Phyton-
International Journal of
Experimental Botany

Tech Science Press

http://doi.org/10.32604/phyton.2026.084482

ARTICLE

Complete Chloroplast Genome of *Pyrola japonica*: Characterization and Phylogenetic Analysis

Chayanee Chairattanawat[1], Junghwa Kang[2], Jaewook Kim[1,*] and Bae Young Choi[3,*]

[1]Department of Biology Education, Korea National University of Education, Cheongju, Republic of Korea
[2]Hantaek Botanical Garden, Hantaek-ro 2, Yongin-si, Gyeonggi-do, Republic of Korea
[3]School of Liberal Arts and Sciences, Korea National University of Transportation, Chungju, Republic of Korea
*Corresponding Authors: Jaewook Kim. Email: jwkim@knue.ac.kr; Bae Young Choi. Email: baeyoung@ut.ac.kr

Received: 23 April 2026; Accepted: 01 June 2026; Published: 30 July 2026

ABSTRACT: *Pyrola japonica*, a member of the Ericaceae family, is a significant medicinal herb and a key model organism in mycorrhizal research, yet its chloroplast (cp) genome has not been fully characterized. Therefore, this study aims to sequence and analyze the complete plastid genome of *P. japonica*. The complete cp genome of *P. japonica* was determined to be 168,146 bp in length, exhibiting a characteristic quadripartite structure with a total GC content of 35.1%. A total of 136 genes were annotated, comprising 65 protein-coding genes, 45 transfer RNA (tRNA) genes, 8 ribosomal RNA (rRNA) genes, and 18 pseudogenes. Amino acid analysis of the protein-coding sequences revealed that Leucine was the most abundant, while Cysteine was the least frequent. Structural analysis identified 789 long dispersed repeat sequences and 112 Simple Sequence Repeats (SSRs), with the majority of SSRs distributed across the large single-copy (LSC), inverted repeat (IR), and small single-copy (SSC) regions, respectively. Phylogenetic reconstruction placed *P. japonica* as a sister group of *P. decorata*, clarifying its evolutionary position within the *Pyrola* lineage. Relative evolutionary divergence analysis revealed that *Pyrola* diverged earlier than the other examined Ericaceae taxa. This study provides the first comprehensive cpDNA map for *P. japonica*, providing essential genetic resources for molecular evolution and species identification.

KEYWORDS: *Pyrola japonica*; Ericaceae; complete chloroplast genome; cpDNA; evolutionary relationship

1 Introduction

Pyrola japonica Klenze, a member of the Ericaceae family, is a prominent medicinal herb widely distributed across East Asia, including South Korea, Japan, and China [1–3]. It is recognized for its diverse secondary metabolites, such as phenolic glycosides, which contribute to its therapeutic properties [4,5]. Specifically, in an osteoclast-induced bone pit formation model, *P. japonica* extracts demonstrated anti-resorptive activity via the transcriptional downregulation of osteoclastogenic genes [6]. Apart from mitigating bone resorption, *P. japonica* extracts show significant antibacterial activity against *Bacillus subtilis* [7]. It also exhibited potent antioxidant activity, highlighting its potential in preventing oxidative stress-related diseases [5]. Collectively, these findings underscore the pharmacological potential of *P. japonica* as a candidate for developing novel natural therapeutics.

Beyond its pharmacological significance, *P. japonica* is a focal point in mycorrhizal research due to its mixotrophic nature as it employs a dual carbon-acquisition strategy, obtaining carbon through both photosynthesis and symbiotic associations with mycorrhizal fungi [2]. *P. japonica* formed arbutoid

mycorrhizal associations, with colonization rates typically increasing under shaded conditions. This correlation suggested that the species undergoes dynamic shifts in both fungal symbiont composition and carbon acquisition strategies as a physiological response to varying light availability [8]. Consequently, investigating photosynthetic variations during these transitions is essential to understanding the physiological mechanisms that compensate for light limitation. Such investigations require comprehensive plastome data to identify the specific genes involved in these adaptive photosynthetic processes.

Despite its importance, the genus *Pyrola* remains one of the most taxonomically challenging groups within the Ericaceae [1]. To address this complexity, chloroplast (cp) genome assembly offers a robust bioinformatic solution. The cpDNA is typically 110–170 kb and highly conserved. In spite of its small size, it is ideal for phylogenetic research due to its uniparental inheritance, which is usually maternal, high copy number, stable structure, and lack of recombination, making it easy to sequence and assemble [9]. Furthermore, due to its highly conserved nature, complete chloroplast genome sequences can act as effective DNA barcodes for precise species identification [10]. Therefore, this method serves as a high-resolution tool for resolving phylogenetic relationships and taxonomic ambiguities. Beyond evolutionary studies, a complete chloroplast genome provides a stable framework for chloroplast genetic engineering, facilitating both fundamental research and the enhancement of plant traits.

To date, despite over 40 species worldwide, only *P. atropurpurea*, *P. rotundifolia*, and *P. decorata* have complete chloroplast genome sequences provided [1,11], whereas the complete chloroplast genome sequences of *P. japonica* has not been reported in existing literature. In this study, we sequenced, assembled, and characterized the complete chloroplast genome of *P. japonica*, providing a genomic resource for future botanical and biotechnological applications.

2 Materials and Methods

2.1 Sampling

Whole plant of *P. japonica* was collected in Yeondang-ri, Nam-myeon, Yeongwol-gun, Gangwon-do, the Republic of Korea (37°09.3100′ N, 128°24.9730′ E) in June 2025. Frozen leaves were used for DNA extraction (Supplementary Fig. S1).

2.2 DNA Extraction and Sequencing

Total genomic DNA was extracted from the collected leaf samples using a SmartGene Plant DNA Extraction kit (SmartGene, Daejeon, Republic of Korea), following the manufacturer's instructions. The genomic DNA library was prepared using an xGen™ DNA Lib Prep EZ kit (Integrated DNA Technologies, Coralville, LA, USA). The generated Illumina libraries were subjected to high-throughput sequencing on an Illumina NovaSeq 6000 instrument (Illumina, San Diego, CA, USA).

2.3 Assembly and Annotation

Quality control of the raw sequence data was executed via Trimmomatic v. 0.39, applying filtering thresholds of 15 for both leading and trailing bases, and retaining reads with a minimum length of 150 bp [12]. The filtered reads were assembled into the chloroplast genome using GetOrganelle v. 1.7.7.1 [13], with the *P. decorata* chloroplast genome (accession number PV200167.1) as a reference. Following assembly, Bandage v. 0. 8. 1 was utilized to refine the contig topologies [14]. Gene annotation was performed using GeSeq ([15], accessed 10 Mar 2026). Any detected annotation inaccuracies were manually corrected with UGENE v. 53.1 [16]. A circular map of the *P. japonica* chloroplast genome was generated using OGDRAW

([17], accessed 19 Mar 2026). The complete chloroplast sequence of *P. japonica* was deposited in GenBank of the National Center for Biotechnology Information (NCBI) under the accession number PZ137749.

2.4 Repeated Sequence Analysis and Codon Usage

To detect dispersed repeats, we utilized vmatch v. 2. 3. 1 [18]. The search criteria were configured with a 30 bp minimum length and a Hamming distance of 3, screening across all four orientations (forward, palindromic, reverse, and complement). For simple sequence repeat (SSR) characterization, MISA v1.0 was deployed [19]. The thresholds for SSR identification required at least 10 repeat units for mononucleotides, 5 for dinucleotides, 4 for trinucleotides, and 3 for tetra-, penta-, and hexanucleotides, alongside a minimum spatial separation of 100 bp between adjacent SSRs. RSCUcaller v. 1.0 was applied to compute and graphically represent the relative synonymous codon usage (RSCU) bias [20].

2.5 Chloroplast Genomes Comparison Analysis

The contraction and expansion of LSC, SSC, and IR region boundaries were visualized using CPJSdraw v. 0. 0. 1 in the four *Pyrola* chloroplast genomes, including *P. atropurpurea*, *P. decorata*, *P. rotundifolia*, and *P. japonica* [21]. Complete chloroplast genome sequences of *Pyrola* species were compared using the mVISTA online tool (http://genome.lbl.gov/vista/mvista/submit.shtml). This analysis was performed using the Shuffle-LAGAN mode [22]. The gene nucleotide diversity (Pi) value of ortholog genes in four *Pyrola* chloroplast genomes was counted using DNAsp v. 5. 0 [23].

