Comparative Survey of Morphological Variations and Plastid Genome Sequencing Reveals Phylogenetic Divergence between Four Endemic *Ilex* Species

Tao Su [1,2] , **Mengru Zhang** [1,2], **Zhenyu Shan** [1], **Xiaodong Li** [1], **Biyao Zhou** [1,2], **Han Wu** [1] and **Mei Han** [1,*]

[1] Co-Innovation Center for Sustainable Forestry in Southern China, College of Biology and the Environment, Nanjing Forestry University, Nanjing 210037, China; tao.su@cos.uni-heidelberg.de (T.S.); zhangmengru@njfu.edu.cn (M.Z.); njfuszy1007@163.com (Z.S.); xiaodong.li@chinahaisheng.com (X.L.); zhoubiyao@njfu.edu.cn (B.Z.); wh010510@163.com (H.W.)

[2] Key Laboratory of State Forestry Administration on Subtropical Forest Biodiversity Conservation, Nanjing Forestry University, Nanjing 210037, China

* Correspondence: sthanmei@njfu.edu.cn; Tel.: +86-158-9598-9551

✓ check for updates

Abstract: Holly (*Ilex* L.), from the monogeneric Aquifoliaceae, is a woody dioecious genus cultivated as pharmaceutical and culinary plants, ornamentals, and industrial materials. With distinctive leaf morphology and growth habitats, but uniform reproductive organs (flowers and fruits), the evolutionary relationships of *Ilex* remain an enigma. To date, few contrast analyses have been conducted on morphology and molecular patterns in *Ilex*. Here, the different phenotypic traits of four endemic *Ilex* species (*I. latifolia*, *I. suaveolens*, *I. viridis*, and *I. micrococca*) on Mount Huangshan, China, were surveyed through an anatomic assay and DNA image cytometry, showing the unspecified link between the examined morphology and the estimated nuclear genome size. Concurrently, the newly-assembled plastid genomes in four *Ilex* have lengths ranging from 157,601 bp to 157,857 bp, containing a large single-copy (LSC, 87,020–87,255 bp), a small single-copy (SSC, 18,394–18,434 bp), and a pair of inverted repeats (IRs, 26,065–26,102 bp) regions. The plastid genome annotation suggested the presence of numerable protein-encoding genes (89–95), transfer RNA (tRNA) genes (37–40), and ribosomal RNA (rRNA) genes (8). A comprehensive comparison of plastomes within eight *Ilex* implicated the conserved features in coding regions, but variability in the junctions of IRs/SSC and the divergent hotspot regions potentially used as the DNA marker. The *Ilex* topology of phylogenies revealed the incongruence with the traditional taxonomy, whereas it informed a strong association between clades and geographic distribution. Our work herein provided novel insight into the variations in the morphology and phylogeography in Aquifoliaceae. These data contribute to the understanding of genetic diversity and conservation in the medicinal *Ilex* of Mount Huangshan.

Keywords: *Ilex* species; Aquifoliaceae; morphological traits; DNA C-value; plastid genome

1. Introduction

The *Ilex* L. (holly) is the only living woody dioecious angiosperm genus, accounting for approximately 700 species within the monogeneric family of Aquifoliaceae [1]. The *Ilex* species are evergreen and deciduous trees, prostrate shrubs, and climbers with a broad distribution from tropics to temperate regions [2,3]. Over 200 species have been documented in the center of East Asia and South America, whereas several species can grow in Europe, tropical Africa, and northern Australia [4]. The unexpected spreading in the oceanic islands as patchy populations prompted the assumption of unspecialized pollination and efficient seed dispersal by birds [5]. Recently, a report in

phylogeny implicated the origin of the Aquifoliaceae as subtropical lowlands with a mesic climate type. This study indicated that the eternal existence of *Ilex* species in humid and warm subtropical monsoon forests of southern China could trace back to the middle Eocene [6,7].

For tens of decades, various *Ilex* species have been used as herbal infusions in traditional Chinese medicine. Some large-leaved species of *Ilex* (e.g., *I. kudingcha* and *I. latifolia*) are processed by brewing to produce a bitter-tasting "Kuding" tea, used for its nutritional merit and nourishment for the daily health stimulation, consumption, and pharmaceutical drugs [8,9]. A traditional Chinese herb, *I. pubescens*, is used particularly to treat cardiovascular diseases [10]. Originating from southern areas of South America, *I. paraguariensis* is consumed commercially as a popular beverage, known as "yerba mate" in America and the Middle East, highly appreciated for its peculiar flavor and health maintenance effects [11,12]. An alternative mate tea, *I. dumosa*, has a similar taste and mild effects; it is available to satisfy the different consumers as a low-caffeine and xanthine-containing substitute [13]. As the natural resources for pharmaceuticals and dietary food, progress in phytochemistry and pharmacology has led to discoveries of terpenoids, saponins, flavonoids, glycosides, amino acids, and other bioactive compounds in many additional *Ilex* species, reflecting the worldwide economic, medicinal, and clinical value of *Ilex* [13–15]. Some larger native *Ilex* species have been developed for timber production and garden use. Other holly species (e.g., *I. aquifolium*, *I. opaca*, and *I. crenata*) retaining the green foliage and bright red drupes are grown as traditional Christmas decorations and ornamentals during the winter holidays [16,17].

Enabling the estimation of absolute genome size, flow cytometry (FCM) has provided relevant clues for taxonomy in various plant species [18]. Independent contrast studies suggested a significant relationship between the phenotypic traits (e.g., cell size and stomata density) and the genome size in angiosperms [19]. However, a proposed correlation between cellular architecture and organelle DNA content remains poorly understood [20]. As all plastids possess the same DNA and a few functional features, the plastid genome comparison between different species of plant branches, integrating basic backgrounds with gene content and bar codes, was significantly relevant to the understanding of the evolution of plastid DNA and the adapted ancient environments [21]. The plastid phylogeny offers an alternative strategy for the investigation of phylogeography due to its maternal inheritance and conserved circular structure as well as small sufficient population size [22]. Increasing evidence suggests that plastid genomes provide more suitable material for species identification and conservation, enabling a significant improvement in the resolution of branches under the frame of the nuclear phylogenies [23,24]. To date, emerging advances in sequencing technologies have increased the availability of plastid genomes to approximately 3000 in the GenBank database (https://www.ncbi.nlm.nih.gov/genbank/). Therefore, contrasting analyses of the plastid genome and nuclear genome using the molecular genetic methodologies may provide the theoretical bases for the phylogenetic reconstruction and exploration of genetic diversity and plant systematics.

The unique subtropical monsoon pattern in Mount Huangshan (Anhui, China) led to high forest diversity and broad distribution of evergreen broad-leaved forest [25]. Based on a dynamic field survey in a large-scale forest plot, a total of 12 documented endemic *Ilex* species in Aquifoliaceae were found to be predominant [26]. However, comparative reports are lacking on the morphology traits in combination with molecular and genomic patterns within the *Ilex* genus. In this work, the four *Ilex* species were chosen to conduct a contrast analysis in terms of their phenotypic variation, DNA-C values, and the newly assembled plastid genomes. The main objectives were: (1) to investigate the possible relationships between the morphological traits and nuclear genome size between the four selected *Ilex* species, (2) to discover the variation and highly divergent regions of the plastome that could be used in the classification and identification of various *Ilex* species, and (3) to analyze the plastid genome structures and reconstruct the plastome-derived phylogenetic relationships of *Ilex* species in Aquifoliaceae.

2. Materials and Methods

2.1. Plant Materials

The respective organs (e.g., leaves and seeds) of *I. latifolia*, *I. suaveolens*, *I. viridis*, and *I. micrococca* were harvested from 10.24 ha (320 m × 320 m) forest plot (30°8′26″ N, 118°6′38″ E) in Mount Huangshan, Anhui, China [26]. The plot ranges altitude from 430 to 565 m with an annual average temperature of 7.8 °C and annual precipitation of 2394.5 mm. The voucher specimens of four *Ilex* species (accession numbers: YL20190417014, YL20190417015, YL20190417016, and YL20190417017) were preserved in the herbarium of Nanjing Forestry University, Nanjing, China.

2.2. Phenotype Quantification and Determination of DNA Content

For each species, more than 60 healthy mature leaves and 300 mature seeds of 5 independent trees of similar age were randomly sampled for the morphological analyses. In total, thirty anatomic sections of epidermis were prepared from the cut leaves (5 mm × 5 mm) macerated by H_2O_2-HAC solutions [27]. To quantitate the size of the upper leaf epidermal cell (LEC), the stomata aperture (STA), and the stomata density (STD), twenty visual fields were captured using the same scale during optical microscopy. The leaf area (LA) was measured from 50 randomly selected leaves by Image J v1.53c (https://imagej.nih.gov/ij/) after scanning with Expression 11000XL (EPSON, Beijing, China). The specific leaf area (SLA) was calculated based on the ratio of leaf area to leaf dry mass. We used 100 air-dried seeds to determine the seed weight (SW). The variance of three perpendicular seed dimensions (VSD) was calculated using the average of 50 seeds according to a previous report [28]. Spherical seeds have a variance of 0, and elongated or flattened seeds have a variance of up to 0.33. Flower size (FS) was measured from the average diameters of 12 female flowers. The significance of phenotypic variation between four *Ilex* samples was statically analyzed using SPSS 24.0 (https://www.ibm.com) based on the ANOVA ($p < 0.05$) and Duncan's multiple range tests.

