- Varshney, R.K.; Sigmund, R.; Borner, A.; Korzun, V.; Stein, N.; Sorrells, M.E.; Langridge, P.; Graner, A. Interspecific transferability and comparative mapping of barley EST-SSR markers in wheat, rye and rice. *Plant. Sci.* **2005**, *168*, 195–202. [[CrossRef](#)]
- Yang, A.; Zhang, J.; Tian, H.; Yao, X. Characterization of 39 novel EST-SSR markers for *Liriodendron tulipifera* and cross-species amplification in *L. chinense* (Magnoliaceae). *Am. J. Bot.* **2012**, *99*, e460–e464. [[CrossRef](#)]
- Li, B.; Lin, F.; Huang, P.; Guo, W.; Zheng, Y. Development of nuclear SSR and chloroplast genome markers in diverse *Liriodendron chinense* germplasm based on low-coverage whole genome sequencing. *Biol. Res.* **2020**, *53*, 21. [[CrossRef](#)]
- Gao, B.; Yuan, L.; Tang, T.; Hou, J.; Pan, K.; Wei, N. The complete chloroplast genome sequence of *Alpinia oxyphylla* Miq. and comparison analysis within the Zingiberaceae family. *PLoS ONE* **2019**, *14*. [[CrossRef](#)]
- Yan, C.; Du, J.; Gao, L.; Li, Y.; Hou, X. The complete chloroplast genome sequence of watercress (*Nasturtium officinale* R. Br.): Genome organization, adaptive evolution and phylogenetic relationships in Cardamineae. *Gene* **2019**, *699*, 24–36. [[CrossRef](#)]
- Cavalier-Smith, T. Chloroplast Evolution: Secondary Symbiogenesis and Multiple Losses. *Curr. Biol.* **2002**, *12*, R62–R64. [[CrossRef](#)]
- Hershberg, R.; Petrov, D.A. Selection on codon bias. *Annu. Rev. Genet.* **2008**, *42*, 287–299. [[CrossRef](#)]

27. Raman, G.; Park, S.; Lee, E.M.; Park, S.J. Evidence of mitochondrial DNA in the chloroplast genome of *Convallaria keiskei* and its subsequent evolution in the Asparagales. *Sci. Rep.* **2019**, *9*, 5028. [CrossRef]
28. Purabi, M.; Rofinayasmin, B.O.; Katharina, M.; Ramakrishnan, N.; Jennifer, A.H. Codon usage and codon pair patterns in non-grass monocot genomes. *Ann. Bot.* **2017**, *120*, 893–909.
29. Wang, W.; Yu, H.; Wang, J.; Lei, W.; Gao, J.; Qiu, X.; Wang, J. The Complete Chloroplast Genome Sequences of the Medicinal Plant *Forsythia suspensa* (Oleaceae). *Int. J. Mol. Sci.* **2017**, *18*, 2288. [CrossRef]
30. Smith, H.C.; Gott, J.M.; Hanson, M.R. A guide to RNA editing. *RNA-Publ. RNA Soc.* **1997**, *3*, 1105–1123.
31. Hoch, B. Editing of a chloroplast mRNA by creation of an initiation codon. *Nature* **1991**, *353*, 178–180. [CrossRef] [PubMed]
32. Maier, R.M.; Zeltz, P.; Kossel, H.; Bonnard, G.; Gualberto, J.M.; Grienenberger, J.M. RNA editing in plant mitochondria and chloroplasts. *Plant. Mol. Biol.* **1996**, *32*, 343–365. [CrossRef] [PubMed]
33. Schmitzlinneweber, C.; Barkan, A. RNA splicing and RNA editing in chloroplasts. *Top. Curr. Genet.* **2007**, *19*, 213–248.