2.6 Phylogenetic Analysis

Complete plastome sequences from 41 taxa (detailed in Supplementary Table S1) were subjected to multiple sequence alignment via MAFFT v. 7.526 [24]. Maximum likelihood (ML) phylogenetic reconstruction was then carried out employing IQ-TREE v. 2.4.0 [25]. The ModelFinder implemented in IQ-TREE automatically determined the optimal nucleotide substitution model. Node reliability was assessed through 1000 ultrafast bootstrap replicates, and the tree was rooted using *Capsicum chacoense* and *Capsicum galapagoense* as outgroup taxa. The generated tree topology was graphically rendered utilizing the Interactive Tree of Life (iTOL) v.7 [26].

2.7 Relative Evolutionary Divergence Analysis

Relative evolutionary divergence analysis was performed according to the previous report with minor modifications [27]. Chloroplast genome sequences from 41 species listed in Supplementary Table S1 were aligned using MAFFT v. 7.526 [24]. Then, the alignment data was trimmed with trimAl v. 1.5 with the automated option [28]. The resulting data were imported into BEAST v. 2.7.8 for Bayesian inference [29]. BEAST analysis parameters were configured with a strict molecular clock model, a Yule speciation tree prior [30], and the HKY nucleotide substitution model. To guarantee robust analytical convergence and thorough sampling of the parameter space, we executed two independent Markov chain Monte Carlo (MCMC) runs [31]. Each chain spanned 100 million generations, with trees and parameters logged every 10,000 states. The Interactive Tree of Life (iTOL) v.7 was utilized to graphically visualize the resulting tree [26].

3 Results

3.1 Chloroplast Genome Structure and Protein-Coding Genes

The assembled chloroplast genome of *P. japonica* is 168,146 bp in length with a GC content of 35.1%. It exhibits a typical quadripartite structure, consisting of a large single-copy (LSC) region (104,725 bp), a small

single-copy (SSC) region (11,758 bp), and a pair of inverted repeat (IR) regions (25,818 bp each) (Fig. 1). A total of 136 genes were annotated, comprising 65 protein-coding genes (PCGs), 45 transfer RNA (tRNA) genes, eight ribosomal RNA (rRNA) genes, and 18 pseudogenes (Table S2). Among the PCGs, 35 are involved in photosynthesis: seven Photosystem I genes, 14 Photosystem II genes, six cytochrome *b/f* complex genes, six ATP synthase subunits, one Rubisco large subunit (*rbcL*), and one photosystem biogenesis factor (*pbf1*). Notably, all 12 NADH-dehydrogenase-like (*ndh*) genes are truncated and were consequently annotated as pseudogenes. An additional 25 genes are categorized as self-replicating, including nine large ribosomal subunit genes, 12 small ribosomal subunit genes, and four DNA-dependent RNA polymerase genes. The *rps12* gene was identified as having a trans-spliced structure, where the gene is fragmented between a 5′ exon in the LSC and duplicate 3′ exons residing in the two IR regions. Four additional functional genes were identified. Among these, the C-type cytochrome synthesis gene (*ccsA*) was annotated as one intact and one truncated copy, situated at the IRb-SSC and SSC-IRa junctions, respectively.

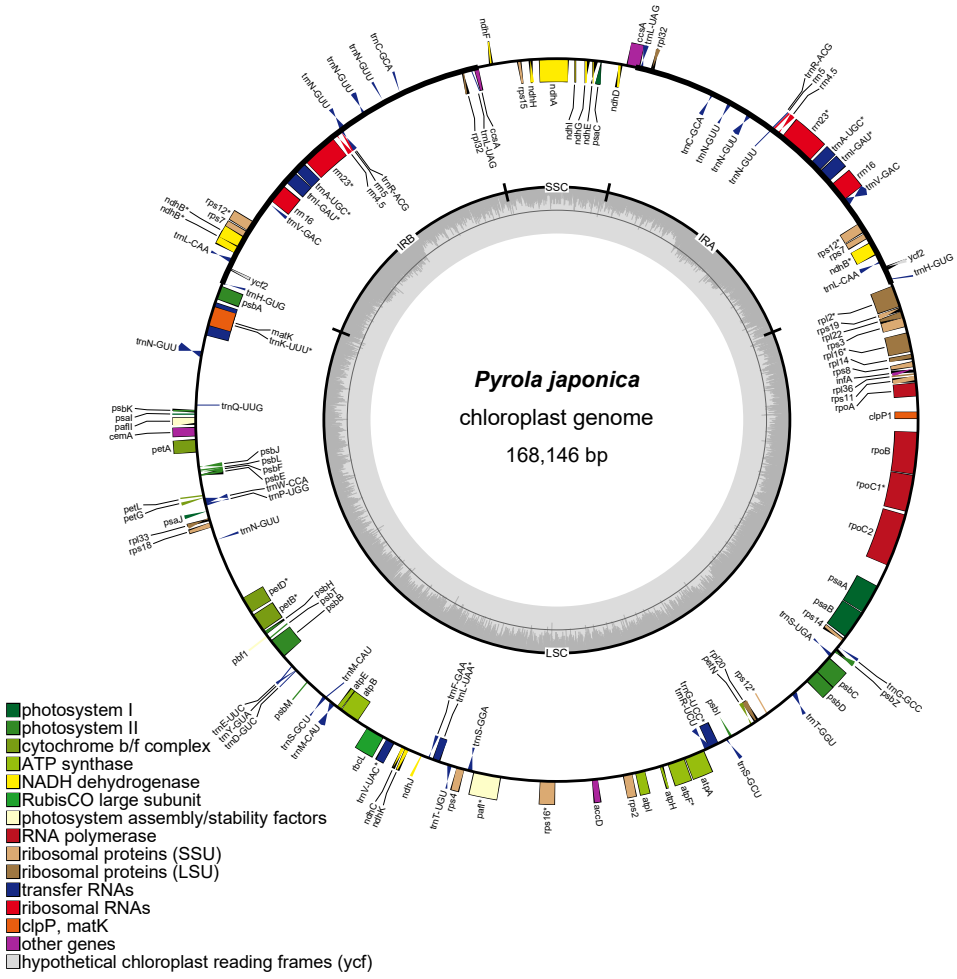

Figure 1: Circular map of the plastid genome generated with OGDRAW. Protein-coding genes, rRNAs, and tRNAs are displayed on the outer circle and are color-coded by functional category according to the OGDRAW scheme. Genes drawn on the outside of the circle are transcribed clockwise, whereas genes on the inside are transcribed counterclockwise. The LSC, SSC, and the two inverted repeat regions (IRa and IRb) are indicated. The inner gray plot represents base-composition variation across the genome (GC content). Intron-containing genes are marked with an asterisk (*).

3.2 Codon Bias in P. japonica Chloroplast Genome

Relative synonymous codon usage (RSCU) values were calculated based on 75 protein-coding sequences (CDS), encompassing a total of 15,457 codons. Leucine (Leu) was the most abundant amino acid, encoded by 1587 codons (10.27%), whereas cysteine (Cys) was the least frequent, represented by only 159 codons (1.03%) (Supplementary Fig. S2). Out of the 64 possible codons, 30 exhibited RSCU values > 1.00, indicating their preferred usage. Notably, all of these preferred codons ended in A or U at the third position, with the sole exception of UUG (Leu). The highest RSCU value was observed for the UUA codon encoding leucine (2.19) (Fig. 2). These results indicate a strong AT-bias at the third codon position. When comparing the codon usage profiles across the four *Pyrola* species, *P. japonica* exhibited highly conserved RSCU patterns with its close relatives. Similar to *P. japonica*, Leucine and Cysteine were consistently the most and least abundant amino acids, respectively, in *P. atropurpurea*, *P. decorata*, and *P. rotundifolia*. Furthermore, the high preference for the UUA codon was a universally shared trait among the four *Pyrola* plastomes (Table S3).