For the determination of the DNA C-value, the young leaves were chopped to isolate the crude nuclear DNA with the addition of woody plant buffer followed by RNase digest. DNA staining of propidium iodide (PI) and FCM analysis were performed based on a previous report [29]. Together with the internal standard (*Solanum lycopersicum*, 2C = 2.00 pg), the resulting suspensions were analyzed with BD Influx^{TM} cell sorter (BD, Piscataway, NJ, USA). The histograms of FCM were generated by the software BD FACS^{TM} 1.0.0.650. The coefficients of variation (CV) of DNA peaks below 5% were considered as reliable. The chromosome numbers in four diploid *Ilex* species were retrieved from the IPCN database (http://legacy.tropicos.org/Project/IPCN). The genome size (DNA C-value or the haploid DNA content) was calibrated by multiplying the standard by the ratio of the mean fluorescent intensity of each sample to that of the standard [30]. The DNA-C value is represented as means ± standard error (±SE) of at least four independent biological replicates.

2.3. Plastome Sequencing, Assembly, Annotation, Codon Usage, and Repeat Analyses

The fresh leaves of *I. suaveolens*, *I. viridis*, and *I. micrococca* were harvested and flash-frozen in liquid nitrogen. DNA was extracted using a method modified from a previous report [31]. The plastid genome sequencing data of *I. latifolia* were obtained from the NCBI database (https://www.ncbi.nlm.nih.gov/) based on the new accession MN688228 [32]. The next-generation sequencing of the whole-plastid genomes was performed by Biodata Biotechnologies Inc. (Hefei, China) for *I. suaveolens*, *I. viridis*, and *I. micrococca* on the BGISEQ-500 platform (BGI, Shenzhen, China). Approximately 50 MB of high-quality clean paired-end reads was generated, and the filtered sequences were assembled by SPAdes assembler 3.14.1 (http://cab.spbu.ru/software/spades/) with default parameters [33]. The genome was annotated using the CPGAVAS (http://47.96.249.172:16014/analyzer/home) program [34] and checked further by DOGMA (https://dogma.ccbb.utexas.edu/) following with nBLAST searches in NCBI based on the reference genome to identify the specific genes [1]. The circular graphical maps of the plastome were drawn using the program OGDRAWv1.3.1 (https://chlorobox.mpimp-golm.mpg.de/OGDraw.html) [35]. The complete

plastome sequences of *I. suaveolens*, *I. viridis*, and *I. micrococca* were submitted to the GenBank with accession numbers MN830249, MN830250, and MN830251, respectively. The relative synonymous codon usage for all protein-encoding genes was analyzed using MEGA X (https://www.megasoftware.net/) [36]. The use frequency of amino acids was calculated by the percentage of the codons divided by the total codons. The simple sequence repeats (SSRs) were analyzed by MISA (http://pgrc.ipk-gatersleben.de/misa/) [37]. The parameters of categorized SSRs (mono-, di-, tri-, tetra-, pena-, and hexanucleotide) sequence length were set up with a minimum number of repeats of 8, 5, 4, 3, 3, and 3, respectively. The long repeat sequences, including the forward, palindrome, reverse, and complement repeats, were analyzed using the online program REPuter (https://bibiserv.cebitec.uni-bielefeld.de/reputer) with adjusted parameters [38].

2.4. Plastome Divergence and Phylogenetic Analyses

Using the annotation of *I. suaveolens* as the reference, we compared of the entire plastid genomes of eight *Ilex* species in Aquifoliaceae using the program GView (https://www.gview.ca/wiki/GView/WebHome) [39] and the mVISTA program (http://genome.lbl.gov/vista/mvista/submit.shtml) in the Shuffle-LAGAN mode [40,41]. For the inverted repeat (IR) expansion and contraction of border genes, eight *Ilex* plastomes were aligned to analyze the variations in the junctions of LSC, IRs, and SSC using IRscope (https://irscope.shinyapps.io/irapp/) [42]. In total, nineteen plastid genome sequences (15 *Ilex* species) were retrieved from the GenBank. *Populus trichocarpa*, *Populus deltoides*, *Quercus acutissima*, and *Helwingia himalaica* were used as outgroups. The multiple sequences were aligned using MAFFT v7.471 (https://mafft.cbrc.jp/alignment/server/index.html) [43]. The phylogenetic topology was constructed using the software MEGA X software by the methods of maximum likelihood (ML) and maximum parsimony (MP) using the nucleotide substitution model of Tamura-Neighbour. The bootstrap values are shown on the branches of the phylogenetic tree based on 1000 replicates.

3. Results

3.1. Variation of the Morphological Trait and Nuclear DNA Content

To investigate the variation in morphology between the four *Ilex* species, we initially performed a comparative analysis of phenotypic traits of the major vegetative and reproductive organs, including LA, LEC, SLA, STA, STD, SW, VSD, and FS (Table 1). Based on the anatomic analyses, the significant variation in LA was observed in all *Ilex* species, with *I. latifolia* having the highest value (79.19 cm^2) (Table 1). Significant differences were also observed in SLA, ranging the values from 40.88 cm^2/g (*I. latifolia*) to 165.4 cm^2/g (*I. micrococca*). Regarding the LEC, the highest value (938.65 μm^2) was found in *I. suaveolens*; however, the value varied insignificantly as *I. viridis* and *I. micrococca* have a similar-sized LEC (Figure 1b). Image analyses of stomata-related traits revealed that STD is drastically different in four *Ilex* species, and *I. micrococca* has the highest density (257.64/mm^2; Table 1 and Figure 1c). The significantly different values in SLA appeared to be related to the STD values in the four *Ilex* species, as both traits showed similar patterns (Table 1). Analyses of VSD indicated that *I. suaveolens* seeds are closer to spherical shape (0.1137). However, for STA, VSD, and FS, significantly statistical distinctions were not found within the four *Ilex* species.

To determine the DNA-C value of four *Ilex* species by FCM, *S. lycopersicum* was used as the internal standard to calculate the genome size [44]. The DNA-associated fluorescence on FCM histograms showed that the CV values for G0/G1 peaks were between 2.92% and 4.67% (Figure 1d). The low CV (< 5%) indicated a constant quality considered reliable for FCM assessments [45]. The inspection of the FCM fluorescent peak revealed that *I. micrococca* has the most abundant nuclear DNA content (3.053 pg), followed by *I. viridis* (2.519 pg). *I. suaveolens* (2.242 pg) and *I. latifolia* (1.910 pg) exhibited similar levels of DNA 2C-value. Further calculation of nuclear genome size (NG) revealed that the mean levels vary with a range from 955 Mb (*I. latifolia*) to 1493 Mb (*I. micrococca*) in the four *Ilex* species (Table 1).

Table 1. Comparison of the morphological traits between four subtropical *Ilex* species.

Traits	*I. latifolia* Thunb.	*I. suaveolens* (H. Lév.) Loes.	*I. viridis* Champ.ex Benth.	*I. micrococca* Maxim.
LA (cm^2)	79.19 ± 2.24 [a]	19.81 ± 1.84 [c]	8.89 ± 0.89 [d]	32.61 ± 1.26 [b]
SLA (cm^{-2}/g)	40.88 ± 2.42 [d]	93.02 ± 3.47 [c]	100.81 ± 3.91 [b]	165.4 ± 3.62 [a]
LEC (μm^2)	848.21 ± 94 [b]	938.65 ± 103 [a]	732.42 ± 85 [cd]	734.21 ± 81 [c]
STA (μm)	32.76 ± 0.35 [a]	30.55 ± 0.29 [c]	31.73 ± 0.29 [b]	32.65 ± 0.27 [ab]
STD (n/mm^2)	118.31 ± 17 [d]	216.90 ± 12 [c]	232.39 ± 11 [b]	257.64 ± 29 [a]
SW (g/100)	6.29 ± 0.46 [b]	3.04 ± 0.36 [c]	8.35 ± 0.69 [a]	2.12 ± 0.10 [d]
VSD	0.083 ± 0.03 [b]	0.114 ± 0.02 [a]	0.118 ± 0.01 [a]	0.136 ± 0.01 [a]
FS (mm)	5.20 ± 0.13 [cd]	7.10 ± 0.21 [a]	5.50 ± 0.11 [bc]	6.50 ± 0.17 [b]
DNA 2C (pg)	1.910 ± 0.021	2.242 ± 0.022	2.519 ± 0.038	3.053 ± 0.047
NG (≈Mb)	955	1121	1232	1493

LA, leaf area; LEC, upper leaf epidermal cells; SLA, specific leaf area; STA, stomata aperture; STD, stomata density; SW, weight of 100 seeds; VSD, variance of seed dimensions; FS, flower size in diameter; NG, nuclear genome size. The data represent mean values ± SE of the independent biological replicates (see materials and methods). According to Duncan's multiple range tests, the superscripted letters indicate statistical significance per ANOVA ($p < 0.05$) between four *Ilex* species.
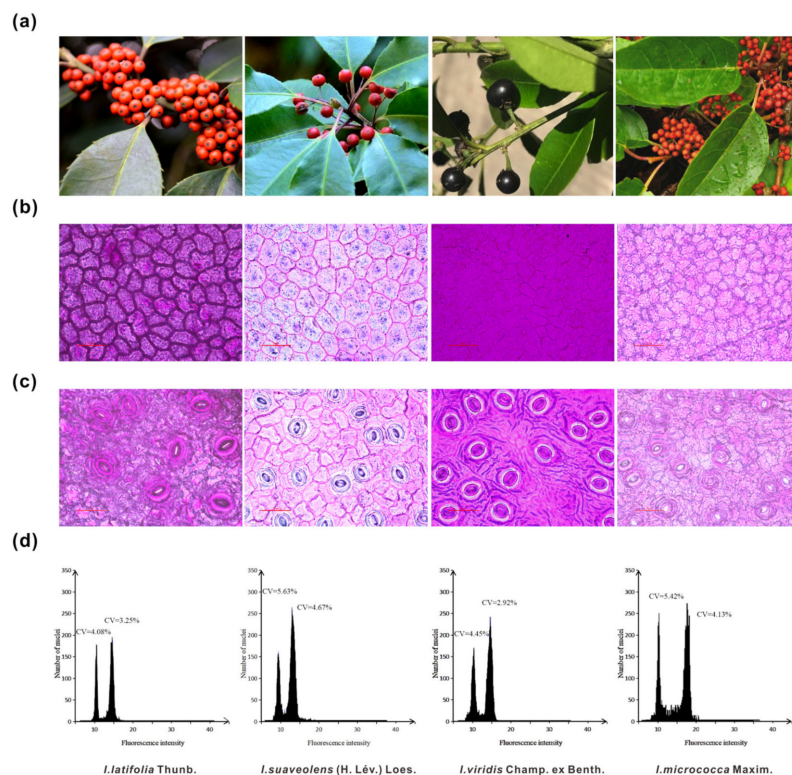

Figure 1. Analyses of the morphological traits and the fluorescent histograms of DNA content in four *Ilex* species. (**a**) The phenotypes of leaves and fruits; (**b**) LEC images; (**c**) STA and STD images; (**d**) flow cytometry (FCM) histograms obtained from leaves of four *Ilex* species. Nuclear DNA was stained with PI, using tomato (*S. lycopersicum*) as the internal standard.

