BMC Ecology and Evolution

RESEARCH

Open Access

Comparative analysis of chloroplast genomes in ten holly (*Ilex*) species: insights into phylogenetics and genome evolution

Jiaxin Hu[1†], Daoliang Yan[1*†], Huwei Yuan[1], Jianhong Zhang[2] and Bingsong Zheng[1]

Abstract

In order to clarify the chloroplast genomes and structural features of ten *Ilex* species and provide insights into the phylogeny and genome evolution of the genus *Ilex*, we conducted a comparative analysis of chloroplast genomes using bioinformatics methods. The chloroplast genomes of ten *Ilex* species were obtained, and their structural features and variations were compared. The results indicated that all chloroplast genomes in the genus *Ilex* exhibit a double-stranded circular structure, with sizes ranging from 157,356 to 158,018 bp, showing minimal differences in size. The chloroplast genomes of the ten *Ilex* species have a relatively conservative gene count, with a total of 134 to 135 genes, including 88 or 89 protein-coding genes, and a conserved number of 8 rRNA genes. Each chloroplast genome contains 3 to 123 SSR (Simple Sequence Repeat) sites, predominantly composed of mononucleotide and trinucleotide repeats, with no detection of pentanucleotide or hexanucleotide repeats. The variation in dispersed repeat sequences among *Ilex* species is minimal, with a total repeat sequence number ranging from 1 to 14, concentrated in the length range of 30 to 42 base pairs. The expansion and contraction of chloroplast genome boundaries among *Ilex* species are relatively stable, with only minor variations observed in individual species. Variations in non-coding regions are more pronounced than those in coding regions, with the variability in the Large Single Copy region (LSC) being the highest, while the variability in the Inverted Repeat region A (IRa) is the lowest. The divergence time among *Ilex* species was estimated using the MCMC-tree module, revealing the evolutionary relationships among these species, their common ancestors, and their differentiation throughout the evolutionary process. The research findings provide a valuable reference for the systematic study and molecular marker development of *Ilex* plants.

Keywords Genus *Ilex*, Chloroplast genome, SSR sites, IR boundary analysis, Divergence time

[†]Hu jiaxin and Yan Daoliang co-authored as the first authors.

*Correspondence:
Daoliang Yan
liangsie@zafu.edu.cn
[1]Zhejiang A&F University State Key Laboratory of Subtropical Silviculture, Zhejiang A&F University, Hangzhou, China
[2]Ningbo Key Laboratory of Characteristic Horticultural Crops in Quality Adjustment and Resistance Breeding, Ningbo Academy of Agricultural Sciences, Ningbo, China

Introduction

Chloroplasts, as cellular organelles in plant cells, possess their own genome that encodes proteins related to photosynthesis, transport proteins, and ribosomal RNA [1]. Despite containing a substantial amount of genetic information, medium-sized chloroplast genomes can be easily sequenced, particularly with the rapid development of high-throughput sequencing technologies [2]. The chloroplast genome exhibits a highly conserved structure, characterized by a short overall sequence length that facilitates sequencing, a stable gene structure, a moderate evolutionary rate, and well-maintained collinearity among cp. genomes across different plant taxa. These features not only facilitate systematic studies at various levels but also confer advantages in comparative analyses [3], and there is good collinearity between chloroplast genomes of different plant groups. This makes chloroplast genomes not only useful for studies at different taxonomic levels but also advantageous in comparative analyses. Plant chloroplast genomes are widely used to analyze the phylogenetic relationships between species, study the origin and domestication, and investigate the structural features, variations, and evolution of plant genes [4]. Therefore, research on plant chloroplast genomes is a crucial aspect of plant genetics, significantly influences the development of phylogenomics.

Plants of *Ilex* are distributed globally, comprising approximately 400 species [5], some of which have significant medicinal value. Several studies have identified compounds in *Ilex*, such as polysaccharides and *Ilex* pubescens triterpenoid saponins (IPTS) [6], exhibiting antioxidant [7], anti-inflammatory [8, 9], anti-tumor, and antimicrobial activities. As the largest genus of dioecious woody plants, *Ilex* is a good system for studying speciation, extinction and biogeography [10]. The classification and phylogenetic relationships of *Ilex* species have been a focus of taxonomic research, crucial for understanding evolutionary history, biodiversity conservation, and potential applications in medicine and horticulture. While past studies relied on morphological characteristics for classification, the development of molecular biology techniques has led to an increasing number of studies using molecular markers for phylogenetic analysis. By comparing the chloroplast genome sequences of ten *Ilex* species, we aim to understand their genome evolutionary history and relationships, revealing the phylogenetic connections among *Ilex* species. This knowledge contributes to a better understanding of species classification, evolutionary history, and aids in the establishment of an accurate classification system.

Table 1 Presents the project numbers and some sequencing details

Name of the species	Accession No.	Read Sum	Q30(%)	Maximal depth	Minimal depth
Ilex bioritsensis	PP541586	14,700,623	92.84	3809x	395x
Ilex buergeri	PP541587	12,660,092	93.15	2598x	220x
Ilex cassine 'Tensaw'	PP541588	19,665,614	92.77	2161x	334x
Ilex centrochinensis	PP541589	16,846,301	93.14	3065x	222x
Ilex chinensis	PP541590	11,396,305	93.29	1571x	61x
Ilex litseifolia	PP541591	20,119,677	93.06	1829x	256x
Ilex macrocarpa	PP541592	11,610,061	92.96	2022x	171x
Ilex pedunculosa	PP541593	21,033,268	92.9	1039x	175x
Ilex rotunda	PP541594	13,101,857	92.86	1063x	36x
Ilex verticillata	PP541595	14,652,492	92.83	998x	42x

Note ReadSum refers to the total number of paired-end reads in the clean data. Q30 indicates the percentage of bases in the clean data with quality scores of 30 or higher

Materials and methods

Test materials

The 10 species of plants belonging to the genus *Ilex* are: *Ilex bioritsensis*、*I.buergeri*、*I.cassine Tensaw*、*I. centrochinensis*、*I.chinensis*、*I.litseifolia*、*I. macrocarpa*、*I.pedunculosa*、*I.otunda* and *I. verticillata*. Tissue samples from these species were collected for chloroplast genome sequencing at Zhejiang A&F University. Genomic DNA extraction was performed using the Tiangen kit (Tiangen, Beijing) for subsequent analysis.