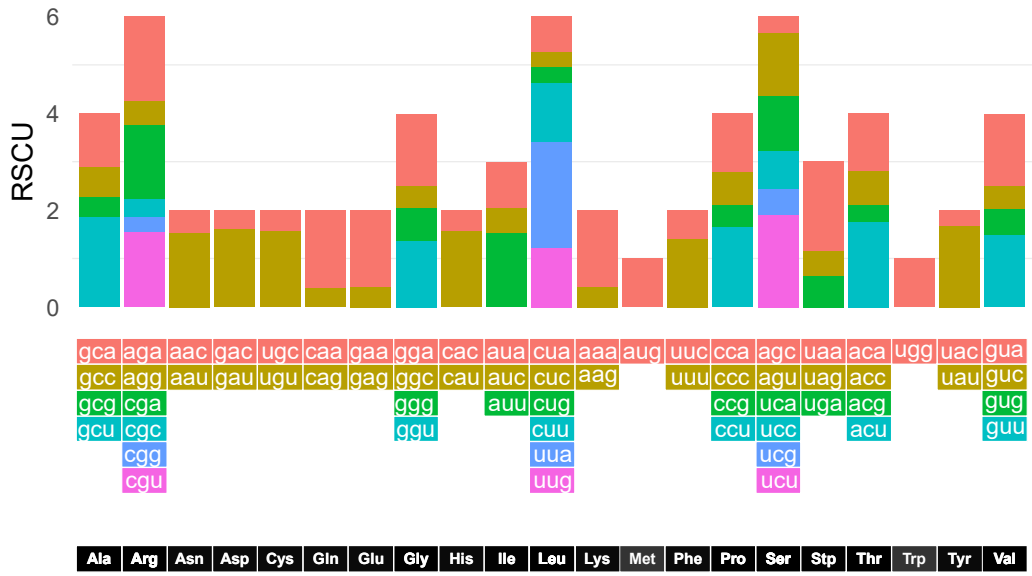

Figure 2: RSCU profile of *P. japonica* chloroplast genome. Different colors represent different codons of the amino acid. Individual amino acid codons are differentiated by distinct colors. The relative height of each colored bar corresponds to the usage frequency of codon.

3.3 Distribution of Repeat Sequences and SSRs in the P. japonica Chloroplast Genome

A total of 789 long dispersed repeat sequences were identified in the *P. japonica* chloroplast genome (Fig. 3). These comprised 337 forward repeats (42.7%) and 452 palindromic repeats (57.3%). No reverse or complement repeats were detected. Notably, the largest palindromic repeat corresponds to the 25,818 bp inverted repeat (IR) region itself, whereas the lengths of the remaining repeats predominantly ranged from 30 to 125 bp.

Furthermore, we characterized SSRs across the chloroplast genome, identifying a total of 112 SSRs (Fig. 4). These consisted of 50 mono-, 14 di-, 27 tri-, 14 tetra-, six penta-, and one hexanucleotide repeats. Mononucleotide and dinucleotide repeats were predominantly composed of A/T motifs, reflecting the strong AT-bias typically observed in chloroplast genomes. Regarding their spatial distribution across the quadripartite structure, the majority of SSRs were located in the LSC region (75), followed by the IR (21)

and SSC (11) regions. Analysis of their genomic contexts revealed that 93% (107) of the SSRs were situated in non-coding regions, while only 7% (5) were located within protein-coding genes (Fig. S3).

For a broader evolutionary context, we compared the repeat sequences and SSRs of *P. japonica* with those of *P. atropurpurea*, *P. decorata*, and *P. rotundifolia* (Table S3). The total numbers of SSRs across the four *Pyrola* species were highly comparable. Across all four species, mononucleotide repeats were dominant. Regarding long dispersed repeats, reverse and complement repeats were absent in all four species. *P. japonica* contained 789 repeats, which was fewer than *P. atropurpurea* (1240) and *P. decorata* (4333), but greater than *P. rotundifolia* (216).

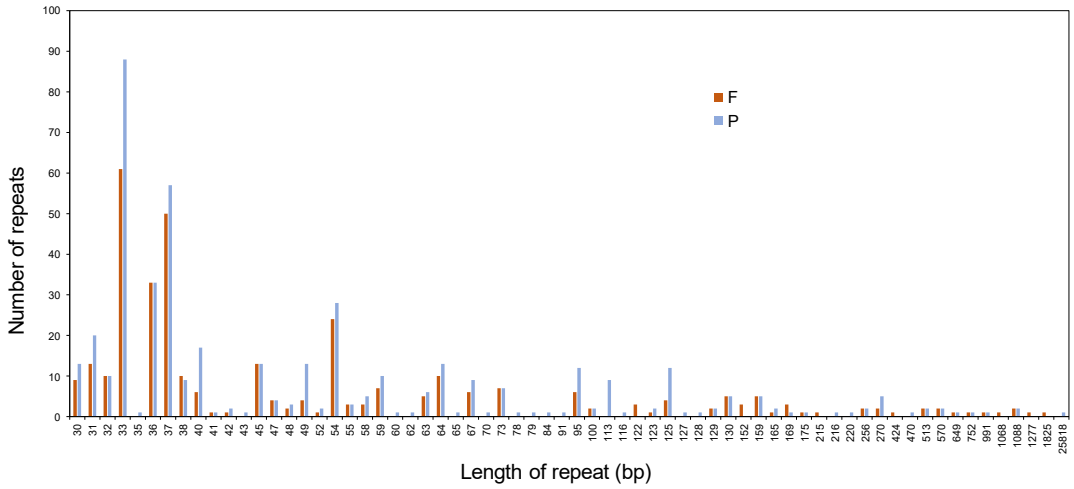

Figure 3: Long fragment repeat distribution in the *P. japonica* chloroplast genome. F: forward repeat; P: palindromic repeats.

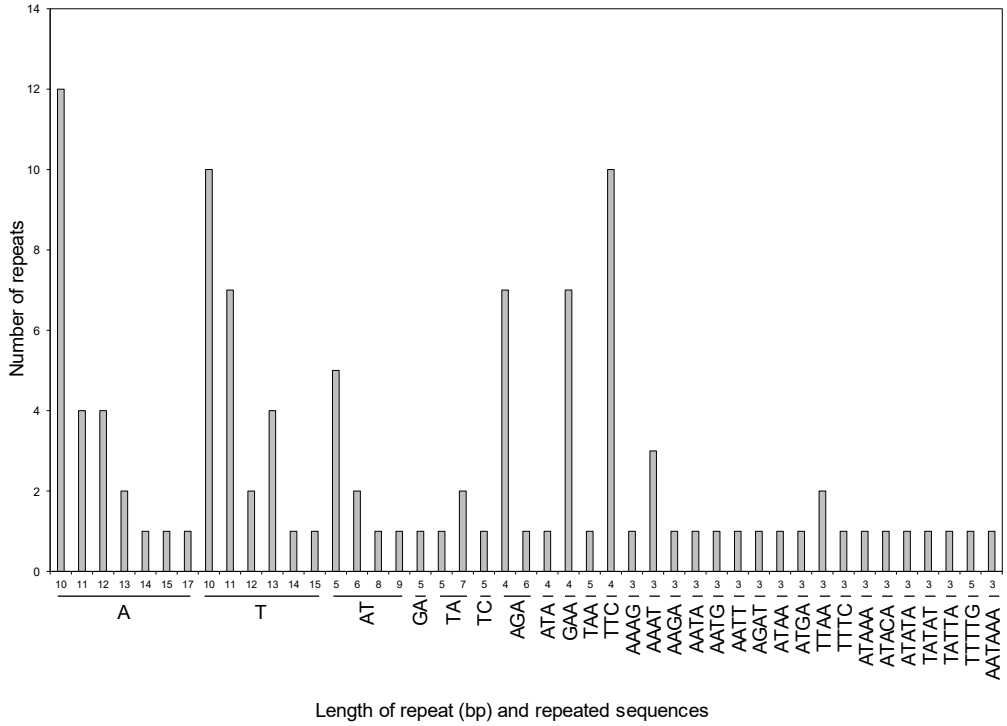

Figure 4: SSR type in the chloroplast genome of *P. japonica*.