3.2. The Plastid Genome Features and Sequence Divergence

The de novo sequencing and reference-guided assembling of the plastid genome of four *Ilex* species revealed a double-stranded circular DNA with a range of length from 157,601 (*I. latifolia*) to 157,857 bp (*I. suaveolens*) (Figure 2). The typical combined quadripartite structures of the plastid genome contain two inverted repeats (IR) regions, IRA and IRB, between 26,065 and 26,102 bp, each separated by a large single-copy (LSC, 87,020–87,255 bp) and a small single-copy (SSC, 18,394–18,434 bp) sections. Except for *I. latifolia*, the plastomes encode 134 genes comprised of 89 putative protein-encoding (PE)

genes, 37 transfer RNA (tRNA) genes, and 8 ribosomal RNA (rRNA) genes in the other three *Ilex* species (Table 2). Among the annotated genes in *Ilex* plastome, eight PE genes (*rps7*, *ndhB*, *ycf15*, *ycf2*, *rpl23*, *rpl2*, *rpl12*, and *ycf1*), five tRNA genes (*trnV-GAC*, *trnL-GAU*, *trnA-UGC*, *trnR-ACG*, and *trnN-GUU*), and all four rRNA genes (*rrn4.5*, *rrn5*, *rrn23*, and *rrn16*) are anchored in the IR regions. Twelve PE genes and one tRNA gene (*trnL-UAG*) characterize the locations in the SSC sections (Figure 2). Most of the gene sequences assembled as a single-copy, whereas 18 genes occur with duplication in IR regions, including seven PE genes, four rRNA genes, and six tRNA genes (Table 2). Interestingly, *ycf1* is the only variable gene identified with an incomplete duplication in the junction of SSC and IRB regions (Figure 2 and Table 2). In *Arabidopsis*, translocon complex of *ycf1/Tic214* and *Tic20* has been proposed as the central component for plastid proteins accumulation [46]; however, a recent report stated that *ycf1/Tic214* might not be involved in the general import machinery [47]. All four *Ilex* plastomes uniformly contain *ycf2*, a duplicated gene, showing a maximum length of 6894 bp in IR regions. Another single-copy gene *rpoC2* has a relatively shorter length (4155 bp) in the LSC section. Besides, a total of 19 genes, including 11 PE genes and 7 tRNA genes, were identified with two exons, while three of PE genes (*clpP*, *ycf3*, and *rps12*) contain three exons. Contrast analyses revealed a slight variation in GC content in various regions of four *Ilex* plastomes (Table S1). Additionally, analyses of the total numbers of universal genetic code for the coding genes showed a similar range from 26,729 (*I. viridis*) to 27,121 (*I. latifolia*) (Table S2). Based on the statistical analysis of the codon usage, leucine (Leu) represents the most abundant amino acid with a frequency of 10.5%, whereas cysteine (Cys) shows the lowest abundance of 1.1%. Overall, the codon usage in all identified genes exhibits similar patterns across the four *Ilex* species. The almost identical gene numbers, annotation, and plastid genome length prompted us to further explore the sequence variation and divergence through the analyses of SSRs and long repeat sequences.
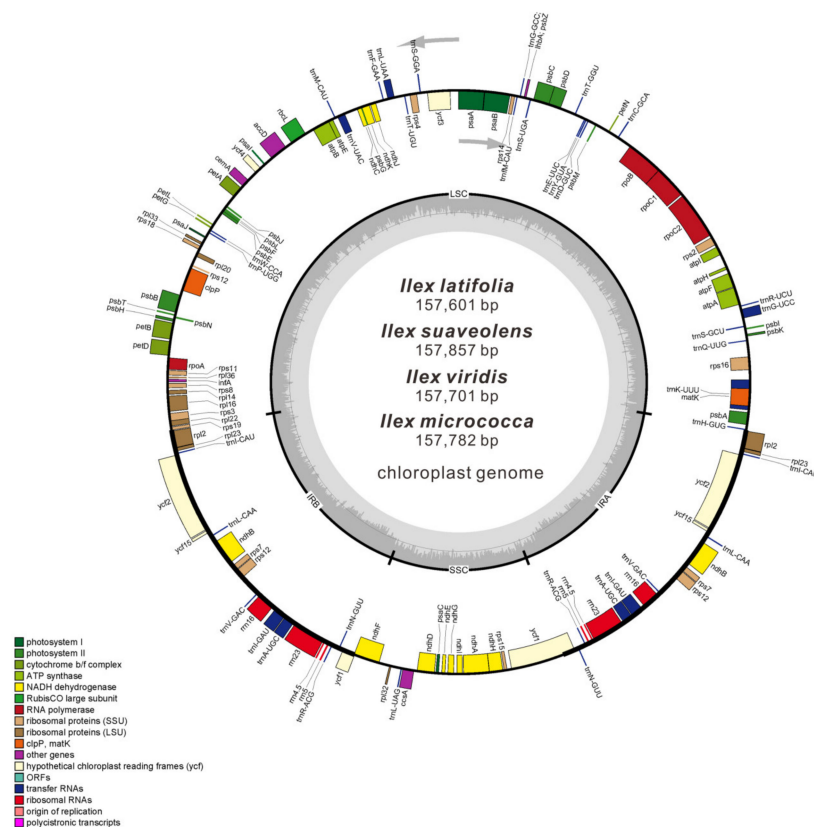

Figure 2. The plastome map of four *Ilex* species. The genes are shown by the gray arrowheads. Genes inside the circle indicate clockwise transcription, and those outsides are transcribed counterclockwise. The protein-encoding genes are marked in different colors. The GC content graphs are included as dark gray bars toward the center of the diagram.

Table 2. List of annotated genes in the plastid genome of *Ilex*.

Category	Gene Family	Gene Name	Numbers
Photosynthesis	Photosystem I	*psaA, psaB, psaC, psaI, psaJ*	5
	Photosystem II	*psbA, psbB, psbC, psbD, psbE, psbF, psbG, psbH, psbI, psbJ, psbK, psbL, psbM, psbN, psbT, psbZ/lhbA*	16
	Cytochrome b/f	*petA, petB[2], petD[2], petL, petG, petN*	6
	ATP synthase	*atpA, atpB, atpE, atpF[2], atpH, atpI*	6
	Cytochrome c	*ccsA*	1
	photosystem I assembly	*ycf3[3], ycf4*	2
	NADH dehydrogenase	*ndhA[2], ndhB[2d], ndhC, ndhD, ndhE, ndhF, ndhG, ndhH, ndhI, ndhJ, ndhK*	12
	Rubisco subunit L	*rbcL*	1
Transcription and translation	RNA polymerase β-subunit	*rpoC1[2], rpoC[2], rpoA, rpoB*	4
	30S ribosomal protein S	*rps2, rps3, rps4, rps7[d], rps8, rps11, rps12[d+], rps14, rps15, rps16[2], rps18, rps19*	14
	50S ribosomal protein L	*rpl2[2d], rpl14, rpl16[2], rpl20, rpl22, rpl23[d], rpl32, rpl33, rpl36*	11
RNA	Ribosomal RNA	*rrn4.5[d], rrn5[d], rrn16[d], rrn23[d]*	8
	Transfer RNA	*trnP-U/GGG, trnH-GUG, trnK-UUU2, trnQ-UUG, trnS-GCU, trnG-GCC2, trnG-UCC, trnR-UCU, trnC-GCA, trnD-GUC, trnY-GUA, trnE-UUC, trnT-GGU, trnM-CAU, trnS-UGA, trnfM-CAU, trnS-GGA, trnT-UGU, trnL-UAA2, trnF-GAA, trnV-UAC2, trnW-CCA, trnP-UGG, trnL-CAA[d], trnV-GAC[d], trnI-GAU2[d], trnR-ACG[d], trnL-UAG, trnN-GUU[d], trnA-UGC2[d], trnI-CAU[d]*	37
Other	RNA splicing	*matK*	1
	Plastid envelope	*cemA*	1
	Acetyl-CoA carboxylase β	*accD*	1
	Serine Protease	*clpP3*	1
	Translational initiation factor 1	*infA*	1
Unknown function	Conserved hypothetical gene	*ycf1[d], ycf2[d], ycf15[d]*	6
		orf56[d], orf42[d], ycf68[d], orf188	7 *

[2] Genes containing two exons; [3] genes containing three exons; [d] two gene copies in the IRs; [d+] two gene copies with one shared exon 1 in the LSC, whereas exon 2 and 3 are in the IRs; * pseudogene.

