Li *et al. BMC Plant Biology* (2025) 25:176
https://doi.org/10.1186/s12870-025-06186-6

BMC Plant Biology

RESEARCH

Open Access

Comparative and phylogenetic analysis of *Potentilla* and *Dasiphora* (Rosaceae) based on plastid genome

Xiaoping Li[1,2], Hao Xu[1,3], Jingya Yu[1,3], Yun Han[1,3], Shuang Han[1,3], Yu Niu[1,3] and Faqi Zhang[1,3,4*]

Abstract

Background *Potentilla* L. and *Dasiphora* L. are predominantly perennial herbs, occasionally manifesting as annuals or shrubs, primarily found in the northern temperate zone. However, taxonomic classification within this group remains contentious, particularly regarding genus boundaries and species delineation. Therefore, this study sequenced and analyzed the complete plastid genomes of 19 species from *Potentilla* and *Dasiphora*, comparing them with five previously published plastid sequences. Our objectives included reconstructing phylogenetic relationships within *Potentilla* and *Dasiphora* and investigating cytonuclear discordance among them.

Results These plastid genomes were highly conserved in structure, GC content, and overall genome composition, comprising 84 protein-coding genes, 37 tRNA genes, and 8 rRNA genes. Notably, all *Dasiphora* plastid genomes lacked the unique intron for *rpl2*. Comparative genomic analyses revealed that variations in plastid genome size were due to differences in the lengths of the LSC, SSC, and IR regions. The IR region was predominantly conserved, while non-coding regions exhibited higher variability than coding regions. We screened SSR and identified seven highly variable loci that serve as potential molecular markers, offering valuable insights into the intergeneric relationships between *Potentilla* and *Dasiphora*. Phylogenetic analyses based on nuclear (ITS, ETS) and cytoplasmic (plastid, mitochondrial) genes confirmed the monophyly of *Potentilla* and *Dasiphora*, with results largely consistent with previous studies and supported by robust reliability metrics. We identified cytonuclear conflicts within *Potentilla*, which frequently disrupt its monophyly. We hypothesize that these conflicts may result from interspecific hybridization or incomplete lineage sorting events during the evolutionary history of the genus.

Conclusions This study offers a theoretical foundation for advancing molecular identification and phylogenetic research on *Potentilla* and *Dasiphora* species. However, future work could benefit from greater detail on the criteria for selecting mitochondrial gene sequences and nrDNA datasets.

Keywords *Potentilla*, *Dasiphora*, Plastid genome, Comparative genomics, Phylogenetic analysis, Cytonuclear discordance

*Correspondence:
Faqi Zhang
fqzhang@nwipb.cas.cn
[1]Key Laboratory of Adaptation and Evolution of Plateau Biota, Northwest Institute of Plateau Biology, Chinese Academy of Sciences, Xining 810001, China
[2]Academy of Animal Science and Veterinary, Qinghai University, Xining 810016, China
[3]College of Life Sciences, University of Chinese Academy of Sciences, Beijing 100039, China
[4]Xining Botanical Garden, Xining 810001, China

Background

Potentilla L. and *Dasiphora* L. are both part of the Rosaceae Juss. family. There are approximately 485 species in the *Potentilla*, primarily found in temperate and frigid zones of the Northern Hemisphere, often in alpine or arid regions [1–3]. Both *Potentilla* and *Dasiphora* are widely used as medicinal plants. Initially, *Dasiphora* was classified within the genus *Potentilla*; however, *Dasiphora* was initially classified within the *Potentilla* but was later distinguished as its genus, separate from *Potentilla*, based on morphological differences [4]. In recent years, DNA molecular markers have enabled more objective assessments of the boundaries within the tribe Potentilleae, including the separation of *Potentilla* and *Dasiphora* [1–3, 5–10]. Phylogenetic analyses based on molecular evidence have demonstrated that *Potentilla* and *Dasiphora* occupy distinct branches on the phylogenetic tree, indicating significant genetic differentiation between them. This supports the view that they are monophyletic relative to each other [1, 11, 12].