3.4 Comparison of Chloroplast Genome Boundaries in the Genus Pyrola

To compare the structural variation of the *P. japonica* chloroplast genome with closely related taxa, we conducted a comparative analysis of the IR and single-copy boundaries among four *Pyrola* species: *P. japonica*, *P. atropurpurea*, *P. decorata*, and *P. rotundifolia* (Fig. 5). Among the analyzed species, the SSC region of *P. rotundifolia* was the largest, whereas that of *P. atropurpurea* was the smallest. *P. decorata* possessed the largest IR and LSC regions. Conversely, the smallest IR and LSC regions were observed in *P. rotundifolia* and *P. japonica*, respectively. Across the four plastomes, the gene arrangements at the four structural junctions—JLB (LSC-IRb), JSB (IRb-SSC), JSA (SSC-IRa), and JLA (IRa-LSC)—were examined. The genes situated in proximity to the JSB and JLA junctions exhibited highly conserved configurations across all analyzed chloroplast genomes. Specifically, the *ccsA* gene traversed the JSB junction, extending from the IRb region into the SSC region. The length of this expansion into the SSC region was strictly conserved at 680 bp in *P. japonica*, *P. atropurpurea*, and *P. decorata*, and only slightly shorter at 671 bp in *P. rotundifolia*. At the JLA junction, the *trnH* and *psbA* genes were consistently located entirely within the IRa and LSC regions, respectively.

In contrast, the gene arrangements at the JLB and JSA junctions exhibited notable structural variations among the four species, indicating dynamic boundary shifts. At the JLB junction, the *trnH* gene was located within the IRb region, while *rpl2* was situated in the LSC region for all species except *P. rotundifolia*. At the JSA junction, *rps15* and *trnL* were distributed in the SSC and IRa regions, respectively, in *P. atropurpurea* and *P. decorata*. Conversely, in *P. japonica* and *P. rotundifolia*, the *ndhF* and *ccsA* genes occupied the SSC and IRa regions, respectively, at this junction. Overall, these findings demonstrate that while the JSB and JLA boundaries are highly conserved, the JLB and JSA junctions have undergone dynamic evolutionary shifts among these four *Pyrola* species.

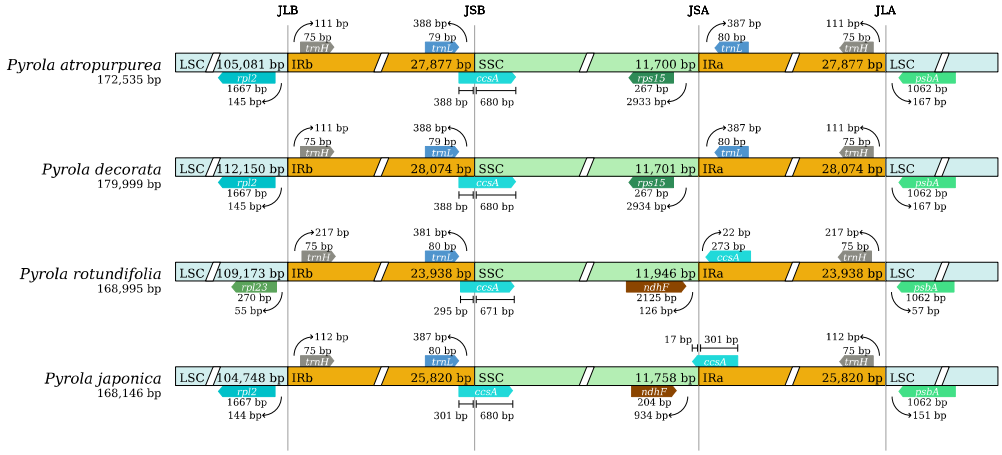

Figure 5: Comparative analysis of the IR and single-copy (LSC/SSC) boundaries across four *Pyrola* chloroplast genomes. The four structural transition zones are denoted as JLB (LSC/IRb), JSB (IRb/SSC), JSA (SSC/IRa), and JLA (IRa/LSC).

3.5 Comparative Sequence Divergence and Nucleotide Diversity among Pyrola Species

To analyze the genomic divergence within the genus *Pyrola*, the complete chloroplast genomes of *P. japonica*, *P. atropurpurea*, and *P. rotundifolia* were aligned and compared using the mVISTA program, with *P. decorata* serving as the reference (Fig. 6). The alignment demonstrated a high degree of synteny and overall sequence similarity across the four *Pyrola* species, reflecting strong evolutionary conservation. As expected, the protein-coding regions (exons, shaded in blue) exhibited significantly higher sequence identity than the

non-coding regions (intergenic spacers and introns, shaded in pink). Notable hypervariable non-coding regions included the intergenic regions between *matK-psbK* and *rps16-accD*.

Next, we calculated the Pi value across the ortholog genes of the four *Pyrola* species to identify regions of high sequence variation (Fig. 7). The total length of the aligned sequences was 48,895 bp, encompassing a total of 703 polymorphic sites. The Pi values ranged from 0 to 0.1479, with an overall average of 0.0075. Regionally, the average Pi values for the LSC, SSC, and IR regions were 0.0082, 0.0138, and 0.0056, respectively, indicating that nucleotide polymorphism in the single-copy regions is considerably higher than in the IR regions. At the individual gene level, the *trnG-GCC* gene, located in the LSC region, exhibited the highest nucleotide diversity (0.1479). This was followed by *trnG-UCC*, *rrn16*, *rps8*, *rps11*, *ccsA*, and *cemA*, which yielded Pi values of 0.0282, 0.0224, 0.0213, 0.0168, 0.0166, and 0.0152, respectively. These highly variable loci represent mutational hotspots.

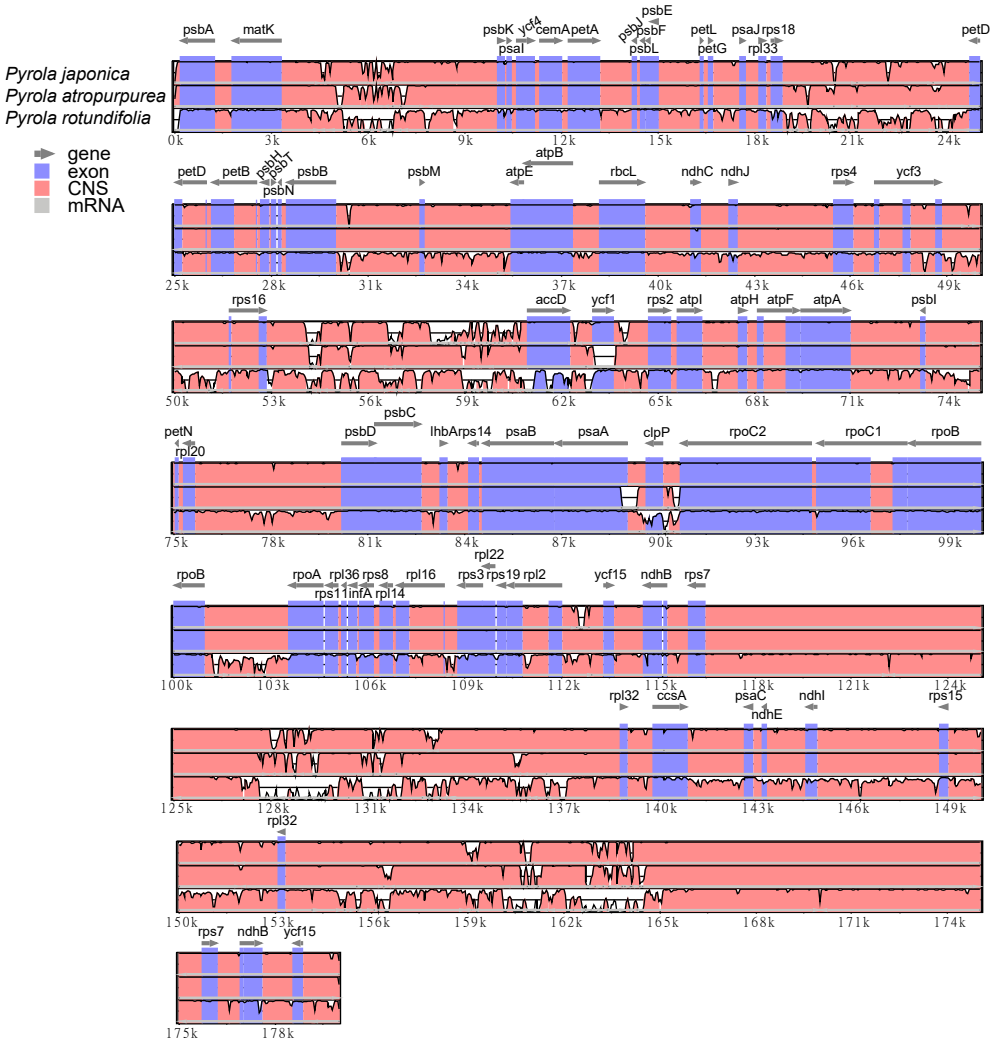

Figure 6: A sequence similarity plot was generated using the mVISTA tool to compare the four chloroplast genomes of *Pyrola* species. *Pyrola decorata* was used as a reference genome. The sequence similarities of *P. japonica*, *P. atropurpurea*, and *P. rotundifolia* relative to *P. decorata* are represented on the *Y*-axis, showing the percentage of sequence identity ranging from 50% to 100%. Gray arrows denote the location and transcriptional orientation of annotated genes. Non-coding sequences (CNS) and coding exons are highlighted in pink and purple, respectively.