Genomic DNA extraction and sequencing

Fresh and healthy young leaves were collected, and total genomic DNA was extracted using the CTAB method [11]. After confirming the quality of the genomic DNA via agarose gel electrophoresis, a mechanical method (ultrasonication) was utilized to fragment the DNA. Subsequently, the fragmented DNA underwent purification, end repair, and sequencing adapter ligation, including 3′-end adenylation.

Gel electrophoresis was performed to size-select DNA fragments, followed by PCR amplification to generate a sequencing library. Each library underwent initial quantification using Qubit 2.0, library dilution, and evaluation on an Agilent 2100 Bioanalyzer. After confirming the insert size met expectations, quantitative PCR (Q-PCR) was employed for accurate determination of library concentration. Following this, the Illumina Nova seq platform was used for paired-end sequencing with a read length of 150 bp [12]. The sequencing data have been deposited in the NCBI database under BioProject ID BankIt2810171. Individual sample accession numbers and additional sequencing details are provided in Table 1.

Chloroplast genome annotation of ten species of *Ilex* Linn. Plant

The raw data were filtered using the fastp software to remove adapter sequences and low-quality reads, resulting in high-quality clean data. The filtered data were assembled into chloroplast genomes using SPAdes v3.10.1 software with k-mer parameters of 55, 87, and 121 [13]. The assembly was performed without relying on a reference genome. SPAdes was also used to assemble the cpDNA sequence into the chloroplast genome's SEED sequence. SSPACE v2.0 software was employed to connect contig sequences into scaffolds, and Gapfiller v2.1.1 software was used to fill gaps [14].The sequencing reads were aligned to a pseudo genome for genome correction. Based on the chloroplast's structure, the corrected pseudo genome underwent coordinate rearrangement to obtain the complete circular chloroplast genome sequence. The visualization of the chloroplast genome was achieved using the OGDRAW online tool.

Chloroplast Genome Annotation of ten species of *Ilex* Linn. Plant

Two methods were employed to annotate the chloroplast genome for increased accuracy. Firstly, Prodigal v2.6.3 was used to annotate the chloroplast's CDS, HMMER v3.1b2 software for rRNA prediction, and Aragorn v1.2.38 was applied for tRNA prediction. Secondly, gene sequences of closely related species available on NCBI were extracted and aligned to the assembled sequences using BLAST v2.6 (https://blast.ncbi.nlm.nih.gov/Blast.cgi)to obtain the second annotation result. The two annotation results were manually inspected for differing genes, and incorrect or redundant annotations were removed. This process determined the boundaries of multiple exons, resulting in the final annotation.

Analysis of cpSSRs and scattered repeat sequences

Simple Sequence Repeats (SSR) markers, composed of a few nucleotides (usually 1 to 6) repeated several times, were analyzed as cpSSR markers on the chloroplast genome using MISA v1.0 (MIcroSAtellite identification tool, http://pgrc.ipk-gatersleben.de/misa/misa.html). Parameters used were 1–8 (single-base repeat of 8 times or more), 2–5, 3–3, 4–3, 5–3, and 6–3.

Scattered repeat sequences, unlike tandem repeat sequences, are dispersed throughout the genome. vmatch v2.3.0 (http://www.vmatch.de/) software, along with Perl scripts, was used to identify repeat sequences, with parameters set to a minimum length=30 bp and a Hamming distance=3.

Genome-wide comparative analysis

The MAFFT (--auto) software was employed for genome-wide comparisons. The IRscope online tool was used to delineate the boundaries of the inverted repeat regions in the chloroplast genomes of the aforementioned ten species. Gene positions on the boundaries were analyzed for contractions and expansions, and the boundary information was visualized using the SVG module in Perl.

Relative synonymous codon usage

Due to the degeneracy of codons, each amino acid corresponds to at least one codon, and at most six codons. There is a significant variation in codon usage across different species and organisms due to synonymous codons. This unevenness in the usage of synonymous codons is referred to as Relative Synonymous Codon Usage (RSCU) [15]. This preference is considered to be a comprehensive result of natural selection, species mutation, and genetic drift. Perl scripts were employed to calculate the RSCU for the chloroplast genomes of 10 *Ilex* species. RSCU is calculated as the ratio of the observed frequency to the average frequency of synonymous codons. RSCU values can be categorized into three types: greater than 1, less than 1, and equal to 1. The value greater than 1 indicates a higher frequency of usage for that codon compared to others, while the value less than 1 suggests a higher frequency for synonymous codons other than the specified one. If the RSCU value is equal to 1, it implies no bias in the usage of that codon. Generally, the origin of codon preference is attributed to the uneven abundance of tRNA corresponding to different codons in the cell, with higher tRNA abundance leading to a higher usage frequency of the corresponding codon.

Phylogenetic analysis and divergence time estimation

To elucidate the intergeneric relationships and phylogenetic positions of 10 *Ilex* species, we conducted sequencing and utilized the complete chloroplast genomes of 13 *Ilex* species available on NCBI to construct a comprehensive evolutionary tree. *Anacardium occidentale* and *Maytenus guangxiensis* (Celastraceae) were chosen as outgroups, and the circular sequences were aligned using the MAFFT software in auto mode. The aligned data were trimmed using trimAl (v1.4.rev15), and the maximum likelihood evolutionary tree was constructed using RAxML (v8.2.10) with the GTRGAMMA model, rapid bootstrap analysis (bootstrap=1000). Additionally, the MCMCtree module in PAML (v4.9) was used to estimate the divergence times among species based on this tree.