The plastid genome typically has a quadripartite structure consisting of a large single-copy (LSC) region, a small single-copy (SSC) region, and two inverted repeats (IRa and IRb) within an independent circular genome. The plastid genome size ranges from approximately 120 to 170 kilobase pairs (kb) and usually encodes 120 to 130 genes [13–15]. Size variations in the plastid genome are mainly due to changes in the IR regions, including expansion, contraction, or even loss [16]. The plastid genome composition in angiosperms is highly conserved, usually uniparentally inherited, and most often matrilineally transmitted, with low nucleotide substitution rates [17–20]. Consequently, the plastid genome provides rich genetic information for phylogenetic studies and is crucial for resolving plant phylogenetic relationships [21, 22].

In recent years, the continuous development of next-generation sequencing (NGS) technologies has made obtaining plastid genome sequences easier and more cost-effective, enabling the resolution of various phylogenetic problems across different taxonomic levels [12, 23–25]. Low-coverage whole-genome sequencing (also known as genome skimming) can provide sequence information for the plastid genome, mitochondrial genome, and nuclear ribosomal regions (rDNA) [26]. This method is more straightforward and more efficient than RNA-seq and RAD-seq. It typically yields sequences for nuclear ribosomal internal transcribed spacers (nrITS), nuclear ribosomal external transcribed spacers (nrETS), various plastid markers, and polymorphic nuclear simple sequence repeats (nSSRs) [27, 28]. Moreover, this method has been utilized in phylogenetic studies at different taxonomic levels [29–32]. Increasingly, studies have revealed conflicting phylogenetic relationships when comparing nuclear and cytoplasmic gene data (chloroplast or mitochondrial), a phenomenon known as cytonuclear discordance [33]. Cytonuclear discordance can arise from various biological processes, including gene duplication, horizontal gene transfer, incomplete lineage sorting, or gene flow [34]. In *Potentilla*, researchers have identified nucleoplasmic conflict using the internal transcribed spacer (ITS) region of ribosomal DNA and the *trn*L-F intergenic spacer of chloroplast DNA, confirming the monophyly of the genus [35]. However, comprehensive analyses of cytonuclear conflicts between *Potentilla* and *Dasiphora* based on whole plastid genomes and nuclear ribosomal DNA (nrDNA) remain limited. Our study aims to address this gap, offering new insights into the taxonomic controversy between these genera.

In this study, we conducted de novo sequencing, assembly, and annotation of the plastid genomes from 11 species of *Potentilla* and 8 species of *Dasiphora*, and we performed comparative analyses of these genomes alongside other published *Potentilla* and *Dasiphora* species on a genome-wide scale. Our primary objectives were: (1) to enhance our understanding of the plastid genome structure in *Potentilla* and *Dasiphora*, shedding light on evolutionary patterns; (2) to conduct a comprehensive comparison of chloroplast genomes to identify structural variations among these taxa; (3) to develop and screen suitable barcodes for *Potentilla* and *Dasiphora*; and (4) to acquire plastome data, as well as nuclear ribosomal DNA (nrDNA, including nrITS and nrETS) and mitochondrial gene data, aiming to elucidate the phylogenetic relationships between *Potentilla* and *Dasiphora*.

Materials and methods

Plant materials, DNA extraction and quality control and sequencing

Field-collected species were archived at the Qinghai-Tibetan Plateau Museum of Biology (QTPMB) (Table S2). Immediately upon collection, fresh leaves were desiccated with silica gel to facilitate DNA extraction. The end-repair method, using 150 bp fragments, was employed for cBOT clustering and subsequent sequencing on the NovaSeq 6000 system. Sequencing data underwent purification with fastp v.0.23.1 [36], targeting (1) reads with undefined nucleotides (N) comprising over 10% of their base count; (2) reads where more than half of the bases were of low quality ($Q \leq 5$); and (3) reads that included adapter sequences. Following these criteria, approximately 3.0 Gbp of high-quality, paired-end clean reads were generated per sample, laying the groundwork for further investigation.

Genome assembly, annotation, and genome structure analysis

In this study, we employed GetOrganelle with specific parameters (-k = 21, 45, 65, 85, 105, 121 -t = 128 -*R* = 15 -F = embplant_pt) to assemble the raw sequencing data derived from Illumina HiSeq [37], facilitating the construction of the circular plastid genome. Subsequently, the CPGAVAS2 [38] annotated its CDS, tRNA genes, and rRNA genes. The precision of start and stop codons [39] and gene locations were enhanced through manual adjustments anchored in the reference plastid genome sequence. Additionally, the tRNA genes underwent further verification via the online annotation resource tRNAscan-SE 2.0 [40], ensuring rigorous accuracy in our genetic annotations. Finally, the plastid genome was visualized and analyzed using Chloroplot [41].