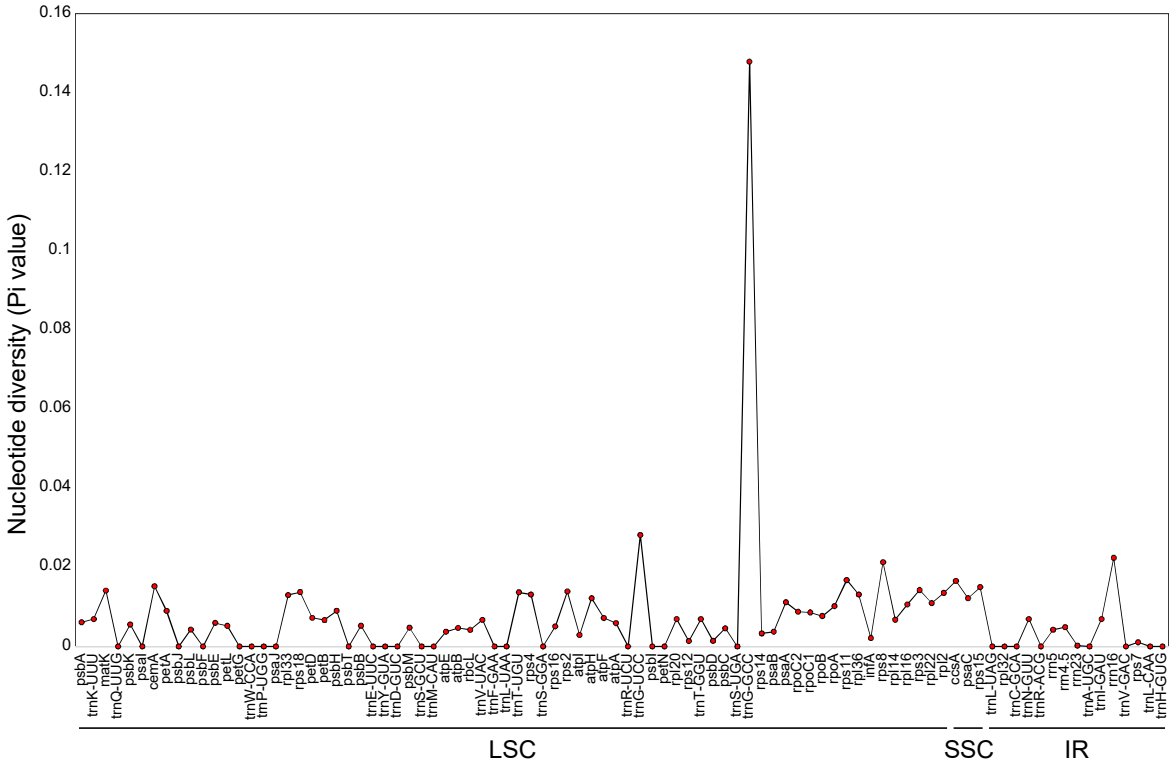

Figure 7: Nucleotide variability across genes in four *Pyrola* chloroplast genomes. The *X*-axis denotes ortholog genes in four *Pyrola* chloroplast genomes and the *Y*-axis represents nucleotide diversity (Pi) for each gene.

3.6 Phylogenetic Inferences of Ericaceae Family Based on Chloroplast Genome Comparative Analysis

To resolve the phylogenetic position of *P. japonica*, its complete chloroplast genome sequence was aligned with those of 40 closely related species from the family Ericaceae. A phylogenetic tree was subsequently constructed using the ML method, with *Capsicum chacoense* and *Capsicum galapagoense* (Solanaceae) designated as outgroups (Fig. 8). The resulting ML tree exhibited robust statistical support across most nodes, with bootstrap values exceeding 96% except for two nodes, effectively resolving the analyzed taxa into four distinct clades. All four *Pyrola* species clustered together into a strongly supported, distinct monophyletic clade (100% bootstrap value). Within this *Pyrola* lineage, *P. japonica* occupies a sister position to *P. decorata*. This close phylogenetic relationship is entirely consistent with our whole-genome comparative sequence analysis, which demonstrated that *P. japonica* shares the highest sequence similarity with *P. decorata* among the analyzed species (Fig. 6).

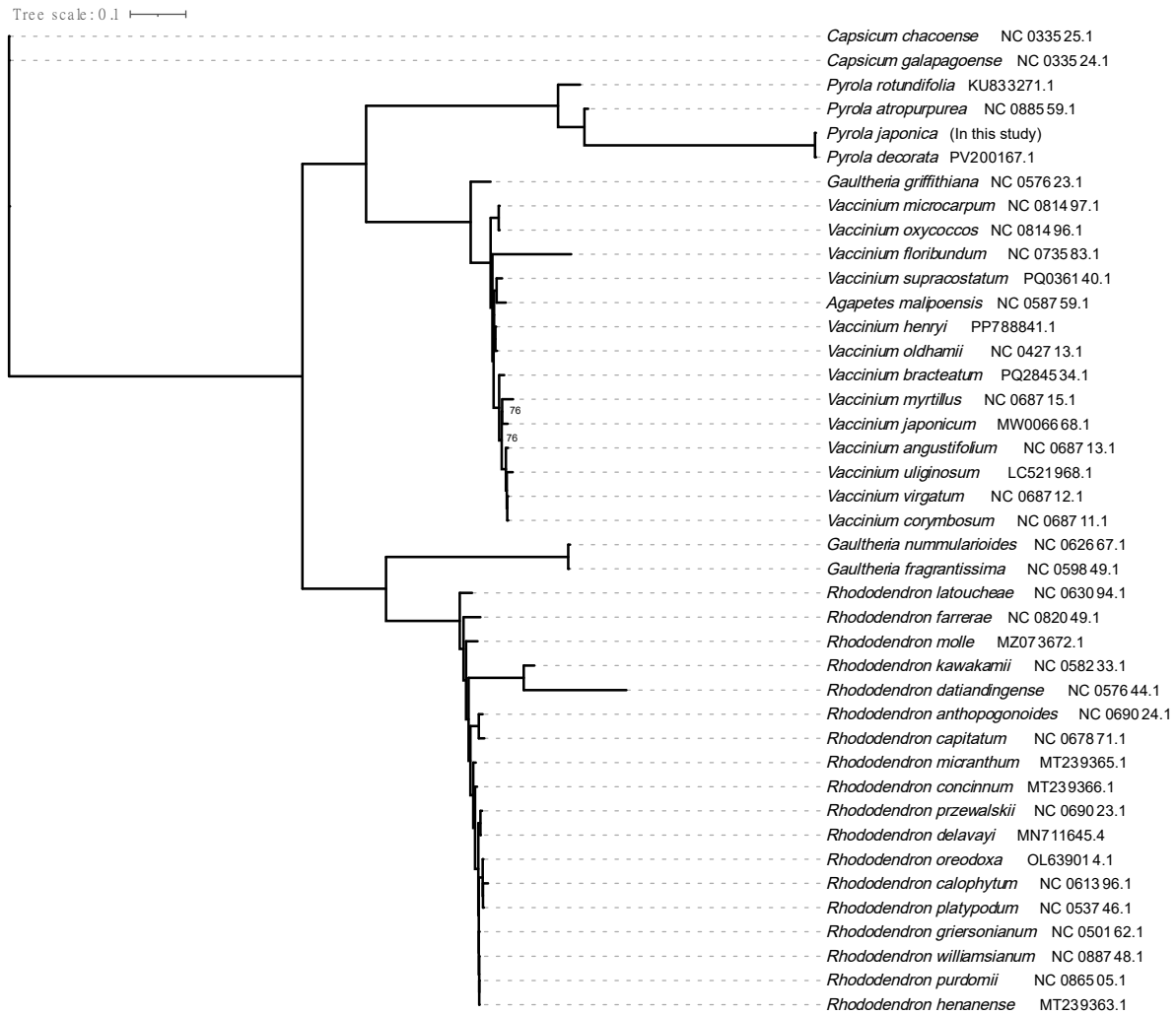

Figure 8: Maximum-likelihood phylogenetic tree inferred from 41 complete chloroplast genome sequences. The tree was rooted using *Capsicum chacoense* and *Capsicum galapagoense* as outgroups. The tree exhibits robust statistical support, with bootstrap values exceeding 96% across all nodes except for two (76%), which are specifically indicated. The scale bar indicates the number of nucleotide substitutions per site.