Results

Chloroplast Genome features of 10 *Ilex* Species

The chloroplast genomes of the 10 *Ilex* species all exhibit a double-stranded circular structure, consisting of four parts: a large single-copy region (LSC), a small single-copy region (SSC), and a pair of inverted repeat regions (IRs). The size of the chloroplast genomes ranges

Table 2 Comparison of chloroplast genomes of 10 *Ilex* species

Name of the species	Total length (bp) and total GC content %	Length (bp)and GC content of LSC %	Length (bp) and GC content of SSC %	Length (bp) and GC content of IR	Total genes	Protein Genes	rRNA	tRNA	Pseudo gene
Ilex bioritsensis	157500,37.64	86826,35.66	18424,31.91	26125,42.95	134	89	8	37	0
Ilex buergeri	157569,37.64	86957,35.66	18426,31.92	26093,42.94	134	89	8	37	0
Ilex cassine 'Tensaw'	157828,37.63	87181,35.68	18459,31.88	26094,42.93	134	89	8	37	0
Ilex centrochinensis	157356,37.69	86730,35.75	18426,31.92	26100,42.93	135	89	8	38	0
Ilex chinensis	157887,37.62	87291,35.64	18388,31.91	26104,42.94	134	89	8	37	0
Ilex litseifolia	157970,37.63	87380,35.66	18382,31.9	26104,42.95	134	89	8	37	0
Ilex macrocarpa	157798,37.62	87232,35.63	18440,31.88	26063,42.97	134	89	8	37	0
Ilex pedunculosa	158018,37.63	87400,35.66	18412,31.89	26103,42.95	134	89	8	37	0
Ilex rotunda	157500,37.64	86826,35.66	18424,31.91	26125,42.95	134	89	8	37	0
Ilex verticillata	157810,37.61	87183,35.62	18435,31.91	26096,42.96	134	88	8	37	1

Table 3 Statistical table of chloroplast gene function classification

Category	Gene group	Gene name	Number
Photosynthesis	Subunits of photosystem I	psaA, psaB, psaC, psaI, psaJ	5
	Subunits of photosystem II	psbA, psbB, psbC, psbD, psbE, psbF, psbG, psbH, psbI, psbJ, psbK, psbL, psbM, psbN, psbT, psbZ	16
	Subunits of NADH dehydrogenase	ndhA*,ndhB*(2),ndhC, ndhD, ndhE, ndhF, ndhG, ndhH, ndhI, ndhJ, ndhK	11
	Subunits of cytochrome b/f complex	petA, petB*,petD*,petG, petL, petN	6
	Subunits of ATP synthase	atpA, atpB, atpE, atpF*,atpH, atpI	6
	Large subunit of rubisco	rbcL	1
	Subunits photochlorophyllide reductase	-	-
Self-replication	Proteins of large ribosomal subunit	rpl14,rpl16*,rpl2*(2),rpl20,rpl22,rpl23(2),rpl32,rpl33,rpl36	9
	Proteins of small ribosomal subunit	rps11,rps12**(2),rps14,rps15,rps16*,rps18,rps19,rps2,rps3,rps4,rps7(2),rps8	12
	Subunits of RNA polymerase	rpoA, rpoB, rpoC1*,rpoC2	4
	Ribosomal RNAs	rrn16(2),rrn23(2),rrn4.5(2),rrn5(2)	4
	Transfer RNAs	trnA-UGC*(2),trnC-GCA, trnD-GUC, trnE-UUC, trnF-GAA, trnG-GCC, trnG-UCC*,trnH-GUG, trnI-CAU(2),trnI-GAU*(2),trnK-UUU*,trnL-CAA(2),trnL-UAA*,trnL-UAG, trnM-CAU, trnN-GUU(2),trnP-UGG, trnQ-UUG, trnR-ACG(2),trnR-UCU, trnS-GCU, trnS-GGA, trnS-UGA, trnT-GGU, trnT-UGU, trnV-GAC(2),trnV-UAC*,trnW-CCA, trnY-GUA, trnfM-CAU	30
Other genes	Maturase	matK	1
	Protease	clpP**	1
	Envelope membrane protein	cemA	1
	Acetyl-CoA carboxylase	accD	1
	c-type cytochrome synthesis gene	ccsA	1
	Translation initiation factor	infA	1
	other	-	-
Genes of unknown function	Conserved hypothetical chloroplast ORF	ycf1(2),ycf15(2),ycf2(2),ycf3**,ycf4	5
Total			115

Note *means gene with one intron; **means gene with two introns; # means pseudo gene; (×2) means number of copies of multi-copy genes

from 157,356 to 158,018 bp, (Table 2) with *I. pedunculosa* being the largest and *I. centrochinensis* being the smallest, differing by 662 bp. The length of the large single-copy region ranged from 86,730 (*I. centrochinensis*) ~ 87,400 bp (*I. pedunculosa*), the length of the small single-copy region was between 18,382 (I. litseifolia) ~ 18,459 bp (*I. cassine Tensaw*), and the length of the single reverse repeat region was between 26,063 (*I. macrocarpa*) ~ 26,125 bp (*I. rotunda*), The lengths of the LSC, SSC, and IR regions show slight variations among the species. The GC content of the chloroplast genomes is around 38.34–38.37%, with similar GC content in the LSC, SSC, and IR regions. While there are differences in the size and GC content of the chloroplast genomes among *Ilex* species, these differences are not pronounced.

Table 3 reveals variations in the length of chloroplast genomes, the number of genes, and the number of protein-coding genes among the 10 *Ilex* species. The total

number of genes, except for *I. centrochinensis*, is 134, with *I. centrochinensis* having 135 genes. The protein-coding genes are either 88 or 89, with *I. verticillata* having 88 and the rest having 89. The rRNA genes are consistently 8 in all species, while the tRNA genes range from 37 to 38, with *I. centrochinensis* having 38 and the others having 37.

The chloroplast genes of the 10 *Ilex* species can be categorized into four classes based on their functions: genes related to photosynthesis (45), genes related to self-replication (59), other genes (6), and genes with unknown functions (5). See Table 4 for detailed classification.