3.3. Analyses of SSRs and Long Repeat Sequences

Using the MISA program, a total of 221 SSRs (mono- /di- /trinucleotide repeats) were characterized in the plastid genomes of *I. latifolia* (46/2/1), *I. suaveolens* (46/3/1), *I. viridis* (53/3/1), and *I. micrococca* (52/2/1). Among the SSRs, the mononucleotide repeats are the most abundant type with a location in the non-coding regions, which is related to AT richness. The proportions of A/T sequences in the mononucleotide repeats appear to be identical in the four *Ilex* species, varying from 92.00% (*I. suaveolens*) to 92.98% (*I. viridis*), whereas C/G only exists in the plastid genomes of *I. latifolia* and *I. micrococca*. The AT/AT sequence of dinucleotide repeats showed similar levels (1.75%–4.08%); however, the AAT/ATT sequence of trinucleotide repeats showed significant variability, ranging in proportion from 1.82% to 6.00% (Figure S1). Other types of SSRs (e.g., tetra- and pentanucleotide repeats) were not identified in any *Ilex* plastid genomes. The long repeats (forward, reverse, complement, and palindrome repeats) were concurrently analyzed by REPuter. A total of 196 unique long repeats

were detected in *I. latifolia* (24/0/0/22), *I. suaveolens* (38/9/3/0), *I. viridis* (41/8/1/0), and *I. micrococca* (38/9/3/0) (Figure S2). The forward repeats appear to be the most common type in the four plastid genomes. We found no reverse and complement repeats identified in *I. latifolia*, but it uniquely contains the palindrome repeats (Figure S2a). Further analyses of various types of long repeats showed that the 20–30 bp sequence length is the typical pattern (Figures S2b–e). In four plastid genomes, the sequence length in reverse and complement repeats usually has a limitation of 30 bp, whereas this length can be extended to 60 bp in the forward and palindrome repeats.

3.4. Comparative Analyses of Complete Plastomes in Ilex Species

We used *I. suaveolens* as a reference to conduct a BLAST comparison of eight *Ilex* species, showing that the entire plastid genomes are well conserved across eight selected species. The LSC and SSC regions were found to be more substantially divergent than the IR regions (Table 3, Figure 3). The non-coding regions appear to have more significant variation compared with the coding regions, in which some genes are relatively conserved. The VISTA analysis resulted in the findings of 12 hotspot regions for genome divergence and variable genes (e.g., *rpoC1*, *rbcL*, *ndhF*, *clpP*, and *psbA*). The highly divergent hotspot regions are particularly located in the intergenic regions than in coding regions, including *trnH-psbA*, *matK-rps16*, *psbK-psbI-trnS-trnG*, *petN-psbM*, *trnE-psbD-trnT*, *trnS-psbZ-psaB*, *trnL-ycf3*, *rbcL-accD-ycf4*, *clpP-rpl33*, *rpl16-rpoA*, *rpl32-ccsA*, and *ycf15-rps12-rrn16* (Figure 4). These divergent sites may facilitate the development of potential DNA markers for the species identification and reconstruction of phylogeny in the genus *Ilex*.

Table 3. List of plastid genome features in *Ilex* species.

Species	Plastome (bp)	LSC (bp)	SSC (bp)	IRs (bp)	PE Genes	tRNA Genes	rRNA Genes	GC (%)
I. szechwanensis	157,900	87,204	18,513	52,183	96	40	8	37.6
I. pubescens	157,741	87,109	18,436	52,196	96	40	8	37.7
I. paraguariensis	157,614	87,144	18,307	52,154	86	37	8	37.6
I. wilsonii	157,918	87,266	18,432	52,220	96	40	8	37.6
I. latifolia	157,610	87,020	18,427	52,154	95	40	8	37.7
I. suaveolens	157,857	87,255	18,398	52,204	89	37	8	37.6
I. viridis	157,701	87,177	18,394	52,130	89	37	8	37.7
I. micrococca	157,782	87,200	18,434	52,148	89	37	8	37.6
I. integra	157,548	86,935	18,426	52,186	86	37	8	37.6

Further sequence comparison of quadripartite borders revealed that the IR regions are extremely conserved in eight *Ilex* species (Figure 5). The various sites of the *rps19*, *rpl2*, *ycf1*, and *ndhF* are generally located in the junction of IRs/SSC (JSA/B) and IRs/LSC (JLA/B). The *rps19* locates within the LSC region, showing 8–13 bp gaps to the JLB in *I. latifolia*, *I. pubescens*, *I. wilsonii*, *I. szechwanensis*, and *I. micrococca*. *rps19* was identified across the JLB with the addition of 4 bp in *I. paraguariensis*, *I. viridis*, and *I. suaveolens*. At the JLA, gene *rpl2* shows the same location of 55 bp away from the LSC in all *Ilex* species. A tRNA gene (*trnH*) showed 11 bp shifts in the LSC region (Figure 5). The JLB and JLA are moderately conserved, whereas the JSA and JSB are strikingly different in all *Ilex* species. One short copy gene, *ycf1* in IRB, shows a significant sequence variation, ranging in length from 1038 bp (*I. wilsonii*) to 1085 bp (*I. suaveolens*). The other long copy gene of *ycf1* in IRB shows an identical length of 5690 bp in most *Ilex* species, excluding *I. paraguariensis* (5693 bp), *I. wilsonii* (5684 bp), and *I. micrococca* (5684 bp). Within the SSC, the gene *ndhF*, which is correlated with photosynthesis, differs in length from 2231 bp, 2258 bp (15 bp shift), and 2270 bp (40 bp shift) from the JSB. The overlapping sequences of the short copy gene *ycf1* and *ndhF* are also commonly found in *I. latifolia*, *I. paraguariensis*, *I. micrococca*, *I. viridis*, and *I. suaveolens*. Other variable sequences were also identified between the gene *psbA* and the JLA (Figure 5). Overall, the compared sequence information of the IRs/SC boundaries suggested that the contractions and expansions in IRs regions have relatively stable patterns in the eight *Ilex* plastid genomes.
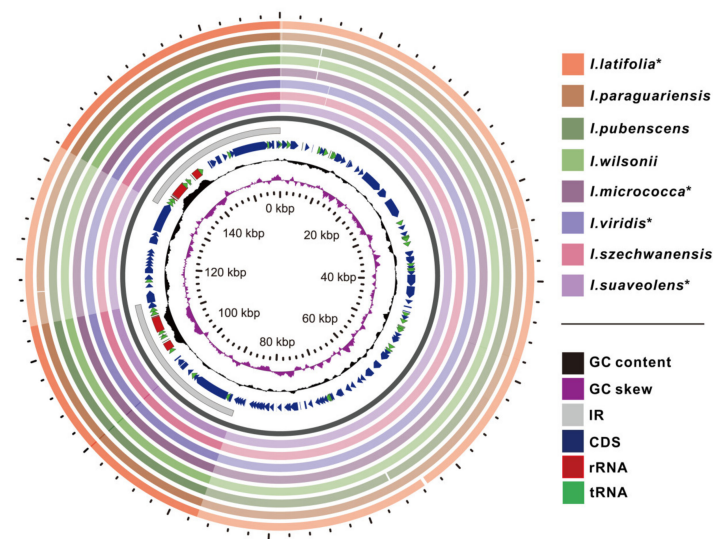
Figure 3. Plastome comparison between eight *Ilex* species in GView. The eight outer circles represent the BLAST results for *I. suaveolens* vs. *I. latifolia*, *I. paraguariensis*, *I. pubescens*, *I. wilsonii*, *I. micrococca*, *I. viridis*, *I. szechwanensis*, and itself, respectively. The clockwise inner cycle shows the CDS, ribosomal RNA (rRNA) genes, and transfer RNA (tRNA) genes in the plastome of *I. suaveolens*. The GC skew in a purple color indicates either G > C or G < C, and the GC content is shown in black.
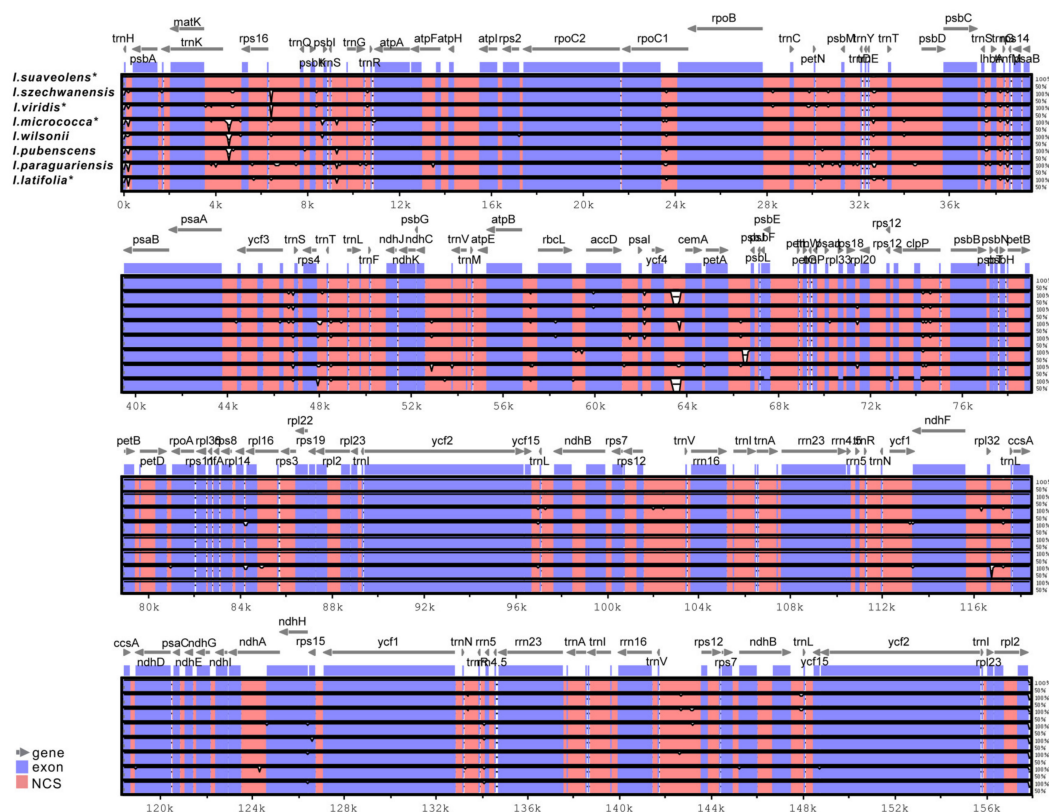
Figure 4. VISTA visualization of the alignment between the eight *Ilex* plastomes. The plots show the sequence identity with the *I.suaveolens* plastome, used as a reference. The vertical and horizontal axes represent the sequence consistency degree (50%–100%) and the sequence length, respectively. The locations of divergent hotspot regions are labeled along the top of the alignment, where the gray arrows indicate the orientation of the annotated genes. The red bars indicate the non-coding sequences (NCS), and white peaks represent differences in chloroplast genomes.
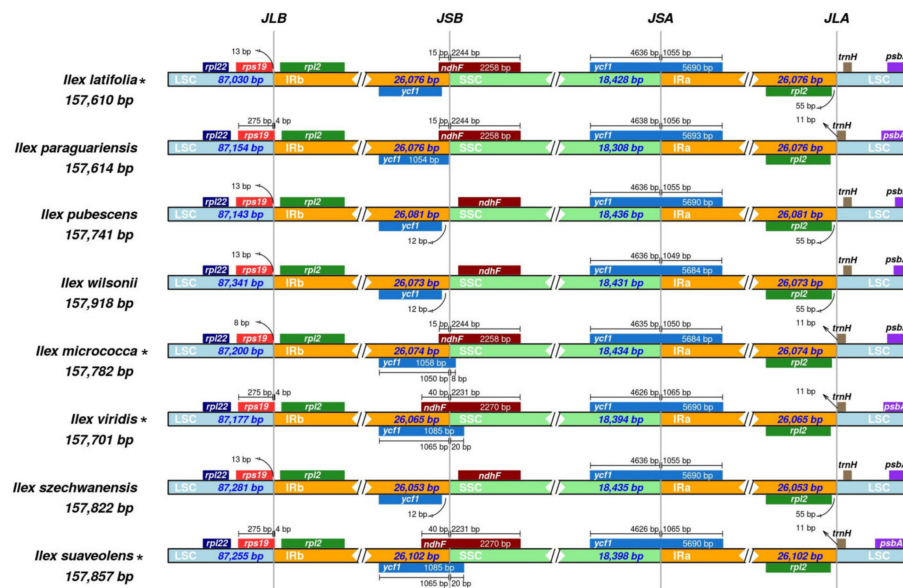
Figure 5. Comparative analyses of the junctions for the SSC/IRs and LSC/IRs regions among eight *Ilex* plastomes. The colored boxes above the strip scaled with sequence length indicate the denoted genes. The gaps in base length (bp) are indicated between boxed genes and boundaries. JLA, the junction of IRa/LSC; JLB, the junction of IRb/LSC; JSA, the junction of IRa/SSC; JSB, the junction of IRb/SSC.