3.7 Relative Evolutionary Divergence Analysis of the Ericaceae Family Based on Chloroplast Genome Comparisons

To trace the evolutionary history of *P. japonica* and its relatives, we performed a Bayesian divergence time estimation using the complete chloroplast genome alignments (Fig. 9). Interestingly, the topology of the resulting Bayesian chronogram exhibited notable differences from the ML phylogeny. In the time-calibrated phylogeny, the *Pyrola* lineage was resolved as the earliest diverging clade among the examined Ericaceae, clearly separating from the rest of the family (Fig. 9). This isolated phylogenetic position, coupled with distinctively long branch lengths, indicates that the plastid genomes of *Pyrola* separated from the other Ericaceae taxa in deep evolutionary time. Within this early-diverging clade, *P. japonica* emerged as the sister species of *P. decorata*, representing a highly recent speciation event characterized by short terminal branch lengths (Fig. 9). Collectively, these temporal and topological insights suggest that the genus *Pyrola* possesses a unique evolutionary trajectory compared to the other examined Ericaceae taxa.

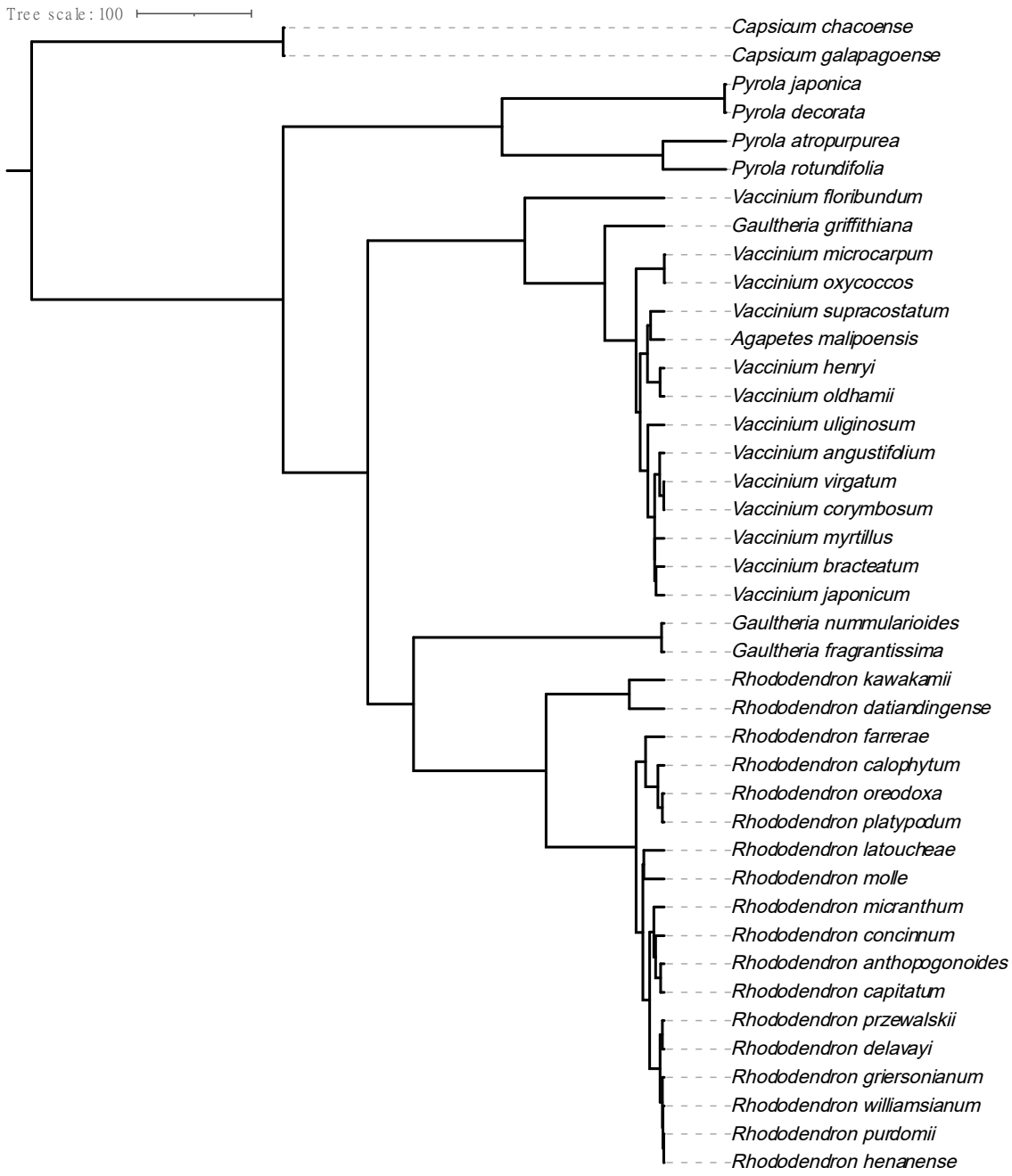

Figure 9: Evolutionary relationships inferred from 41 complete chloroplast genomes. The phylogenetic tree was constructed using a molecular-clock model implemented in BEAST. *Capsicum galapagoense* and *Capsicum chacoense* were used as outgroups. The horizontal axis represents relative evolutionary divergence inferred from branch lengths.

4 Discussion

Chloroplast genomes are essential tools in plant science. They are often used as DNA barcodes to identify different species and study how they are classified [1,11]. In this study, we successfully sequenced and analyzed the complete chloroplast genome of *P. japonica*. Our results showed that the *P. japonica* cp genome is 168,146 bp long with a GC content of 35.1% (Fig. 1). The high degree of similarity in GC content (ranging from 34.8% to 35.7%) across the four analyzed plastomes strongly underscores the conserved

nature of the chloroplast genome within the genus *Pyrola*. Moreover, slight fluctuations in overall genome sizes, ranging from 168,146 bp in *P. japonica* to 179,999 bp in *P. decorata*, were detected (Table S4). This consistency makes cpDNA a reliable resource for identifying species and understanding the evolutionary relationships within the *Pyrola* lineage.

Codons are essential for translating genetic information into functional proteins. Codon usage bias, which is the tendency of an organism to prefer specific codons, is a key indicator of genome evolution [32]. In this study, we used RSCU values from 75 CDS to measure this preference (Fig. 2). Our analysis showed that Leucine was the most frequent amino acid, while Cysteine was the least frequent. This is consistent with other species in the genus; for example, in *P. atropurpurea*, Leucine represents the highest proportion (10.4%), whereas Cysteine, the least abundant amino acid, accounts for only 1.1% [11]. Moreover, our study showed that the highest RSCU value was found in the UUA codon (which codes for Leucine). This high value points to a strong AT bias at the third codon position, which is a common characteristic of chloroplast DNA [1]. We found a similar trend in *P. decorata*, which shows a very high A+T content of 92.1% at the third codon position [1]. These results support the evolutionary theory that closely related species share very similar codon usage patterns [33].