SSR and scattered repeat sequence analysis

Table 3 indicates the presence of SSR sites in the chloroplast genomes of the 10 *Ilex* species, predominantly consisting of single-nucleotide repeats and tri-nucleotide repeats, with no detected penta-nucleotide and hexa-nucleotide repeats. The number of detected mononucleotide repeats ranges from 118 to 128, with *I. cassine* having the highest and I. rotunda having the lowest. The differences in the overall quantities of di-, tri-, and tetra-nucleotide repeats are relatively small. Analyzing nucleotide types, mononucleotide repeats are the most abundant, with A and T repeat types being predominant. This result indicates that the types and quantities of simple repeat sequences in *Ilex* species are relatively conservative. Additionally, these sites are unevenly distributed in the chloroplast genome, but the differences among the 10 *Ilex* species are not significant.

Use of the vmatch v2.3.0 software (http://www.vmatch.de/) combined with Perl scripts was employed to analyze and identify scattered repeat sequences in the chloroplast genomes of ten species of *Ilex* (holly) plants. The analysis revealed that both forward and reverse repeats were detected. Palindromic repeats were only identified in *I. buergeri*, *I. litseifolia*, *I. macrocarpa*, *I. rotunda*, and *I. verticillata*. Complementary repeats were only detected in I. rotunda. These findings suggest minimal variation in scattered repeat sequences among *Ilex* species. In terms of the total number of scattered repeat sequences detected, the total number of repeats in the chloroplast genomes of the ten *Ilex* species ranged from 42 to 50.

IR boundary expansion and contraction

During evolution, the boundaries of the Inverted Repeat (IR) regions undergo certain expansion and contraction. The change in the length of the chloroplast genome is the main reason for the contraction and expansion of the IR region [16]. The analysis of boundary expansion and contraction (Fig. 1.) shows that, except for *I. macrocarpa* (located between the *rps19* and *rpl2* genes), JLB (LSC-IRb) is located within the *rps19* gene; JSB (IRb-SSC) is located within ycf1 and ndhF; JSA (SSC-IRa) is located within ycf1; JLA (IRa-LSC) is located between *rpl2* and *trnH* genes.

Analysis of IR boundaries in other *Ilex* species revealed that *rps19* in *I. asprella* [17], *I. pubescens*, *I. wilsonii*, *I. micrococca*, *I. latifolia*, and *I. macrocarpa* underwent some degree of contraction. Investigations indicate that these hollies generally grow in high-altitude areas [18]. In particular, *I. pubescens* and *I. wilsonii* exhibited contraction in the *ndhF* gene, and the survey results suggest that these two holly species prefer dense forests with high humidity.

I. integra replaces the *ycf1* gene with *trnN*. In our investigation, *I. integra* was found to be distributed in the coastal areas of southeastern Zhejiang and Fujian provinces in China, as well as in southern Japan and the Korean Peninsula [19]. Given its long-term growth in high-salinity and barren coastal areas, multiple studies have confirmed that *I. integra* has high salt tolerance [20, 21], and the *trnN* gene may be an important breakthrough in research on salt-alkali tolerance in hollies.

Overall, the chloroplast genomes of *Ilex* species exhibit conservative characteristics in terms of sequence length, genome composition, and GC content. Their evolutionary relationships are conservative, with minimal structural differences. Boundary expansion and contraction

Table 4 List of simple sequence repeats analysis results

SSR(number)	Region, %			Types of ssr			
speices	LSC	SSC	IR	p1	p2	p3	p4
Ilex bioritsensis	67.50	13.90	18.60	123	4	73	3
Ilex buergeri	66.80	14.50	18.70	121	3	72	3
Ilex cassine 'Tensaw'	68.00	13.70	18.30	128	4	69	6
Ilex centrochinensis	67.50	13.60	18.80	119	4	71	4
Ilex chinensis	67.90	13.50	18.70	123	3	69	4
Ilex litseifolia	68.40	13.30	18.40	122	5	70	4
Ilex macrocarpa	69.00	13.00	18.00	126	4	73	4
Ilex pedunculosa	67.70	13.80	18.50	123	4	71	4
Ilex rotunda	66.80	13.50	19.70	118	4	72	6
Ilex verticillata	67.30	14.90	17.80	127	4	70	4