In this research, we conducted a comprehensive comparative analysis of the 24 plastid genomes sequenced. Codon usage, including counts and RSCU in protein-coding genes, was assessed using CodonW v.1.4.4 [42]. SSRs were analyzed with MISA v2.1 [43], while long repeats (forward, palindromic, reverse, and complementary) were identified using the REPuter online tool under default settings [44]. The IRScope tool facilitated the visualization of contractions and expansions across LSC, IRb, SSC, and IRa regions [45]. Plastid genome alignments were performed using PhyloSuite v1.2.3, and nucleotide diversity (Pi) was calculated with DnaSP [46], employing a step size of 200 bp and a sliding window of 800 bp. Additionally, mVISTA was utilized for the comparative analysis and visualization of these genomes, enabling the detection of highly variable regions [47].

Acquisition of nrITS, nrETS, and mitochondrial gene sequences

We analyzed nrITS, nrETS, and mitochondrial gene sequences for 19 species. The nrITS sequences from *P. anserina*, *D. davurica*, *Sibbaldianthe bifurca*, and nrETS sequences from *D. fruticosa*, *D. glabra*, *P. lignosa*, alongside mitochondrial genome sequences from *Malus domestica* and *Rosa chinensis* (GenBank accession numbers U90788, FJ356159, KP994565 for nrITS; KP875279, KP875277 for nrETS; NC018554 NC065236 for mitochondrial genomes), served as references. Sequence alignment was performed using MAFFT v7.505 [48], followed by nrITS and nrETS sequences extraction using Easy353 v2.0.2 [49]. Mitochondrial gene sequences were extracted based on *cob*, *cox*1, and *mat*R gene reference sequences.

Positive selection analysis

We extracted non-redundant genes shared among species to assess selection pressure on plastid protein-coding genes and aligned them using MAFFT v7.505. The synonymous (Ks) and non-synonymous (Ka) substitution rates, along with the Ka/Ks ratios, were subsequently calculated using ParaAT 2.0 integrated with KaKs Calculator 2.0 [50], employing default parameters.

Phylogenetic analysis

In this study, we sequenced plastid genomes from 11 *Potentilla* and 8 *Dasiphora* species. We incorporated 4 *Potentilla*, 1 *Dasiphora*, and 4 outgroup plastid genomes from NCBI (*Ampelocera hottlei*, *Aphananthe aspera*, *Cannabis sativa*, and *Celtis tetrandra*). We performed phylogenetic analyses using both ML and BI on 8 distinct datasets, including plastid genome, nrDNA, LSC, SSC, IR, CDS, DNA barcoding (*rbc*L, *mat*K, and *trn*H-*psb*A), and mitochondrial gene sequences. The consistency of ITS, ETS, mitochondrial gene sequences, and other species-specific datasets was maintained, except where data were unavailable. Phylogenetic analyses were conducted using PhyloSuite v1.2.3 for comprehensive bioinformatics analysis and MAFFT v7.505 for initial multi-sequence alignment, employing the "auto" strategy to select optimal algorithms for different sequence types. Sequence alignments were refined using trimAl v.1.2 [51] to remove poorly aligned positions and divergent regions. For PCGs, batch refinement was performed using MACSE v2.0.3 [52]. The refined data were concatenated into eight matrices for subsequent analysis. Model selection and partitioning for each dataset were determined by ModelFinder [53]. ML analysis was carried out using IQ-TREE [54], with 1000 ultrafast bootstrap replicates to assess branch support, while BI was conducted using MrBayes v3.2.7a [55], with two independent Markov chain Monte Carlo (MCMC) runs of 10 million generations, sampling every 1,000 generations until the mean standard deviation of split frequencies dropped below 0.01. The first 25% of trees from each MCMC run were discarded as burn-in. Phylogenetic trees were visualized and edited using the online tool chiplot (https://www.chiplot.online/).