Repetitive sequences are essential genetic markers that provide deep insights into the origin and evolution of a species [34]. In this study, we identified 789 long dispersed repeat sequences within the *P. japonica* chloroplast genome (Fig. 3). Additionally, we detected 112 SSR loci (Fig. 4). SSR markers are highly valued in genetic research and polymorphisms investigations [35]. Our analysis showed that mononucleotide and dinucleotide repeats were predominantly composed of A/T motifs. This finding supports the point that polyA and polyT repeats are the most common SSRs in chloroplasts, while G or C tandem repeats are relatively rare [35]. These SSRs represent promising molecular tools for future research on the genetic diversity and population structure of *P. japonica*.

The expansion and contraction of the IR regions are primary drivers of genome size variation in chloroplast [36]. IR regions of *P. japonica* (25,820 bp) exhibited an intermediate length compared to the larger IRs of *P. decorata* (28,074 bp) and *P. atropurpurea* (27,877 bp), and the significantly smaller IR of *P. rotundifolia* (23,938 bp) (Fig. 5). These results highlight the dynamic nature of plastome boundaries in this lineage. Consequently, these structural variations provide robust molecular evidence that can be utilized to refine the taxonomic classification of *Pyrola* species [1]. Moreover, our comparison of four *Pyrola* species revealed that while the JSB (IRb-SSC) and JLA (IRa-LSC) junctions are highly stable, the JLB (LSC-IRb) and JSA (SSC-IRa) boundaries are hotspots for evolutionary shifts. The strict conservation of the *ccsA* gene expansion (680 bp) across most species (*P. japonica*, *P. atropurpurea*, and *P. decorata*) suggests a functional constraint at the JSB junction.

The mVISTA alignment confirmed a high degree of synteny and overall sequence similarity across the four *Pyrola* species, demonstrating a strong evolutionary conservation (Fig. 6). Moreover, protein-coding regions are more conserved than non-coding regions, which is typical for chloroplast genomes due to selective pressure on functional genes. We also could identify hypervariable non-coding regions, including *matK-psbK* and *rps16-accD*. Our analysis of nucleotide diversity further identified several mutational hotspots. The overall average Pi value was low (0.0075); however, nucleotide polymorphism was notably higher in the single-copy regions (LSC and SSC) than in the IR regions. This pattern reflects the characteristic stabilizing effect of IR on the plastid genome [37]. The *trnG-GCC* gene was identified as the most divergent locus (Pi = 0.1479), followed by a suite of highly variable genes including *trnG-UCC*, *rrn16*, *rps8*, *rps11*, *ccsA*, and *cemA*. These hotspots offer superior potential as DNA barcodes for *Pyrola*. Unlike universal

barcodes, these lineage-specific markers can provide the high resolution needed to distinguish between morphologically similar species or to study the population genetics of these medicinal herbs [11].

The phylogenetic analysis provides robust evidence for the taxonomic positioning of *P. japonica* within the Ericaceae family. The clear resolution of the four *Pyrola* species into distinct clades underscores the reliability of using complete plastome sequences for resolving complex relationships within this family (Fig. 8). Our findings confirm that all *Pyrola* species form a strongly supported monophyletic group, reinforcing the evolutionary integrity of the genus. Furthermore, our relative evolutionary divergence analysis revealed that the *Pyrola* lineage diverged earlier than the other examined Ericaceae taxa, branching off in deep evolutionary time (Fig. 9). This ancient divergence scenario is consistent with previous multi-locus studies utilizing *rbcL* and *matK* sequences, which similarly positioned *Pyrola* as an early-diverging lineage within the broader Ericaceae [38]. Evolutionarily, *Pyrola* species exhibit a mixotrophic life cycle [2]. Because they supplement their carbon intake via fungi, their plastomes experience relaxed purifying selection, leading to the gradual degeneration of photosynthesis-related pathways as they shift toward heterotrophy [39]. Within Ericaceae, the evolutionary transition to acquire carbon from fungi is a specialized trait predominantly reported in specific lineages, notably the subfamilies Monotropoideae and Pyroloideae [40]. Both our ML phylogeny and Bayesian chronogram firmly place *P. japonica* as the sister species of *P. decorata* (Fig. 8; Fig. 9). Given that *P. decorata* is already documented to be experiencing active genomic degradation within its photosynthetic apparatus [39], this close phylogenetic relationship suggests that *P. japonica* is subject to similar evolutionary pressures driving the reduction of its photosynthetic capacity. Indeed, our annotation of the *P. japonica* plastome revealed widespread truncation and pseudogenization of the *ndh* gene family (Table S2), providing genomic evidence for the degradation of photosynthesis-related genes within *P. japonica*. While the *Pyrola* species formed a robust monophyletic clade, our phylogenetic tree also revealed topological complexities within the broader Ericaceae, particularly among the other examined lineages. For instance, we observed the paraphyletic nature of the *Vaccinium* group, with *Agapetes* and *Gaultheria* species positioned within the *Vaccinium* clade (Fig. 8). This clustering is consistent with previous molecular phylogenetic studies indicating that the large genus *Vaccinium* is not strictly monophyletic [41–43].

Beyond these broad phylogenetic implications, certain limitations remain to be addressed in future research. First, our findings rely primarily on computational and bioinformatic analyses. Although the widespread pseudogenization of photosynthesis-related genes provides genomic evidence for mixotrophy, exploring the *in vivo* functional implications of these degraded pathways requires further experimental validations. Furthermore, while cpDNA is an excellent tool for lineage tracking due to its maternal inheritance, its well-known limitation is a low nucleotide substitution rate relative to the nuclear genome. Because cpDNA is highly conserved, relying solely on plastome data may occasionally lack the resolution required to distinguish extremely closely related varieties or recently diverged populations [44]. To fully understand the evolutionary biology, functional adaptations, and fine-scale population genetics of *P. japonica*, future research must integrate these chloroplast resources with nuclear genomic data, experimental functional assays, and broader ecological studies.

5 Conclusion

This study successfully characterized the complete chloroplast genome of *P. japonica*, which comprises 168,146 bp and displays a typical quadripartite structure. The genome contained 136 genes (65 protein-coding, 45 tRNA, 8 rRNA, and 18 pseudogenes) with an overall GC content of 35.1%. Genomic analysis revealed a distribution of 789 long dispersed repeats and 112 SSRs, primarily located within the LSC, IR, and SSC

regions. Phylogenetic reconstruction showed that *P. japonica* is a sister group of *P. decorata* within the *Pyrola* lineage. Furthermore, our relative evolutionary divergence analysis revealed that *Pyrola* diverged earlier than the other examined Ericaceae taxa, representing an ancient split within the family. These findings provide essential genomic resources that will facilitate future research in species identification, population genetics, and the molecular evolution of mixotrophic lineages within the family Ericaceae.

Acknowledgement: Not applicable.

Funding Statement: This was supported by Korea National University of Transportation Industry-Academy Cooperation Foundation in 2024.

Author Contributions: CRediT: Chayanee Chairattanawat: Conceptualization, Writing—original draft, review & editing; Junghwa Kang: Investigation, Resources; Jaewook Kim: Conceptualization, Formal analysis, Supervision, Writing—review & editing; Bae Young Choi: Conceptualization, Formal analysis, Funding acquisition, Supervision, Writing—original draft, review & editing. All authors reviewed and approved the final version of the manuscript.

Availability of Data and Materials: The complete chloroplast genome sequence of *P. japonica* in this study was submitted to the NCBI database under the accession number PZ137749.

Ethics Approval: Not applicable.

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

Supplementary Materials: The supplementary material is available online at https://www.techscience.com/doi/10.32604/phyton.2026.084482/s1.