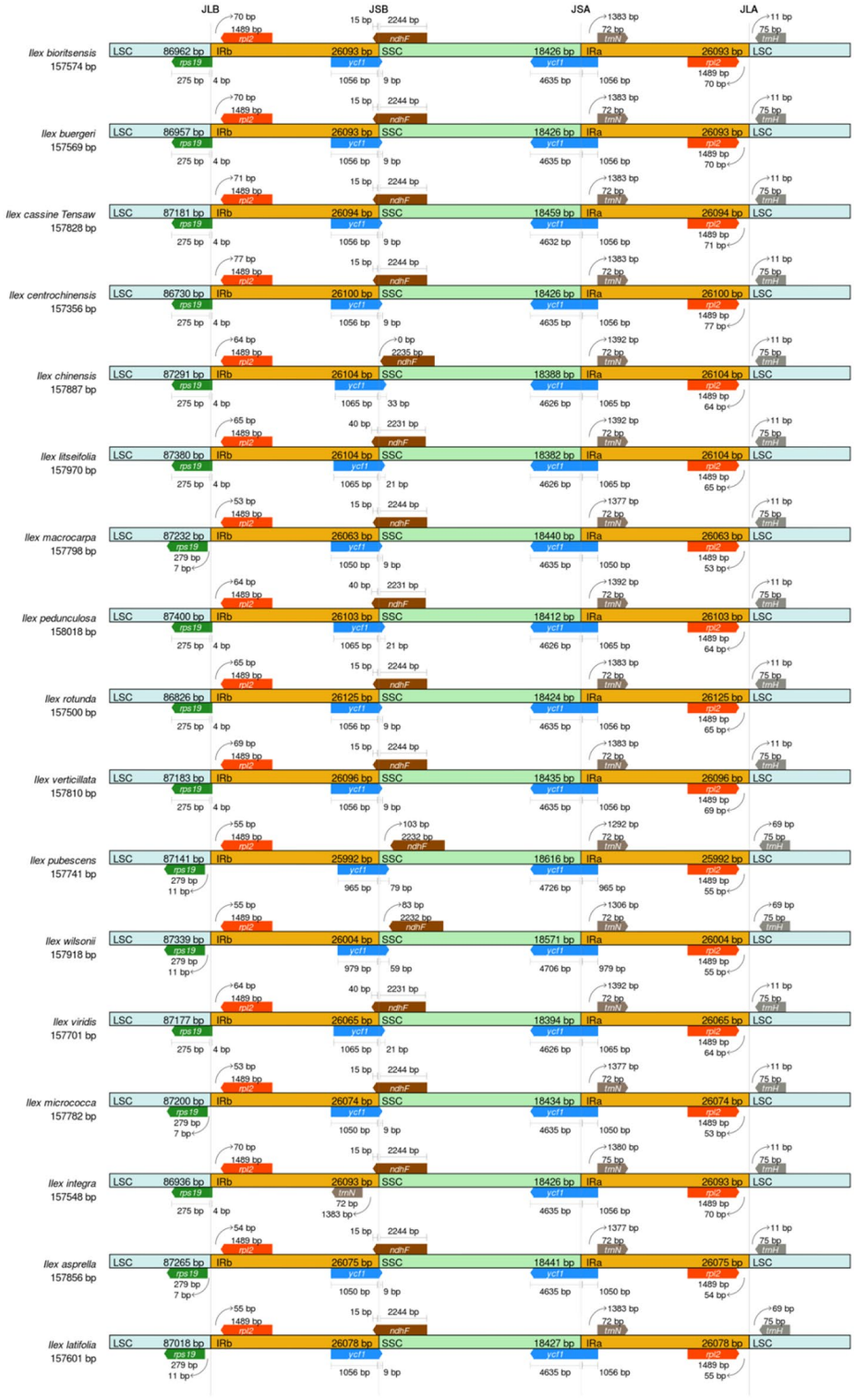

Fig. 1 Visualization of the alignment of the ten *Ilex* chloroplast genomes. LSC indicates longsingle copy region; SSC indicates short single copy region; IRa and IRb indicate two inverted regions.Locationsof divergent hotspot regions are labeled above alignment *Note* The thin lines represent the junction points for each region, and the diagram presents information about genes near the junctions

are relatively stable and similar, with only minor variations observed in individual species.

Relative synonymous codon usage analysis

The frequency of codon usage (RSCU) was analyzed, and a total of 889 protein-coding gene sequences meeting the criteria were selected. Each *Ilex* species had 65 such sequences, and the RSCU values for codons were similar across all species (Fig. 2.). Among the codons with RSCU > 1, there were 29, accounting for 44.62% of the total, indicating a higher frequency of usage, with a preference for A and U endings in the chloroplast genomes of *Ilex* species. The codon for tryptophan (UGG) occurred only once with an RSCU value of 1, suggesting no preference, while stop codons tended to end with UAA. The analysis suggests that the preference for codons among the ten *Ilex* species remains highly consistent.

Phylogenetic analysis of *Ilex* species in the family Aquifoliaceae

Ilex comprises numerous species distributed worldwide, making their evolutionary relationships potentially complex. To elucidate the evolutionary positions of the ten *Ilex* species, a BLASTP comparison of the chloroplast genomes of 23 *Ilex* species with those of *Anacardium* occidentale and *Maytenus guangxiensis* (Celastraceae) was performed, and a phylogenetic tree was constructed (Fig. 3). Using the NODE TIME dataset (http://www.timetree.org), key branch evolution times were determined, including between *I. pedunculosa* and *I.chinensis* (12.3–36.4 MYA), *Maytenus guangxiensis* and *Anacardium occidentale* (102.0–112.5 MYA), and *M. guttatus* and *V. viniera* (12.3–36.4 MYA), used for time calibration. Species divergence times were estimated using the MCMCtree module in PAML v4.9. The tree was then beautified using FigTree v1.4.4 and iTOL (Interactive Tree Of Life, embl.de). This tree represents the evolutionary relationships among 23 *Ilex* species, illustrating their common ancestors and evolutionary differentiation. Branch lengths provide an estimate of the time gaps between species. The phylogenetic analysis of the complete cp. genome indicates that extant species within the *Ilex* genus began diversifying during the Middle Miocene of the Tertiary Period, around 22.71 MYA. This period of divergence is likely correlated with the substantial fluctuations in global average temperatures occurring between 15 and 25 MYA [22].

Among the ten species of *Ilex* that we examined, *I.latifolia* and *I. chinensis* demonstrate a closer relationship, diverging approximately 8.3 MYA. Conversely, the divergence time for *I. bioritsensis* and *I.buergeri* is estimated to be around 1.03 MYA, indicating a closer genetic affinity.

Discussion and outlook

In this study, we employed bioinformatics tools to analyze the chloroplast genomes of ten *Ilex* species. Our results revealed that the genome size range from approximately 157,356 to 158,018 kb, exhibiting higher AT content compared to GC content. Moreover, these genomes possess a typical chloroplast genome structure comprising about 134 genes alongside intergenic regions, with rRNA being the most conserved. Through genomic sequence analysis, we identified core genes encoding chloroplast functions, including those associated with photosynthesis, respiration, tRNA, and rRNA genes. Studies have indicated that repetitive sequences significantly influence sequence variations within cp. genomes [23]. Their abundant distribution and high polymorphism hold substantial significance for genetic diversity analysis in plants and for molecular marker identification in germplasm resources [24]. Consistent with findings in other angiosperms, SSR sites within the *Ilex* chloroplast genome exhibit non-uniform distribution, mostly concentrated in the LSC region [25]. A/T mononucleotide repeat types are prevalent, potentially attributed to polyadenylation at the mRNA termini in chloroplast genes. The overall variation in the type and quantity of simple repeat sequences remains relatively conservative. Additionally, during chloroplast replication, A/T chain separation is relatively more prone to slippage mismatch than the G/C chain [26].Previous studies indicate the significant role of the Inverted Repeat (IR) region in preserving vital genes and stabilizing the cp. DNA structure [27], representing the most conserved region within the chloroplast genome. All of the *Ilex* chloroplast genomes show a similar arrangement in the boundary regions. In brief, SSC/IR boundaries were found within ycf1 in all of the Ilex species. However, slight differences in the distance between each gene and the boundaries across the ten species imply a higher degree of conservation in the chloroplast genome within this genus. Analyzing the boundary information of the IR-LSC/SSC regions holds significance for understanding differences in chloroplast genome structure and species evolution.