Results

Plastome characteristics

The plastid genome lengths of the 24 species ranged from 152,883 bp (*D. parvifolia* p158) to 156,461 bp (*P. potaninii* p5). All species sequenced in this study exhibited the typical quadripartite structure (Fig. 1), consisting of a large single-copy (LSC) region (84,117 – 85,808 bp), a small single-copy (SSC) region (18,128 – 18,709 bp), and two inverted repeats (IR) regions (25,291 – 26,276 bp). The GC content in the plastid genomes of *Potentilla* species ranged from 36.7% (*P. tanacetifolia* p99) to 37.3% (*D.* mandshurica p3). The GC content distribution among the four regions was uneven, with the IR regions having the highest GC content (42.60-42.92%), followed by the

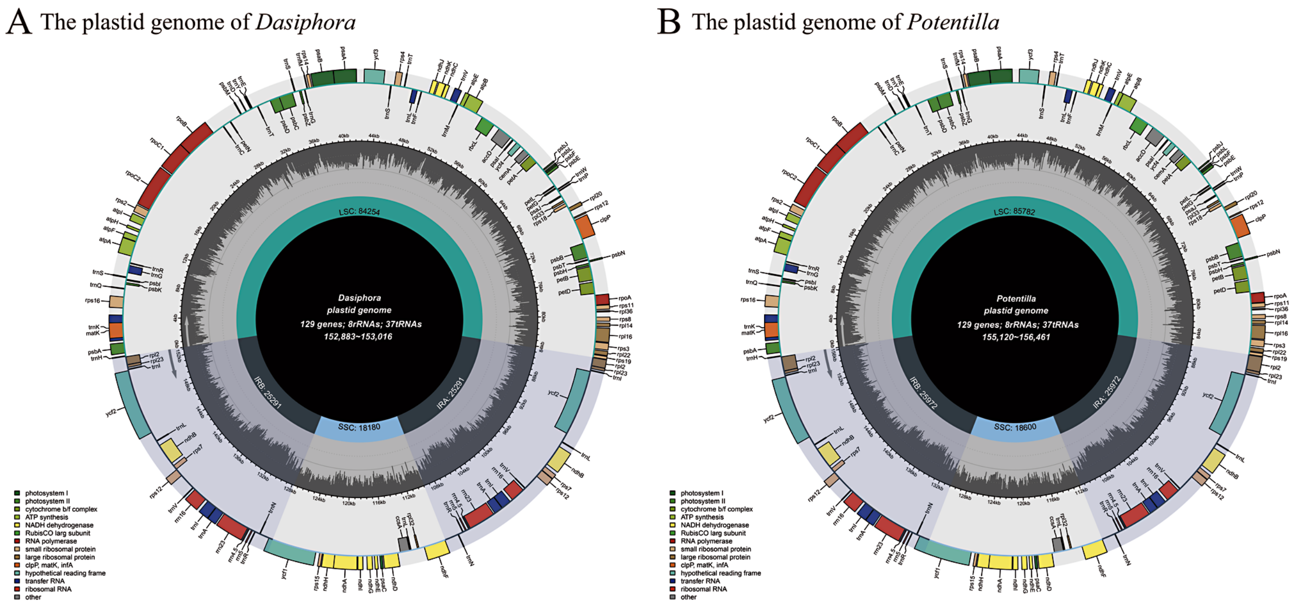

Fig. 1 Plastid genome map of *Dasiphora* (**A**) and *Potentilla* (**B**). Genes located inside the circle are transcribed clockwise, whereas those positioned outside are transcribed counterclockwise. Within the inner circle, darker gray represents the GC content, while lighter gray indicates the AT content of the genome

LSC region (34.54-35.17%) and the SSC region (30.55-31.08%) (Table 1). The higher GC content in the IR regions is likely related to the presence of all rRNA genes within these regions [56].