3.5. Phylogenetic Analyses of Ilex in Aquifoliaceae

To investigate the phylogenies, the entire plastid genome sequences of 15 *Ilex* accessions were aligned using MAFFT. The phylogenetic topology was generated by MEGA X using the ML method supported with 1000 bootstrap values. *P. trichocarpa*, *P. deltoides*, *Q. acutissima*, and *H. himalaica* were used as out groups in the phylogenetic analyses. In Figure 6, the phylogenetic tree shows that four clades were deduced from all examined *Ilex* species. *I. latifolia* and *I. integra* cluster together with an identical high value in clade I, which also includes three additional *Ilex* species (*I. delavayi*, *I.sp. XY-2016*, and *I. cornuta*). All of them are evergreen trees with leathery leaves and broad distribution in subtropical Asia. The endemic *I. viridis* and *I. suaveolens* show close relationships in the clade III. *I. micrococca* clusters around *I. wilsonii* and *I. asprella* in clade IV, and the latter two are deciduous shrubs or trees. Yerba mate *I. paraguariensis* and *I. dumosa*, originating from South America, are located in clade II. The plastid phylogeny revealed that the species might not have originated from a single ancestor in Aquifoliaceae. The full encoding sequences were used to construct the phylogenetic tree, showing a distinctive clade (Figure S3).
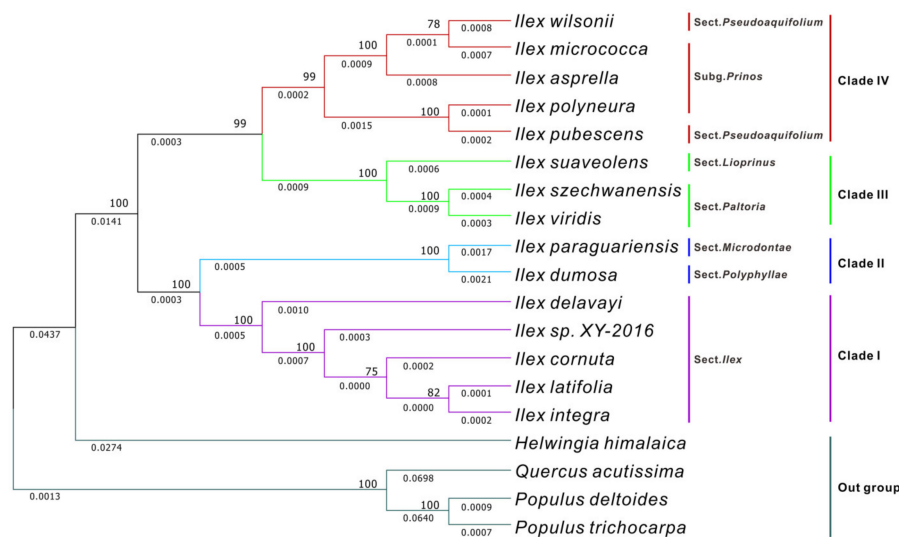
Figure 6. The phylogenetic analyses of 15 *Ilex* species using the entire plastid genomes. The evolutionary tree was constructed using MEGA X with the maximum likelihood (ML) method and the Tamura-Neighbour model. The percentage of trees in which the associated taxa clustered together is shown next to the branches. This analyses involved 19 plastid entire plastomes of *I. latifolia*, *I. suaveolens*, *I. viridis*, *I. micrococca*, *I. paraguariensis* (KP016928), *I. dumosa* (KP016927), *I. integra* (MK335537), *I. cornuta* (MK335536), *I. sp. XY-2016* (KX426469), *I. delavayi* (KX426470), *I. szechwanensis* (KX426466), *I. wilsonii* (KX426471), *I. polyneura* (KX426468), *I. asprella* (NC_045274.1), and *I. pubescens* (KX426467). The plastomes of *P. trichocarpa* (NC_009143), *P. deltoides* (MK267316), *Q. acutissima* (MF593895), and *H. himalaica* (KX434807) were used as outgroups.

4. Discussion

The holly *Ilex* mostly grows in mesic environments with global distribution. A recent report on phylogeny and biogeography suggested the origin of *Ilex* being in subtropical Asia [6]. Subsequently, it colonized other areas (e.g., South and North America, Australia, Europe, Africa, and some ocean islands) with divergence time from 4 to 30 million years ago [6]. Approximately 204 *Ilex* species (149 endemic species) have been described in China [48]. Located in a transition zone of north–south flora of eastern China, Mount Huangshan is considered as a priority spot for biodiversity conservation, in which all of 20 recorded *Ilex* species exhibited many medicinal properties and economic importance for garden and industry use [8,49,50]. Among these *Ilex* species, *I. suaveolens* is the most abundantly native species. The deciduous *I. micrococca* is locally grown as popular ornamental trees and for providing superior materials for paper and tannin production, in addition to herbal medicine use. The "Kuding" tea represented by *I. latifolia*, and pharmaceutical *I. viridis*, show potential clinical functions for scavenging heat, anti-inflammation, and detoxification, are moderately distributed in the dynamic forest plot [26].