References

1. Kang R, Kang S, Yu K, Jiang Y, Qin Z, Hu Y, et al. Exploring the complete chloroplast genome of *Pyrola decorata* andres: structure, variability, phylogenetic relationship. Curr Issues Mol Biol. 2025;47(9):688. [CrossRef].
2. Uesugi T, Nakano M, Selosse MA, Obase K, Matsuda Y. *Pyrola japonica*, a partially mycoheterotrophic Ericaceae, has mycorrhizal preference for russulacean fungi in central Japan. Mycorrhiza. 2016;26(8):819–29. [CrossRef].
3. Huh MK, Chung SD, Huh HW. Allozyme variation and population structure of *Pyrola japonica* in Korea. Bot Bull Acad Sin. 1998;39:107–12.
4. Kim JS, Shim SH, Xu YN, Kang SS, Son KH, Chang HW, et al. Phenolic glycosides from *Pyrola japonica*. Chem Pharm Bull. 2004;52(6):714–7. [CrossRef].
5. Jang JH, Sim MO, Lee KH, Woo KW, Jung HK, An B, et al. Effect of antioxidant activity of *Pyrola japonica* on H_2O_2-induced oxidative stress in HSEC cells. Planta Med. 2016;81(S 01):S1–S381. [CrossRef].
6. Park JS, Lim HH. Effects of *Pyrola japonica* extracts on osteoclast differentiation and bone resorption. J Korean Med Rehabil. 2019;29(2):135–47. [CrossRef].
7. Park HG, Cha MR, Hwang JH, Kim JY, Park MS, Choi SU, et al. Antimicrobial activity of the extract from *Pyrola japonica* against *Bacillus subtilis*. J Life Sci. 2006;16(6):989–93. [CrossRef].
8. Matsuda Y, Shimizu S, Mori M, Ito S, Selosse MA. Seasonal and environmental changes of mycorrhizal associations and heterotrophy levels in mixotrophic *Pyrola japonica* (Ericaceae) growing under different light environments. Am J Bot. 2012;99(7):1177–88. [CrossRef].
9. Jin L, Qu S, Zhu W, Li Q, Tan Y, Cao Z, et al. Chloroplast-genome-based insights into the systematic relationship of *Eriobotrya* (*Rhaphiolepis*). iScience. 2025;28(11):113855. [CrossRef].
10. Go AR, Takahashi Y, Iwasaki T, Yoo KO. The complete chloroplast genome sequence and phylogenetic analysis of *Viola grayi* (Violaceae). Mitochondrial DNA B Resour. 2025;10(12):1195–9. [CrossRef].
11. Sheng W. Chloroplast genome sequence characterization and phylogenetic analysis of Pyrola atropurpurea franch. Phyton. 2025;94(2):331–45. [CrossRef].

12. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30(15):2114–20. [CrossRef].
13. Jin JJ, Yu WB, Yang JB, Song Y, DePamphilis CW, Yi TS, et al. GetOrganelle: a fast and versatile toolkit for accurate *de novo* assembly of organelle genomes. Genome Biol. 2020;21(1):241. [CrossRef].
14. Wick RR, Schultz MB, Zobel J, Holt KE. Bandage: interactive visualization of *de novo* genome assemblies. Bioinformatics. 2015;31(20):3350–2. [CrossRef].
15. Tillich M, Lehwark P, Pellizzer T, Ulbricht-Jones ES, Fischer A, Bock R, et al. GeSeq–versatile and accurate annotation of organelle genomes. Nucleic Acids Res. 2017;45(W1):W6–W11. [CrossRef].
16. Okonechnikov K, Golosova O, Fursov M, UGENE team. Unipro UGENE: a unified bioinformatics toolkit. Bioinformatics. 2012;28(8):1166–7. [CrossRef].
17. 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(W1):W59–64. [CrossRef].
18. Kurtz S. The Vmatch large scale sequence analysis software. Ref Type Comput Program. 2003;412:297.
19. Beier S, Thiel T, Münch T, Scholz U, Mascher M. MISA-web: a web server for microsatellite prediction. Bioinformatics. 2017;33(16):2583–5. [CrossRef].
20. Maździarz M, Zając S, Paukszto Ł, Sawicki J. RSCUcaller: an R package for analyzing differences in relative synonymous codon usage (RSCU). BMC Bioinform. 2025;26(1):141. [CrossRef].
21. Li H, Guo Q, Xu L, Gao H, Liu L, Zhou X. CPJSdraw: analysis and visualization of junction sites of chloroplast genomes. PeerJ. 2023;11:e15326. [CrossRef].
22. Brudno M, Do CB, Cooper GM, Kim MF, Davydov E, NISC Comparative Sequencing Program, et al. LAGAN and Multi-LAGAN: efficient tools for large-scale multiple alignment of genomic DNA. Genome Res. 2003;13(4):721–31. [CrossRef].
23. Librado P, Rozas J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics. 2009;25(11):1451–2. [CrossRef].
24. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013;30(4):772–80. [CrossRef].
25. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, et al. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. Mol Biol Evol. 2020;37(5):1530–4. [CrossRef].
26. Letunic I, Bork P. Interactive Tree of Life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool. Nucleic Acids Res. 2024;52(W1):W78–82. [CrossRef].
27. Sebastin R, Kim J, Jo IH, Yu JK, Jang W, Han S, et al. Comparative chloroplast genome analyses of cultivated and wild *Capsicum* species shed light on evolution and phylogeny. BMC Plant Biol. 2024;24(1):797. [CrossRef].
28. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics. 2009;25(15):1972–3. [CrossRef].
29. Baele G, Ji X, Hassler GW, McCrone JT, Shao Y, Zhang Z, et al. BEAST X for Bayesian phylogenetic, phylogeographic and phylodynamic inference. Nat Methods. 2025;22(8):1653–6. [CrossRef].
30. Yule GU. II.—A mathematical theory of evolution, based on the conclusions of Dr. J. C. Willis, F. R. S. Philos Trans R Soc Lond. 1925;213(402–410):21–87. [CrossRef].
31. Hasegawa M, Kishino H, Yano T. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. J Mol Evol. 1985;22(2):160–74. [CrossRef].
32. Zhang Y, Ma Y, Yu H, Han Y, Yu T. Deciphering codon usage patterns in the mitochondrial genome of the *Oryza* species. Agronomy. 2024;14(11):2722. [CrossRef].
33. Parvathy ST, Udayasuriyan V, Bhadana V. Codon usage bias. Mol Biol Rep. 2022;49(1):539–65. [CrossRef].
34. Yang C, Wang K, Zhang H, Guan Q, Shen J. Analysis of the chloroplast genome and phylogenetic evolution of three species of *Syringa*. Mol Biol Rep. 2023;50(1):665–77. [CrossRef].
35. Lei W, Ni D, Wang Y, Shao J, Wang X, Yang D, et al. Intraspecific and heteroplasmic variations, gene losses and inversions in the chloroplast genome of *Astragalus membranaceus*. Sci Rep. 2016;6:21669. [CrossRef].
36. Hejazi FA, Mohammadi P, Soorni A. Comparative chloroplast genomics of *Teucrium* species reveals genome evolution, phylogenetic relationships, and candidate molecular markers. Sci Rep. 2025;15(1):44318. [CrossRef].

37. Wang ZX, Wang DJ, Yi TS. Does IR-loss promote plastome structural variation and sequence evolution? Front Plant Sci. 2022;13:888049. [CrossRef].
38. Schwery O, Onstein RE, Bouchenak-Khelladi Y, Xing Y, Carter RJ, Linder HP. As old as the mountains: the radiations of the Ericaceae. New Phytol. 2015;207(2):355–67. [CrossRef].
39. Yu J, Wang C, Gong X. Degeneration of photosynthetic capacity in mixotrophic plants, *Chimaphila japonica* and *Pyrola decorata* (Ericaceae). Plant Divers. 2017;39(2):80–8. [CrossRef].
40. Lallemand F, Gaudeul M, Lambourdière J, Matsuda Y, Hashimoto Y, Selosse MA. The elusive predisposition to mycoheterotrophy in Ericaceae. New Phytol. 2016;212(2):314–9. [CrossRef].
41. Kron KA, Powell EA, Luteyn JL. Phylogenetic relationships within the blueberry tribe (Vaccinieae, Ericaceae) based on sequence data from MATK and nuclear ribosomal ITS regions, with comments on the placement of *Satyria.* Am J Bot. 2002;89(2):327–36. [CrossRef].
42. Becker AL, Crowl AA, Luteyn JL, Chanderbali AS, Judd WS, Manos PS, et al. A global blueberry phylogeny: evolution, diversification, and biogeography of Vaccinieae (Ericaceae). Mol Phylogenet Evol. 2024;201:108202. [CrossRef].
43. Zhidkin RR, Matveeva TV. Phylogeny problems of the genus *Vaccinium* L. and ways to solve them. Ecol Genet. 2022;20(2):151–64. [CrossRef].
44. Teske D, Peters A, Möllers A, Fischer M. Genomic profiling: the strengths and limitations of chloroplast genome-based plant variety authentication. J Agric Food Chem. 2020;68(49):14323–33. [CrossRef].