34. Shikanai, T. RNA editing in plant organelles: Machinery, physiological function and evolution. *Cell. Mol. Life Sci.* **2006**, *63*, 698–708. [CrossRef] [PubMed]
35. Dong, W.; Xu, C.; Li, C.; Sun, J.; Zuo, Y.; Shi, S.; Cheng, T.; Guo, J.; Zhou, S. *ycf1*, the most promising plastid DNA barcode of land plants. *Sci. Rep.* **2015**, *5*, 8348. [CrossRef]
36. Neubig, K.M.; Whitten, W.M.; Carlsward, B.S.; Blanco, M.A.; Endara, L.; Williams, N.H.; Moore, M. Phylogenetic utility of *ycf1* in orchids: A plastid gene more variable than *matK*. *Plant. Syst. Evol.* **2009**, *277*, 75–84. [CrossRef]
37. Dugas, D.V.; Hernandez, D.; Koenen, E.J.M.; Schwarz, E.; Straub, S.; Hughes, C.E.; Jansen, R.K.; Nageswara-Rao, M.; Staats, M.; Trujillo, J.T. Mimosoid legume plastome evolution: IR expansion, tandem repeat expansions, and accelerated rate of evolution in *clpP*. *Sci. Rep.* **2015**, *5*, 16958. [CrossRef]
38. Yu, X.; Tan, W.; Zhang, H.; Gao, H.; Tian, X. Complete Chloroplast Genomes of *Ampelopsis humulifolia* and *Ampelopsis japonica*: Molecular Structure, Comparative Analysis, and Phylogenetic Analysis. *Plants* **2019**, *8*, 410. [CrossRef]
39. Yang, Z.; Bielawski, J.P. Statistical methods for detecting molecular adaptation. *Trends Ecol. Evol.* **2000**, *15*, 496–503. [CrossRef]
40. Swanson, W.J.; Wong, A.; Wolfner, M.F.; Aquadro, C.F. Evolutionary Expressed Sequence Tag Analysis of *Drosophila* Female Reproductive Tracts Identifies Genes Subjected to Positive Selection. *Genetics* **2004**, *168*, 1457–1465. [CrossRef]
41. Gitzendanner, M.A.; Soltis, P.S.; Wong, G.K.S.; Ruhfel, B.R.; Soltis, D.E. Plastid phylogenomic analysis of green plants: A billion years of evolutionary history. *Am. J. Bot.* **2018**, *105*. [CrossRef]
42. Du, Y.P.; Bi, Y.; Yang, F.P.; Zhang, M.F.; Zhang, X.H. Complete chloroplast genome sequences of *Lilium*: Insights into evolutionary dynamics and phylogenetic analyses. *Sci. Rep.* **2017**, *7*, 1–10. [CrossRef]
43. Porebski, S.; Bailey, L.G.; Baum, B.R. Modification of a CTAB DNA extraction protocol for plants containing high polysaccharide and polyphenol components. *Plant. Mol. Biol. Rep.* **1997**, *15*, 8–15. [CrossRef]
44. Andrews, S. FastQC A Quality Control Tool for High Throughput Sequence Data. In Babraham Institute. 2015. Available online: <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/> (accessed on 1 December 2020).
45. Nicolas, D.; Patrick, M.; Guillaume, S. NOVOPlasty: De novo assembly of organelle genomes from whole genome data. *Nucleic Acids Res.* **2017**, *4*, e18.
46. Wang, K.; Li, L.; Zhao, M.; Li, S.; Sun, H.; Lv, Y.; Wang, Y. Characterization of the complete chloroplast genome of longan (*Dimocarpus longan* Lour.) using illumina paired-end sequencing. *Mitochondrial DNA Part. B* **2017**, *2*, 904–906. [CrossRef] [PubMed]
47. Bankevich, A.; Nurk, S.; Antipov, D.; Gurevich, A.A.; Dvorkin, M.; Kulikov, A.S.; Lesin, V.M.; Nikolenko, S.I.; Pham, S.; Pribelski, A.D. SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **2012**, *19*, 455–477. [CrossRef] [PubMed]
48. Walker, B.J.; Abeel, T.; Shea, T.; Priest, M.; Abouelliel, A.; Sakthikumar, S.; Cuomo, C.A.; Zeng, Q.; Wortman, J.R.; Young, S. Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS ONE* **2014**, *9*, e112963. [CrossRef]
49. Michael, T.; Pascal, L.; Tommaso, P.; Ulbricht-Jones, E.S.; Axel, F.; Ralph, B.; Stephan, G. GeSeq – versatile and accurate annotation of organelle genomes. *Nucleic Acids Res.* **2017**, *45*, W6–W11.
50. Shi, L.; Chen, H.; Jiang, M.; Wang, L.; Wu, X.; Huang, L.; Liu, C. CPGAVAS2, an integrated plastome sequence annotator and analyzer. *Nucleic Acids Res.* **2019**, *47*, W65–W73. [CrossRef]
51. Chan, P.P.; Lowe, T.M. tRNAscan-SE: Searching for tRNA Genes in Genomic Sequences. *Methods Mol. Biol.* **2019**, *1962*, 1–14.