Our findings align with those of Xin Yao et al. [28], confirming the evolutionary relationships of genus Ilex plants based on chloroplast genome comparisons and evolutionary analyses. At the same time, our study found that *Ilex asprella* is more closely related to *Ilex macrocarpa* than previously reported by Yan Chen et al. [29]. Compared to Yancai Shi's study [30], we identified two species, *Ilex bioritsensis* and *Ilex buergeri*, that are more similar to *Ilex latifolia*. These species began to diverge from *Ilex latifolia* approximately 1.03 MYA. Building upon the findings of Zhenyu Shan et al. [31], we added *Ilex rotunda* and *Ilex verticillata* to the phylogenetic tree. These species are more closely related to *Ilex pubescens*

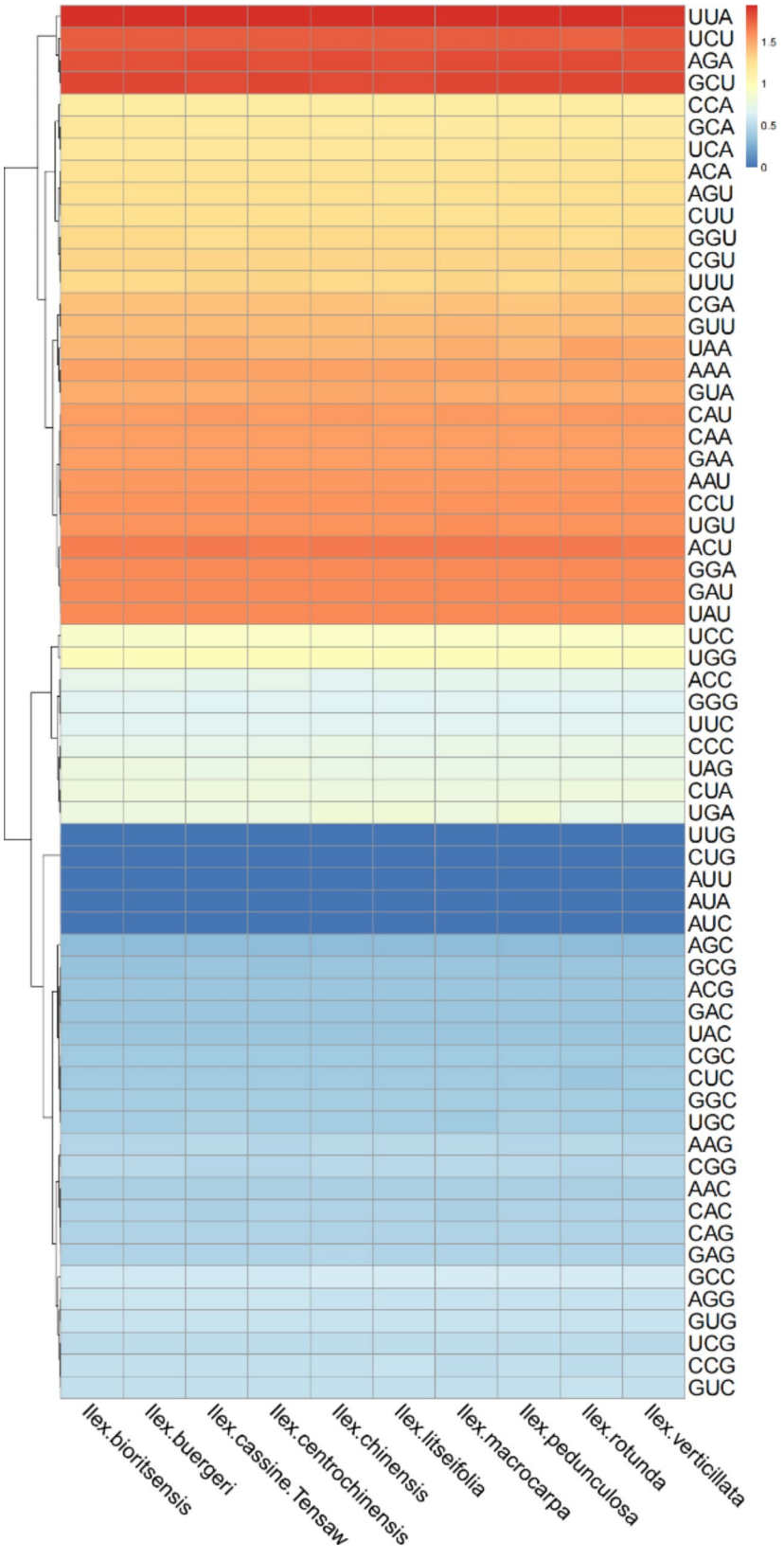

Fig. 2 Relative Synonymous Codon Usage (RSCU) analysis heatmap

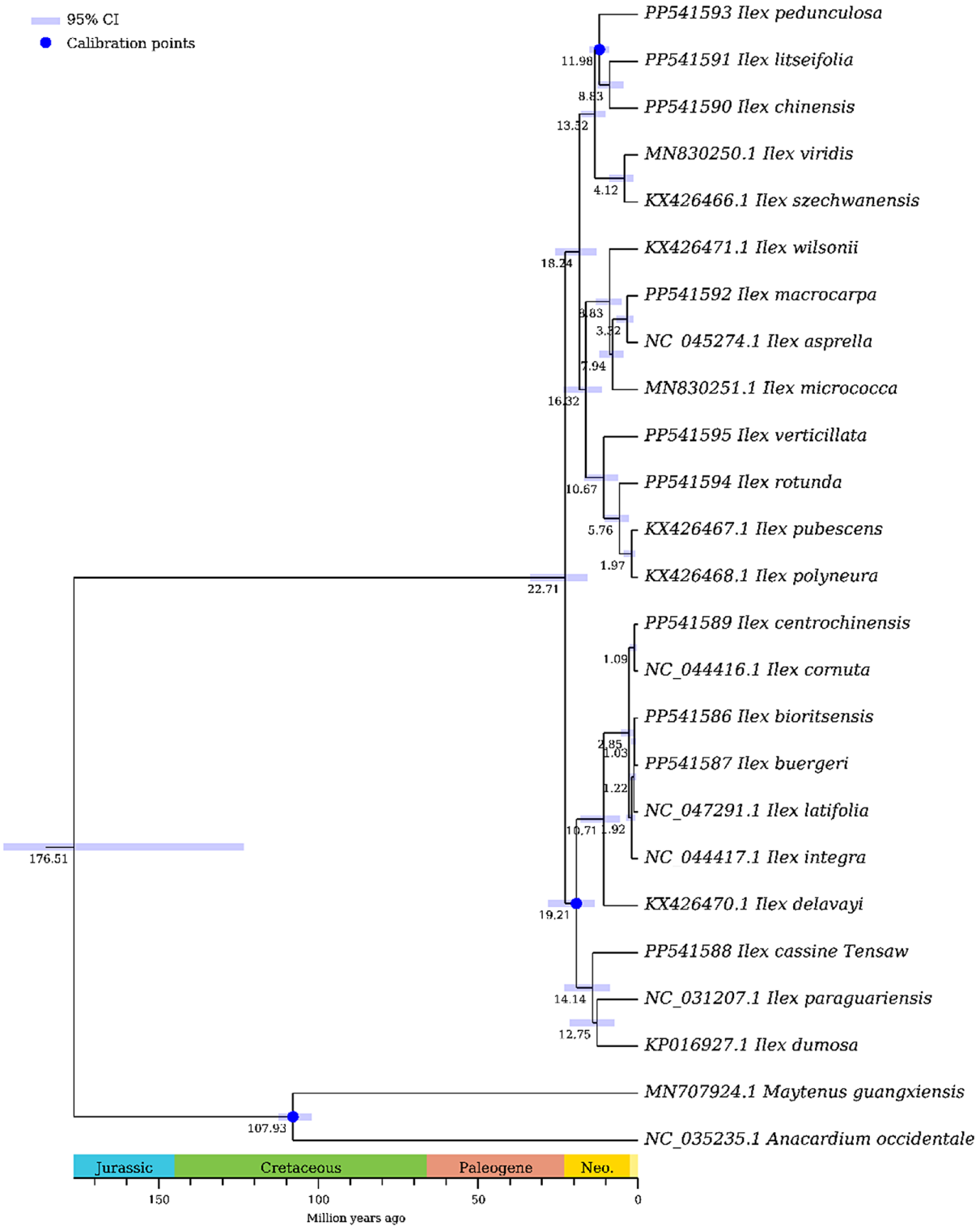

Fig. 3 Divergence time estimation based on chloroplast genome sequences. Each node shows the divergence times, and the blue bars provide the 95% greatest posterior density interval for each node age

than *Ilex wilsonii*. Our results are generally consistent with those of several previous studies; however, we have expanded the dataset by including 10 new species of holly and studying their divergence times. These investigations aid in inferring the origin and evolutionary history of Ilex plants. Moreover, they facilitate the analysis of genomic differences and variations among different species within the genus *Ilex*, enabling a deeper understanding of their evolutionary and phylogenetic relationships.

Through the analysis of genome sequences and structures, we can uncover species relationships, elucidate species evolutionary histories, and delve into the adaptive mechanisms and biodiversity of *Ilex* plants. Furthermore, this study serves as a vital reference for genetic improvement, germplasm resource conservation, and taxonomic research within the genus *Ilex*. However, current research still has limitations, such as the need for further investigation into functional genes within the chloroplast, the evolutionary mechanisms of the genome, and the interactions between the genome and the environment. Future studies demand more samples and comprehensive genomic information to deepen our understanding of the evolution of *Ilex* plants.

Acknowledgements
The authors are grateful to Zhejiang A&F University.We thank the Ningbo Academy of Agricultural Sciences for providing access to the plant materials under their care.We are grateful to the editor and reviewers for their constructive comments and suggestions, which have improved the quality of the manuscript.

Author contributions
Hu Jiaxin is responsible for the processing of experimental materials, data measurement and analysis and the writing of articles, Yan Daoliang is responsible for the overall design of the research project, data analysis and revision of follow-up articles, Yuan Huwei is responsible for sample collection, Zhang Jianhong is responsible for sample collection and identification, and Zheng Bingsong is responsible for the formulation of experimental design ideas.