The plasticity and diversity in leaf traits, including venation, anatomy, stomatal distribution, and stomatal conductance, were drastically affected by both cellular factors and environmental cues [51–53]. Hence, this extensively phenotypic survey of the vegetative and reproductive organs in plants contributes to the understanding of plant physiological and ecological adaptation. In our work, initial comparative analyses of leaf phenotypes revealed the significantly different values in LA, SLA, STD, and SW between four *Ilex* species, suggesting that these morphological traits could be potentially used as candidates to distinguish different *Ilex* species (Table 1). The persistent fruits and distinctive leave morphology in *Ilex* result from complicated genetic variation and ecological adaptability to various environments; therefore, the underlying molecular mechanisms require an extensive understanding of genome architecture and genome size diversity [54]. The transcriptome assembly was previously only conducted in *I. paraguariensis* [13]. Unfortunately, the draft of the full

nuclear genome sequencing and annotation remains unavailable in *Ilex*. Using the tomato as the internal standard, analyses of DNA image cytometry in four *Ilex* species revealed that the calculated NG showed approximately 955 Mb in *I. latifolia*. This NG value is higher than *I. cornuta* (642 Mb) [55]. The examined NG in *I. suaveolens* and *I. viridis* showed similar levels to *I. mucronata* (1073 Mb) [56]. The most abundant NG was characterized in *I. micrococca*, slightly lower than *I. paraguariensis* (1078 Mb) [29]. The estimated DNA C-value (NG) in four *Ilex* species had not been previously reported. We found that *I. micrococca* had the largest size in NG but showed the lowest SW levels in comparison with other *Ilex*. A similar phenomenon was reported in a recent study, which proposed a negative association between the seed mass and nuclear genome size in the diploid *Aesculus* species [57]. In contrast with all detected morphological traits, the calculated NG appears to have a linear relationship with SLA and STD in four *Ilex* species, which is inconsistent with the associated NG/STD patterns from a large-scale comparative analysis in angiosperms [19]. We hypothesized that the statistical associations between the NG, LEC, and STD might not be evaluated adequately with small sample capacity, which could reason the discordance above. Perhaps, the relationships of NG size and various leaf traits appeared to be not conservative among angiosperms due to the diversified environmental adaptation [58–60].

The advances in novel biotechnologies, particularly in whole-genome sequencing, created the opportunity to explore the phylogenies, species identification, and molecular markers among taxa [23,61]. In our work, using high-throughput sequencing technologies combined with de novo and reference-guided assembly, the complete plastomes of four *Ilex* species were constructed, showing the conserved quadripartite circular structure, comprising of LSC, SSC, and two copies of IR regions. A total of 144 putative genes (96 PE, 40 tRNA, and 8 rRNA genes) were previously annotated in *Ilex* plastid genomes [1]. The number of identified PE genes and tRNA appeared to be variable in various *Ilex* species [62,63]. Further sequence analyses indicated that *ycf68*, *orf42*, and *ycf68* were located between two exons of *trnA-UGC*, and *orf188* sequence overlap with one exon of *ndhA*. *trnP-GGG* shows a sequence overlapping with *trnP-UGG*. Thus, the total numbers of genes identified were normalized into 134 (89 PE, 37 tRNA, and 8 rRNA genes) in the plastomes of *I. suaveolens*, *I. viridis*, and *I. micrococca*. The most distinctive length in plastome is 370 bp between various *Ilex* species, whereas the maximum sequence difference of LSC is 331 bp, suggesting that the divergence in the LSC region may cause the varied plastome lengths, which might depend on the IR contraction and expansion [64]. Plastid genome sequences exhibit extensive variations in the length, number, and distribution of SSRs. These variations have potential importance for the outcome of genomic diversity [65]. Plastid SSRs have been used extensively in the taxonomy, phylogenies, and the maternal structures in the community, diversity, and differentiation [66]. In total, 221 SSRs and 196 unique long repeats were characterized in four *Ilex* plastomes. The mononucleotide repeats (A/T) in all SSRs represent the dominant type distributed in the non-coding regions, which is related to the AT abundance of the nucleotide composition. This typical pattern was also identified in previous reports [67–69]. The forward repeat and 20–30bp in sequence length were the most general features within the different types of long repeats.

In our study, both GView and mVISTA analyses of eight plastid genomes illustrated numerous divergent hotspots that are primarily located in the SSC regions, suggesting more variable sites in intergenic non-coding sequences than in coding genes. These variable sequences could potentially be used to develop new molecular markers for the identification and taxonomy in *Ilex*. The findings of these hotspot regions are compatible with a previous study that reported the presence of at least 11 divergent regions [1]. Although several divergent genes (*rbcL*, *atpB*, and *matK*) helped inform the reconstruction of the phylogenetic tree among distantly-related species, they might provide suitable resolution for studies within *Ilex* due to the relative lower divergence than the hotspot regions [2,6,22,70]. Recently, *trnH-psbA*, *rbcL-accD*, and *trnS-trnG* have been developed as genetic markers in other species [71,72]. The plastome-divergent hotspot regions were considered powerful for species-level identification. However, whether these divergence hotspot regions could be extensively used for classification and taxonomy in *Ilex* remains to be assessed experimentally.

The comparative detection of sequence variability in the junctions of IRs/SSC and IRs/LSC showed relative fluid patterns in eight *Ilex* species. Mostly, gene *rps19* in LSC regions is nearest to the JLB; however, it was found to span the JLB into IRB with a 4 bp extension in *I. paraguariensis*, *I. viridis*, and *I. suaveolens*. The typical variability in JSB was also observed for *ycf1* in the IRB region and *ndhF* in the SSC region. A similar phenomenon of the extent in IR regions has been recognized in other *Ilex* species [1]. The sequence variations in four junctions of SC/IRs appeared to frequently occur during genome evolution, which resulted in alteration of the plastome size [73]. Hence, the contraction and expansion at the boundaries of IR regions could explain the variability in sequence length between different plastid genomes [74].

Although the Aquifoliaceae has good fossil records, several systematic studies revealed a high incongruity of phylogenies between the nuclear trees and plastid trees, suggesting that the evolutionary and phylogenetic patterns of Aquifoliaceae remain to be elucidated [1,6,22]. Some molecular genetic factors, including inter-lineage hybridization, lineage sorting, introgression, and gene duplication and loss, were reported to significantly influence the phylogenetic incongruence, which broadly occurred in Angiosperms [75]. In contrast to the nuclear trees, the plastid trees strongly reflect the biogeographic distribution of extant species [22]. In our work, the phylogenetic topology revealed the presence of four clades within *Ilex*, in agreement with the fossil record [6], and also fit well with recent reports on plastid phylogenetic analysis [1,32,63]. All five *Ilex* species in the clade I belong to section *Ilex*, showing leathery morphology in the leaf, which is consistent with the traditional classification [48]. *I. paraguariensis* and *I. dumosa* are both located in clade II, suggesting a similar geographic distribution in Southern America. However, *I. suaveolens* in section *Lioprinus* clusters together with *I. szechwanensis* and *I. viridis* in clade III, while the latter two species are categorized into section *Paltoria*. The findings that evergreen species surprisingly have a close relationship with deciduous species in clade IV, further reflects the incongruence between the plastid phylogeny and traditional taxonomy [1,3,67]. Nevertheless, three endemic *Ilex* species in clade III show a relative similarity of the leaf morphology and growth pattern (600–1600 m altitude). Both subtropical deciduous *I. micrococca* and *I. asprella* in clade IV have very similar leaf margins, growing at an altitude of 400–1000 m. Additionally, the topology constructed using the entire plastomes significantly improved the quality of phylogenies compared to the use of full coding sequences (Figure S3). Therefore, the plastid topology provides more reliable clues to infer the species' geographic patterns and origins. Overall, when more complete plastome sequences of *Ilex* species become accessible, plastid trees combined with the nuclear markers will contribute to the resolution of the deeper branches of the biogeography and phylogeny in Aquifoliaceae.

5. Conclusions

The unspecific pollination and weakness of reproductive isolation resulted in the frequently intraspecific genetic exchange, confusing the phylogenies and biogeography in the *Ilex* genus. Dioecious *Ilex* species have similar flowers and fruits, but the variable morphology in leaves is commonly affected by climate and season, creating an obstacle to the identification and classification of the *Ilex* specimens [22,75]. Hitherto, the anatomical features and plastid genomes of four endemic *Ilex* species in Mount Huangshan were undocumented. In this work, we performed comparative analyses of a variety of phenotypic traits (e.g., leaves, flowers, and seeds) and the molecular profiles, including nuclear genome size, plastid genome structures, variable patterns, and phylogenetic relationships, in four *Ilex* species. The reconstruction of the plastid phylogenies verified the significant usage of the complete plastomes in identifying the phylogeography and phylogenetic evolution of *Ilex*. However, the evolved phenotypic variation and variable molecular patterns might be tightly related to the adapted ecology and environments. The limited accessions in the plastid genome restrained the extensive exploration of the phylogenies in *Ilex*, which necessitates additional sequencing samples. In summary, the morphological and molecular data in the present study provided informative resources for the in-depth mining of the phylogenies, biogeography, and genetic diversity in the family Aquifoliaceae.