The genes encoded in the plastid genome of *Potentilla* are mainly composed of self-replication-related, photosynthesis-related, other, and *ycf* genes. Statistically, it was found that there were 129 coding genes in the plastid genome of *Potentilla* and *Dasiphora*, among which 18 genes existed in double copies, including all rRNA, tRNA genes *trnV-GAC*, *trnS-GCU*, *trnR-ACG*, *trnN-GUU*, *trnI-GAU*, *trnL-CAA*, *trnI-CAU*, *trnA-UGC*, and protein-coding genes *rps12*, *rps7*, *rpl2*, *rpl23*, ndhB, and *ycf2*. We found that 24 plastid genomes of commensal plants were consistent regarding the number (84, 37, 8) and type of CDSs, tRNAs, and rRNAs. 8 tRNA genes and 8 protein-coding genes contained a single intron. *clpP* and *ycf3* genes contained two introns (Table 1). By comparing the plastid genome sequences of the *Potentilla* obtained in this study, we found that they are highly conserved in terms of gene content, number of genes, orientation, and number of introns.

The boundaries of inverted repeat (IR) regions are hotspots for gene duplication or deletion [57], contributing to plastid genome size variation [16, 58, 59]. In this study, we analyzed the expansion and contraction of IR regions in 24 plastid genomes of the *Potentilla*. The results revealed a high degree of sequence similarity and structural conservation among these plastid genomes (Fig. 2). The size variation in *Potentilla* plastid genomes was attributed to differences in the lengths of the large single-copy (LSC), small single-copy (SSC), and IR regions. We further compared the exact IR boundary locations and neighbouring genes between the plastid genomes of *P. multicaulis* p93 and *P. tanacetifolia* p99. Notably, the *ndh*F genes in *P. multicaulis* p93 and *P. tanacetifolia* p99 were within the SSC/IRb boundary region. Additionally, the *rpl2* gene in the plastid genomes of all *Dasiphora* lacked one intron compared to *Potentilla*. In *Potentilla*, the *rpl2* gene is 61–75 bp from the LSC/IRb boundary, and the *ycf1* gene spans 1,076 – 1,350 bp within the IRa region. Conversely, in *Dasiphora*, the *rpl2* gene is 43 bp from the LSC/IRb boundary, and the *ycf1* gene spans 417-1,098 bp within the IRa region (Fig. 2).

Relative synonymous codon usage

Codon usage frequency and relative synonymous codon usage (RSCU) were calculated based on protein-coding genes from 24 plastid genomes of *Potentilla* and *Dasiphora*. A total of 64 codons encoding 20 amino acids were identified in this study. This includes six codons each for leucine (Leu), serine (Ser), and arginine (Arg); four codons each for proline (Pro), threonine (Thr), alanine (Ala), valine (Val), and glycine (Gly); and three codons for isoleucine (Ile). Methionine (Met) and tryptophan (Trp) are each encoded by a single codon, resulting in an RSCU value of 1 for these amino acids (Table S3). Leucine (Leu) exhibited the highest RSCU values (10.55-10.61%) among all protein-coding genes in the plastid genomes, whereas cysteine (Cys) had the lowest frequency (1.14-1.15%) (Fig. 3), consistent with previous reports on plastid

Table 1 List of genes in the plastid genome of the 24 *Potentilla* and *Dasiphora* species

Category of genes	Function of genes	Name of genes
rRNA	Self-replication	*rrn16*(2), *rrn23*(2), *rrn4.5*(2), *rrn5*(2)
tRNA		*trnA-UGC*(2), *trnC-GCA*, *trnD-GUC*, *trnE-UUC*, *trnF-GAA*, *trnfM-CAU*, *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*
Small subunit of ribosome		*rps2*(2), *3*, *4*, *7*(2), *8*, *11*, *12*(2), *14*, *15*, *16*, *18*, *19*
Large subunit of ribosome		*rpl2*(2), *14*, *16*, *20*, *22*, *23*(2), *32*, *33*, *36*
RNA polymerase		*rpoA*, *B*, *C1*, *C2*
NADH dehydrogenase	photosynthesis	*ndhA*, *B*(2), *C*, *D*, *E*, *F*, *G*, *H*, *I*, *K*, *J*
ATP synthase		*atpA*, *B*, *E*, *F*, *H*, *I*
Photosystem I gene		*psaA*, *B*, *C*, *I*, *J*
Photosystem II gene		*psbA*, *B*, *C*, *E*, *F*, *G*, *H*, *I*, *J*, *K*, *L*, *M*, *N*, *T*, *Z*
Cytochrome b/f complex		*petA*, *B*, *D*, *G*, *L*, *N*
RubisCO large subunit		*rbcL*
Other genes	Other genes	*accD*, *ccsA*, *cemA*, *clpP*, *matK*
Hypothetical chloroplast reading frames	Unknown	*Ycf1*, *2*(2), *3*, *4*

genomes [60]. Generally, codons with RSCU > 1 are preferred [42, 61]. Notably, nearly half of the codons analyzed had RSCU > 1 (30/64), and all these codons end in the bases A or U (Table S3).