52. Lohse, M.; Drechsel, O.; Bock, R. OrganellarGenomeDRAW (OGDRAW): A tool for the easy generation of high-quality custom graphical maps of plastid and mitochondrial genomes. *Curr. Genet.* **2007**, *52*, 267–274. [CrossRef]
53. Beier, S.; Thiel, T.; Munch, T.; Scholz, U.; Mascher, M. MISA-web: A web server for microsatellite prediction. *Bioinformatics* **2017**, *33*, 2583–2585. [CrossRef] [PubMed]
54. Li, Q.; Wan, J.M. SSRHunter: Development of a local searching software for SSR sites. *Yi Chuan* **2005**, *27*, 808–810. [PubMed]
55. Amiryousefi, A.; Hyvonen, J.; Poczai, P. IRscope: An online program to visualize the junction sites of chloroplast genomes. *Bioinformatics* **2018**, *34*, 3030–3031. [CrossRef] [PubMed]
56. Afgan, E.; Baker, D.; Den Beek, M.V.; Blankenberg, D.; Bouvier, D.; Cech, M.; Chilton, J.; Clements, D.; Coraor, N.; Eberhard, C. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. *Nucleic Acids Res.* **2016**, *44*, W3–W10. [CrossRef]

57. Wright, F. The effective number of codons used in a gene. *Gene* **1990**, *87*, 23–29. [[CrossRef](#)]
58. Mower, J.P. The PREP suite: Predictive RNA editors for plant mitochondrial genes, chloroplast genes and user-defined alignments. *Nucleic Acids Res.* **2009**, *37*, 253–259. [[CrossRef](#)]
59. Wang, Y.; Yuan, X.; Zhang, J. The complete chloroplast genome sequence of *Pometia tomentosa*. *Mitochondrial DNA Part B* **2019**, *4*, 3950–3951. [[CrossRef](#)]
60. Yang, B.; Li, M.; Ma, J.; Fu, Z.; Tian, J. The complete chloroplast genome sequence of *Sapindus mukorossi*. *Mitochondrial DNA Part A* **2016**, *27*, 1825–1826. [[CrossRef](#)]
61. Saina, J.K.; Gichira, A.W.; Li, Z.Z.; Hu, G.W.; Wang, Q.F.; Liao, K. The complete chloroplast genome sequence of *Dodonaea viscosa*: Comparative and phylogenetic analyses. *Genetica* **2017**, *146*, 101–113. [[CrossRef](#)]
62. Du, X.; Xin, G.; Ren, X.; Liu, H.; Hao, N.; Jia, G.; Liu, W. The complete chloroplast genome of *Eurycorymbus cavaleriei* (Sapindaceae), a Tertiary relic species endemic to China. *Conserv. Genet. Resour.* **2019**, *11*, 283–285. [[CrossRef](#)]
63. Kim, S.C.; Baek, S.H.; Hong, K.N.; Lee, J.W. Characterization of the complete chloroplast genome of *Koelreuteria paniculata* (Sapindaceae). *Conserv. Genet. Resour.* **2018**, *10*, 69–72. [[CrossRef](#)]
64. Chen, S.Y.; Zhang, X.Z. Characterization of the complete chloroplast genome of *Xanthoceras sorbifolium*, an endangered oil tree. *Conserv. Genet. Resour.* **2017**, *9*, 595–598. [[CrossRef](#)]
65. Frazer, K.A.; Pachter, L.; Poliakov, A.; Rubin, E.M.; Dubchak, I. VISTA: Computational tools for comparative genomics. *Nucleic Acids Res.* **2004**, *32*, W273–W279. [[CrossRef](#)] [[PubMed](#)]
66. Brudno, M.; Malde, S.; Poliakov, A.; Do, C.B.; Couronne, O.; Dubchak, I.; Batzoglou, S. Glocal alignment: Finding rearrangements during alignment. *Bioinformatics* **2003**, *19*, 54–62. [[CrossRef](#)]
67. Zhang, Z.; Xiao, J.; Wu, J.; Zhang, H.; Liu, G.; Wang, X.; Dai, L. ParaAT: A parallel tool for constructing multiple protein-coding DNA alignments. *Biochem. Biophys. Res. Commun.* **2012**, *419*, 779–781. [[CrossRef](#)]
68. Wang, D.; Zhang, Y.; Zhang, Z.; Zhu, J.; Yu, J. KaKs_Calculator 2.0: A Toolkit Incorporating Gamma-Series Methods and Sliding Window Strategies. *Genom. Proteom. Bioinform.* **2010**, *8*, 77–80. [[CrossRef](#)]
69. Nei, M.; Gojobori, T. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol. Biol. Evol.* **1986**, *3*, 418–426.
70. Katoh, K.; Rozewicki, J.; Yamada, K.D. MAFFT online service: Multiple sequence alignment, interactive sequence choice and visualization. *Brief. Bioinform.* **2019**, *20*, 1160–1166. [[CrossRef](#)]
71. Darriba, D.; Taboada, G.L.; Doallo, R.; Posada, D. jModelTest 2: More models, new heuristics and parallel computing. *Nat. Methods* **2012**, *9*, 772. [[CrossRef](#)]
72. Cummings, M.P. PAUP* (Phylogenetic Analysis Using Parsimony (and Other Methods)). In *Dictionary of Bioinformatics Computational Biology*; John Wiley & Sons Inc.: Hoboken, NJ, USA, 2004.
73. Azim, M.K.; Khan, I.A.; Zhang, Y. Characterization of mango (*Mangifera indica* L.) transcriptome and chloroplast genome. *Plant. Mol. Biol.* **2014**, *85*, 193–208. [[CrossRef](#)]