Funding
The project was supported by the Open Project of Ningbo Key Laboratory of Characteristic Horticultural Crops in Quality Adjustment and Resistance Breeding. (NBYYL20230003)

Data availability
Sequence data that support the findings of this study have been deposited in the NCBI database with the primary accession code PRJNA1090273.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 26 February 2024 / Accepted: 7 October 2024
Published online: 31 October 2024

References

1. Dobrogojski J, Adamiec M, Robert L. The chloroplast genome: a review. Acta Physiol Plant, 2020, 42(6).
2. Liu C, Shi L, Zhu Y, et al. CpGAVAS, an integrated web server for the annotation, visualization, analysis, and GenBank submission of completely sequenced chloroplast genome sequences. BMC Genomics. 2012;13(1):715–715.
3. Daniell H, Lin CS, Yu M, et al. Chloroplast genomes: diversity, evolution, and applications in genetic engineering. Genome Biol. 2016;17(1):1–29.
4. Yagi Y, Shiina T. Recent advances in the study of chloroplast gene expression and its evolution. Front Plant Sci. 2014;5(1):61.
5. Wu ZY, Raven P, Beijing. & St Louis: Science Press & Missouri Botanical Garden, 2020:1994–2009.
6. Zhang CX, Lin CZ, Xiong TQ, et al. New triterpene saponins from the root of *Ilex pubescens*. Fitoterapia. 2010;81(7):788–92.
7. Li L, Xu LJ, Ma GZ, et al. The large-leaved Kudingcha (*Ilex Latifolia* 'Thunb' and *Ilex Kudingcha* C.J. Tseng): a traditional Chinese tea with plentiful secondary metabolites and potential biological activities. J Nat Med. 2013;67(3):425–37.
8. Wan P, Peng YJ, Chen GJ, et al. Dicaffeoylquinic acids from *Ilex Kudingcha* attenuate dextran sulfate sodium-induced colitis in C57BL/6 mice in association with the modulation of gut microbiota. J Funct Foods. 2019;61(C):1.
9. Wan P, Xie MH, Chen GJ et al. Anti-inflammatory effects of dicaffeoylquinic acids from *Ilex Kudingcha* on lipopolysaccharide-treated RAW264.7 macrophages and potential mechanisms. Food Chem Toxicol, 2019, 126.
10. Shi L, Li NW et al. Molecular evidence for the Hybrid Origin of *Ilex dabieshanensis* (Aquifoliaceae). PLoS ONE, 2016, 11(1).
11. DOYLE J. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem Bull. 1987;19:11–5.
12. Goodwin S, Mcpherson JD, Mccombie WR. Coming of age: ten years of next-generation sequencing technologies. Nat Rev Genet, 2016.
13. Bankevich A, Nurk S, Antipov D, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol. 2012;19(5):455–77.
14. Boetzer M, Pirovano W. Toward almost closed genomes with GapFiller. Genome Biology,13,6(2012-06-25), 2012, 13(6):R56.
15. Alexandra D, Tamir T. The effect of tRNA levels on decoding times of mRNA codons. Nucleic Acids Res, 2014(14):9171.
16. Raubeson LA, Peery R, Chumley TW, et al. Comparative chloroplast genomics: analyses including new sequences from the angiosperms Nuphar advena and Ranunculus macranthus. BMC Genomics. 2007;8(1):174.
17. Liu LR. *Ilex Asprella* (Hook.et Arn.) Champ. Jiangxi Forestry Sci Technol, 2010(1):22–22. (in Chinese).
18. Lin KM. Investigation and Analysis of Wild Plant Resources of *Ilex* in Daiyunshan Nature Reserve. Inner Mongolia Forestry Surv Des. 2017;40(6):4. (in Chinese).
19. Xu BF, Wang GM, Wang MO, et al. Distribution and propagation of *Ilex integra* 'Thunb'. And *Ilex crenata* 'Thunb'. Chin Wild PlantResour. 2007;26(4):63–5. (in Chinese).
20. Zhang Q, Zhou P, Liu C, et al. Comparison of transcriptomic activity of *Ilex Integra* and *I. purpurea* roots with NaCl treatments. J Nanjing Forestry Univ (Natural Science). 2022;46(3):99–108. (in Chinese).
21. Du YL, Chen YZ, Qian AG et al. Salt tolerance of four typical coastal plants. J Appl Oceanogr 2022,41(1):15–24. (in Chinese).
22. Scotese CR, Song H, Mills B et al. Phanerozoic Paleotemperatures: The Earth's Changing Climate during the Last 540 million years. Earth Sci Rev, 2021(11).
23. Bausher MG, Singh ND, Lee SB, et al. The complete chloroplast genome sequence of Citrus sinensis (L.) Osbeck var 'Ridge Pineapple': organization and phylogenetic relationships to other angiosperms. BMC Plant Biol. 2006;6(1):21.
24. Wenpan D, Chao X, Tao C, et al. Sequencing Angiosperm Plastid genomes made Easy: A Complete Set of Universal primers and a case study on the phylogeny of Saxifragales. Genome Biology & Evolution; 2013. pp. 989–97. 5.
25. Luo C, Huang W, Sun H, et al. Comparative chloroplast genome analysis of Impatiens species (Balsaminaceae) in the karst area of China: insights into genome evolution and phylogenomic implications. BMC Genomics. 2021;22:571.
26. Keller J, Rousseau-Gueutin M, Martin GE, et al. The evolutionary fate of the chloroplast and nuclear rps16 genes as revealed through the sequencing and comparative analyses of four novel legume chloroplast genomes from Lupinus. Dna Research An International Journal for Rapid Publication of Reports on Genes & Genomes; 2017. pp. 343–58. 4.

27. Palmer JD, Thompson WF. Chloroplast DNA rearrangements are more frequent when a large inverted repeat sequence is lost. Cell. 1982;29(2):537–50.
28. Yao X, Tan YH, Liu YY, et al. Chloroplast genome structure in *Ilex* (Aquifoliaceae). Sci Rep. 2016;6(1):28559.
29. Chen Y, Chen HX, Li HL et al. Complete plastome sequence of *Ilex Asprella* (Hooker and Arnott) Champion ex Bentham (Aquifoliaceae), a Chinese folk herbal medicine. Mitochondrial DNA Part B, 2019.
30. Shi Y, Liu B. Complete chloroplast genome sequence of *Ilex Latifolia* (Aquifoliaceae), a traditional Chinese tea. Mitochondrial DNA Part B, 2020.
31. Shan Z, Zhou B, Li Y, et al. Complete chloroplast genome sequence and phylogenetic analysis of Ilex Viridis Champ. Ex Benth. Mitochondrial DNA Part B. 2020;5(1):914–5.

Publisher's note