References

1. Yao, X.; Tan, Y.-H.; Liu, Y.-Y.; Song, Y.; Yang, J.-B.; Corlett, R.T. Chloroplast genome structure in Ilex (Aquifoliaceae). *Sci. Rep.* **2016**, *6*, 28559. [CrossRef] [PubMed]

2. Cuénoud, P.; Spichiger, R.; Andrews, S.; Manen, J.-F.; Martinez, M.A.D.P.; Loizeau, P.-A. Molecular Phylogeny and Biogeography of the Genus Ilex L. (Aquifoliaceae). *Ann. Bot.* **2000**, *85*, 111–122. [CrossRef]

3. Manen, J.-F.; Barriera, G.; Loizeau, P.-A.; Naciri, Y. The history of extant Ilex species (Aquifoliaceae): Evidence of hybridization within a Miocene radiation. *Mol. Phylogenetics Evol.* **2010**, *57*, 961–977. [CrossRef] [PubMed]

4. Gottlieb, A.M.; Giberti, G.C.; Poggio, L. Molecular analyses of the genus Ilex (Aquifoliaceae) in southern South America, evidence from AFLP and ITS sequence data. *Am. J. Bot.* **2005**, *92*, 352–369. [CrossRef] [PubMed]

5. Tsang, A.C.; Corlett, R.T. Reproductive biology of the Ilex species (Aquifoliaceae) in Hong Kong, China. *Can. J. Bot.* **2005**, *83*, 1645–1654. [CrossRef]

6. Yao, X.; Song, Y.; Yang, J.; Tan, Y.; Corlett, R.T. Phylogeny and biogeography of the hollies (*Ilex*, L., Aquifoliaceae). *J. Syst. Evol.* **2020**. [CrossRef]

7. Xie, Y.; Wu, F.; Fang, X. Middle Eocene East Asian monsoon prevalence over southern China: Evidence from palynological records. *Glob. Planet. Change* **2019**, *175*, 13–26. [CrossRef]

8. Hao, D.; Gu, X.; Xiao, P.; Liang, Z.; Xu, L.; Peng, Y. Research progress in the phytochemistry and biology of Ilex pharmaceutical resources. *Acta Pharm. Sin. B* **2013**, *3*, 8–19. [CrossRef]

9. Zhao, X.; Pang, L.; Li, J.; Song, J.-L.; Qiu, L.-H. Apoptosis Inducing Effects of Kuding Tea Polyphenols in Human Buccal Squamous Cell Carcinoma Cell Line BcaCD885. *Nutrients* **2014**, *6*, 3084–3100. [CrossRef]

10. Cao, D.; Xu, C.; Xue, Y.; Ruan, Q.; Yang, B.; Liu, Z.; Cui, H.; Zhang, L.; Zhao, Z.; Jin, J. The therapeutic effect of Ilex pubescens extract on blood stasis model rats according to serum metabolomics. *J. Ethnopharmacol.* **2018**, *227*, 18–28. [CrossRef]

11. Berté, K.A.S.; Beux, M.R.; Spada, P.K.W.D.S.; Salvador, M.; Hoffmann-Ribani, R. Chemical Composition and Antioxidant Activity of Yerba-Mate (*Ilex paraguariensis* A.St.-Hil., Aquifoliaceae) Extract as Obtained by Spray Drying. *J. Agric. Food Chem.* **2011**, *59*, 5523–5527. [CrossRef] [PubMed]

12. Bracesco, N.; Sanchez, A.G.; Contreras, V.; Menini, T.; Gugliucci, A. Recent advances on Ilex paraguariensis research: Minireview. *J. Ethnopharmacol.* **2011**, *136*, 84–378. [CrossRef] [PubMed]

13. Debat, H.J.; Grabiele, M.; Aguilera, P.M.; Bubillo, R.E.; Otegui, M.B.; Ducasse, D.A.; Zapata, P.D.; Marti, D.A. Exploring the Genes of Yerba Mate (*Ilex paraguariensis* A. St.-Hil.) by NGS and De Novo Transcriptome Assembly. *PLoS ONE* **2014**, *9*, e109835. [CrossRef] [PubMed]

14. Hao, D.C.; Gu, X.-J.; Xiao, P.G. Phytochemistry and biology of Ilex pharmaceutical resources. In *Medicinal Plants*; Elsevier: Amsterdam, The Netherlands, 2015; pp. 531–585.

15. Kim, H.K.; Saifullah; Khan, S.; Wilson, E.G.; Kricun, S.D.P.; Meissner, A.; Goraler, S.; Deelder, A.M.; Choi, Y.H.; Verpoorte, R. Metabolic classification of South American Ilex species by NMR-based metabolomics. *Phytochemistry* **2010**, *71*, 773–784. [CrossRef]

16. Evens, Z.; Stellpflug, S. Holiday Plants with Toxic Misconceptions. *West. J. Emerg. Med.* **2012**, *13*, 538–542. [CrossRef]

17. Rendell, S.; Ennos, R.A. Chloroplast DNA diversity of the dioecious European tree *Ilex aquifolium* L. (English holly). *Mol. Ecol.* **2003**, *12*, 2681–2688. [CrossRef]

18. Spooner, D.M.; Van Den Berg, R.G.; Rivera-Peña, A.; Velguth, P.; Del Rio, A.; Salas-López, A. Taxonomy of Mexican and Central American Members of Solanum series Conicibaccata (sect. Petota). *Syst. Bot.* **2001**, *26*, 743–756.

19. Beaulieu, J.M.; Leitch, I.J.; Patel, S.; Pendharkar, A.; Knight, C.A. Genome size is a strong predictor of cell size and stomatal density in angiosperms. *New Phytol.* **2008**, *179*, 975–986. [CrossRef]

20. Smith, D.R. Does Cell Size Impact Chloroplast Genome Size? *Front. Plant Sci.* **2017**, *8*. [CrossRef]

21. Sabater, B. Evolution and Function of the Chloroplast. Current Investigations and Perspectives. *Int. J. Mol. Sci.* **2018**, *19*, 3095. [CrossRef]

22. Manen, J.F.; Boulter, M.C.; Naciri-Graven, Y. The complex history of the genus Ilex, L. (Aquifoliaceae): Evidence from the comparison of plastid and nuclear DNA sequences and from fossil data. *Plant. Syst. Evol.* **2002**, *235*, 79–98. [CrossRef]

23. Nock, C.J.; Waters, D.L.E.; Edwards, M.A.; Bowen, S.G.; Rice, N.; Cordeiro, G.M.; Henry, R.J. Chloroplast genome sequences from total DNA for plant identification. *Plant. Biotechnol. J.* **2011**, *9*, 328–333. [CrossRef] [PubMed]

24. Parks, M.; Cronn, R.; Liston, A. Increasing phylogenetic resolution at low taxonomic levels using massively parallel sequencing of chloroplast genomes. *BMC Biol.* **2009**, *7*, 84. [CrossRef] [PubMed]

25. Xu, H.; Cao, M.; Wu, Y.; Cai, L.; Cao, Y.; Ding, H.; Cui, P.; Wu, J.; Wang, Z.; Le, Z.; et al. Optimized monitoring sites for detection of biodiversity trends in China. *Biodivers. Conserv.* **2017**, *26*, 1959–1971. [CrossRef]

26. Ding, H.; Fang, Y.; Yang, X.; Yuan, F.; He, L.; Yao, J.; Wu, J.; Chi, B.; Li, Y.; Chen, S.; et al. Community characteristics of a subtropical evergreen broad-leaved forest in Huangshan, Anhui Province, East China. *Biodivers. Sci.* **2016**, *24*, 875–887. [CrossRef]

27. Da Silva, N.R.; Oliveira, M.W.D.S.; Filho, H.A.D.A.; Pinheiro, L.F.S.; Rossatto, D.R.; Kolb, R.M.; Bruno, O.M. Leaf epidermis images for robust identification of plants. *Sci. Rep.* **2016**, *6*, 25994. [CrossRef]

28. Thompson, K.; Band, S.R.; Hodgson, J.G. Seed Size and Shape Predict Persistence in Soil. *Funct. Ecol.* **1993**, *7*, 236–241. [CrossRef]

29. Gottlieb, A.M.; Poggio, L. Quantitative and qualitative genomic characterization of cultivated Ilex, L. species. *Plant. Genet. Resour.* **2015**, *13*, 142–152. [CrossRef]

30. Doležel, J.; Greilhuber, J.; Suda, J. Estimation of nuclear DNA content in plants using flow cytometry. *Nat. Protoc.* **2007**, *2*, 2233–2244. [CrossRef]

31. Su, T.; Han, M.; Min, J.; Cao, D.; Pan, H.; Liu, Y. The complete chloroplast genome sequence of Populus deltoides 'Siyang-2'. *Mitochondrial DNA Part. B* **2020**, *5*, 283–285. [CrossRef]

32. Shi, Y.; Liu, B. Complete chloroplast genome sequence of Ilex latifolia (Aquifoliaceae), a traditional Chinese tea. *Mitochondrial DNA Part. B* **2020**, *5*, 190–191. [CrossRef]

33. Bankevich, A.; Nurk, S.; Antipov, D.; Gurevich, A.A.; Dvorkin, M.; Kulikov, A.S.; Lesin, V.M.; Nikolenko, S.I.; Pham, S.; Prjibelski, A.D.; et al. SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *J. Comput. Biol.* **2012**, *19*, 455–477. [CrossRef] [PubMed]

34. Liu, C.; Shi, L.; Zhu, Y.; Chen, H.; Zhang, J.; Lin, X.; Guan, X. CpGAVAS, an integrated web server for the annotation, visualization, analysis, and GenBank submission of completely sequenced chloroplast genome sequences. *BMC Genomics* **2012**, *13*, 715. [CrossRef] [PubMed]

35. Greiner, S.; Lehwark, P.; Bock, R. OrganellarGenomeDRAW (OGDRAW) version 1.3.1: Expanded toolkit for the graphical visualization of organellar genomes. *Nucleic Acids Res.* **2019**, *47*, W59–W64. [CrossRef] [PubMed]

36. Kumar, S.; Stecher, G.; Li, M.; Knyaz, C.; Tamura, K. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol. Biol. Evol.* **2018**, *35*, 1547–1549. [CrossRef]

37. Beier, S.; Thiel, T.; Münch, T.; Scholz, U.; Mascher, M. MISA-web: A web server for microsatellite prediction. *Bioinformatics* **2017**, *33*, 2583–2585. [CrossRef] [PubMed]

38. Kurtz, S. REPuter: The manifold applications of repeat analysis on a genomic scale. *Nucleic Acids Res.* **2001**, *29*, 4633–4642. [CrossRef]

39. Petkau, A.; Stuart-Edwards, M.; Stothard, P.; Van Domselaar, G. Interactive microbial genome visualization with GView. *Bioinformatics* **2010**, *26*, 3125–3126. [CrossRef]

40. Brudno, M.; Malde, S.; Poliakov, A.; Do, C.B.; Couronne, O.; Dubchak, I.; Batzoglou, S. Glocal alignment: Finding rearrangements during alignment. *Bioinformatics* **2003**, *19*, i54–i62. [CrossRef]

41. 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]

42. Amiryousefi, A.; Hyvönen, J.; Poczai, P. IRscope: An online program to visualize the junction sites of chloroplast genomes. *Bioinformatics* **2018**, *34*, 3030–3031. [CrossRef] [PubMed]

43. 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] [PubMed]

44. Obermayer, R. Nuclear DNA C-values in 30 Species Double the Familial Representation in Pteridophytes. *Ann. Bot.* **2002**, *90*, 209–217. [CrossRef] [PubMed]

45. Dolezel, J. Plant DNA Flow Cytometry and Estimation of Nuclear Genome Size. *Ann. Bot.* **2005**, *95*, 99–110. [CrossRef]

46. Kikuchi, S.; Bédard, J.; Hirano, M.; Hirabayashi, Y.; Oishi, M.; Imai, M.; Takase, M.; Ide, T.; Nakai, M. Uncovering the Protein Translocon at the Chloroplast Inner Envelope Membrane. *Sci.* **2013**, *339*, 571–574. [CrossRef]

47. Bölter, B.; Soll, J. Ycf1/Tic214 Is Not Essential for the Accumulation of Plastid Proteins. *Mol. Plant.* **2017**, *10*, 219–221. [CrossRef]

48. Chen, S.K.; Ma, H.; Feng, Y.; Barriera, G.; Loizeau, P.A. Aquifoliaceae. In *Flora of China*; Wu, Z.Y., Raven, P.H., Hong, D.Y., Eds.; Science Press: Beijing, China; Missouri Botanical Garden Press: St. Louis, MO, USA, 2008; Volume 11, pp. 359–438.

49. Qian, Y.; Tian, R. Research Advance of Ilex Germplasm Resources and Their Application to Landscape. *World For. Res.* **2016**, *29*, 40–45.

50. Yi, F.; Zhao, X.; Peng, Y.; Xiao, P. *Genus Ilex* L.: Phytochemistry, Ethnopharmacology, and Pharmacology. *Chinese Herb. Med.* **2016**, *8*, 209–230. [CrossRef]

51. Hetherington, A.M.; Woodward, F.I. The role of stomata in sensing and driving environmental change. *Nature* **2003**, *424*, 901–908. [CrossRef]

52. Maricle, B.R.; Koteyeva, N.K.; Voznesenskaya, E.V.; Thomasson, J.R.; Edwards, G.E. Diversity in leaf anatomy, and stomatal distribution and conductance, between salt marsh and freshwater species in the C4 genus Spartina (Poaceae). *New Phytol.* **2009**, *184*, 216–233. [CrossRef]

53. Scoffoni, C.; Kunkle, J.; Pasquet-Kok, J.; Vuong, C.; Patel, A.J.; Montgomery, R.A.; Givnish, T.J.; Sack, L. Light-induced plasticity in leaf hydraulics, venation, anatomy, and gas exchange in ecologically diverse Hawaiian lobeliads. *New Phytol.* **2015**, *207*, 43–58. [CrossRef]

54. Dodsworth, S.; Leitch, A.R.; Leitch, I.J. Genome size diversity in angiosperms and its influence on gene space. *Curr. Opin. Genet. Dev.* **2015**, *35*, 73–78. [CrossRef]

55. Zhang, L.; Cao, B.; Bai, C. New reports of nuclear DNA content for 66 traditional Chinese medicinal plant taxa in China. *Caryologia* **2013**, *66*, 375–383. [CrossRef]

56. Bai, C.; Alverson, W.S.; Follansbee, A.; Waller, D.M. New reports of nuclear DNA content for 407 vascular plant taxa from the United States. *Ann. Bot.* **2012**, *110*, 1623–1629. [CrossRef]

57. Krahulcová, A.; Trávníček, P.; Krahulec, F.; Rejmánek, M. Small genomes and large seeds: Chromosome numbers, genome size and seed mass in diploid Aesculus species (Sapindaceae). *Ann. Bot.* **2017**, *119*, 957–964.

58. Bertolino, L.T.; Caine, R.S.; Gray, J.E. Impact of Stomatal Density and Morphology on Water-Use Efficiency in a Changing World. *Front. Plant. Sci.* **2019**, *10*. [CrossRef]

59. Jordan, G.J.; Carpenter, R.J.; Koutoulis, A.; Price, A.; Brodribb, T.J. Environmental adaptation in stomatal size independent of the effects of genome size. *New Phytol.* **2015**, *205*, 608–617. [CrossRef]

60. Morgan, H.D.; Westory, M. The Relationship between Nuclear DNA Content and Leaf Strategy in Seed Plants. *Ann. Bot.* **2005**, *96*, 1321–1330. [CrossRef]

61. Zong, D.; Gan, P.; Zhou, A.; Zhang, Y.; Zou, X.; Duan, A.; Song, Y.; He, C. Plastome Sequences Help to Resolve Deep-Level Relationships of Populus in the Family Salicaceae. *Front. Plant. Sci.* **2019**, *10*, 5. [CrossRef]

62. Cascales, J.; Bracco, M.; Garberoglio, M.; Poggio, L.; Gottlieb, A. Integral Phylogenomic Approach over Ilex, L. Species from Southern South America. *Life* **2017**, *7*, 47. [CrossRef]

63. Park, J.; Kim, Y.; Nam, S.; Kwon, W.; Xi, H. The complete chloroplast genome of horned holly, Ilex cornuta Lindl. & Paxton (Aquifoliaceae). *Mitochondrial DNA Part. B* **2019**, *4*, 1275–1276.

64. Kim, K.; Lee, S.-C.; Lee, J.; Yu, Y.; Yang, K.; Choi, B.-S.; Koh, H.-J.; Waminal, N.E.; Choi, H.-I.; Kim, N.-H.; et al. Complete chloroplast and ribosomal sequences for 30 accessions elucidate evolution of Oryza AA genome species. *Sci. Rep.* **2015**, *5*, 15655. [CrossRef]

65. George, B.; Bhatt, B.S.; Awasthi, M.; George, B.; Singh, A.K. Comparative analysis of microsatellites in chloroplast genomes of lower and higher plants. *Curr. Genet.* **2015**, *61*, 665–677. [CrossRef]

66. Provan, J. Novel chloroplast microsatellites reveal cytoplasmic variation in Arabidopsis thaliana. *Mol. Ecol.* **2000**, *9*, 2183–2185. [CrossRef] [PubMed]

67. Li, D.-M.; Zhu, G.-F.; Xu, Y.-C.; Ye, Y.-J.; Liu, J.-M. Complete Chloroplast Genomes of Three Medicinal Alpinia Species: Genome Organization, Comparative Analyses and Phylogenetic Relationships in Family Zingiberaceae. *Plants* **2020**, *9*, 286. [CrossRef] [PubMed]

68. Nguyen, V.B.; Linh Giang, V.N.; Waminal, N.E.; Park, H.-S.; Kim, N.-H.; Jang, W.; Lee, J.; Yang, T.-J. Comprehensive comparative analysis of chloroplast genomes from seven Panax species and development of an authentication system based on species-unique single nucleotide polymorphism markers. *J. Ginseng Res.* **2020**, *44*, 135–144. [CrossRef] [PubMed]

69. Yang, Y.; Zhou, T.; Duan, D.; Yang, J.; Feng, L.; Zhao, G. Comparative Analysis of the Complete Chloroplast Genomes of Five Quercus Species. *Front. Plant. Sci.* **2016**, *07*, 959. [CrossRef]

70. Rohwer, J.G. Toward a Phylogenetic Classification of the Lauraceae: Evidence from matK Sequences. *Syst. Bot.* **2000**, *25*, 60–71. [CrossRef]

71. Shaw, J.; Shafer, H.L.; Leonard, O.R.; Kovach, M.J.; Schorr, M.; Morris, A.B. Chloroplast DNA sequence utility for the lowest phylogenetic and phylogeographic inferences in angiosperms: The tortoise and the hare IV. *Am. J. Bot.* **2014**, *101*, 1987–2004. [CrossRef]

72. Yang, J.; Vázquez, L.; Chen, X.; Li, H.; Zhang, H.; Liu, Z.; Zhao, G. Development of Chloroplast and Nuclear DNA Markers for Chinese Oaks (Quercus Subgenus Quercus) and Assessment of Their Utility as DNA Barcodes. *Front. Plant. Sci.* **2017**, *8*, 816. [CrossRef]

73. Liu, X.; Chang, E.-M.; Liu, J.-F.; Huang, Y.-N.; Wang, Y.; Yao, N.; Jiang, Z.-P. Complete Chloroplast Genome Sequence and Phylogenetic Analysis of Quercus bawanglingensis Huang, Li et Xing, a Vulnerable Oak Tree in China. *Forests* **2019**, *10*, 587. [CrossRef]

74. Wang, R.-J.; Cheng, C.-L.; Chang, C.-C.; Wu, C.-L.; Su, T.-M.; Chaw, S.-M. Dynamics and evolution of the inverted repeat-large single copy junctions in the chloroplast genomes of monocots. *BMC Evol. Biol.* **2008**, *8*, 36. [CrossRef]

75. Shi, L.; Li, N.; Wang, S.; Zhou, Y.; Huang, W.; Yang, Y.; Ma, Y.; Zhou, R. Molecular Evidence for the Hybrid Origin of Ilex dabieshanensis (Aquifoliaceae). *PLoS ONE* **2016**, *11*, e0147825. [CrossRef] [PubMed]