Repeat sequences analysis

Simple sequence repeats (SSRs), which are short tandemly repeated DNA sequences composed of repetitive units 1–6 bp in length, are widely distributed throughout the plastid genome and commonly used as molecular markers for species identification [62–64]. We detected 52 (*D. mandshurica* MW752148) to 113 (*P. tanacetifolia* p99) SSRs in 24 plastid genomes (Fig. 4D). These included 39–89 mononucleotides, 6–21 dinucleotides, 0–5 trinucleotides, 2–9 tetranucleotides, and 0–1 pentanucleotide, with no hexanucleotides observed (Fig. 4A). Additionally, we found a specific base preference in the repeat sequence composition, with A/T being the only single-nucleotide SSR type across all 24 species. The repeat units of the other four SSR types also primarily consist of A or T (Fig. 4B). Using tandem repeat sequence detection technology and REPuter, we identified forward (F), reverse (R), palindromic (P), and complementary (C) repeat sequences in 24 plastid genomes of the *Potentilla*. We identified a total of 31 (*P. saundersiana* p131) to 48 (*P. saundersiana* NC072931) long repetitive elements, which included 12–24 forward repeats, 1–5 reverse repeats, 15–19 palindromic repeats, and 0–3 complementary repeats (Fig. 4C). These findings suggest that SSRs and long repetitive sequences vary among species, providing opportunities to develop new molecular markers for the identification of *Potentilla* and *Dasiphora* species.

Comparisons of the plastid genome

We compared the structure of 24 plastid genomes from *Potentilla* and *Dasiphora* sequenced in this study. Genomic divergence and sequence identity were assessed using mVISTA with *D. fruticosa* p1 as a reference. The results indicated that the LSC and SSC regions diverged more than the IR regions. Variation was primarily concentrated in the LSC and SSC regions, whereas the IR regions were less divergent. Additionally, coding regions were more conserved than non-coding regions. Significant differences in the non-coding regions appeared in the intergenic spacer (IGS) regions, including *rps16-trnQ-UUG*, *rpoB-trnC-GCA*, *trnT-GGU-psbD*, *rps15-ycf1*, *rpl32-trnL-UAG*, *rps4-trnT-UGU*, and *trnT-UGU-trnL-UAA* (Fig. 5). Only a few variations occurred in coding genes, with notable differences in genes such as *rpl22*, *ndhA*, and *psbB* (Fig. 5). In contrast, all rRNA genes were highly conserved.

Nucleotide diversity

To further understand DNA polymorphism (Pi) in the plastid genomes of *Potentilla* and *Dasiphora*, we analyzed sliding windows of these genomes and observed a high degree of variability. Nucleotide variation values were calculated for 89 coding genes and 56 IGS regions in the plastid genomes of 24 *Potentilla* and *Dasiphora* species. The IR regions were more conserved than the LSC and SSC regions, with almost all divergent regions located in the non-coding regions (Fig. 6). The Pi values for coding genes ranged from 0.00,000 to 0.09,077 (*ycf1*), with a mean value of 0.03. The Pi values for the IGS regions were significantly higher, with a maximum value of 0.15,419 (*rpl32-trnL-UAG*) and a mean value of 0.08, 2.67 times higher than the coding genes. We identified seven highly divergent regions (Pi > 0.12): rpl32-trnL-UAG (0.15,419), *rps4-trnT-UGU* (0.15,350), *trnH-GUG-psbA* (0.14,599), *rps16-trnQ-UUG* (0.14,465), *rps15-ycf1* (0.13,714), *trnT-UGU-trnL-UAA* (0.12,523), and *petA-psbJ* (0.12,014) (Fig. 6). These findings are consistent with the results of the mVISTA analysis. These hotspot regions can serve as

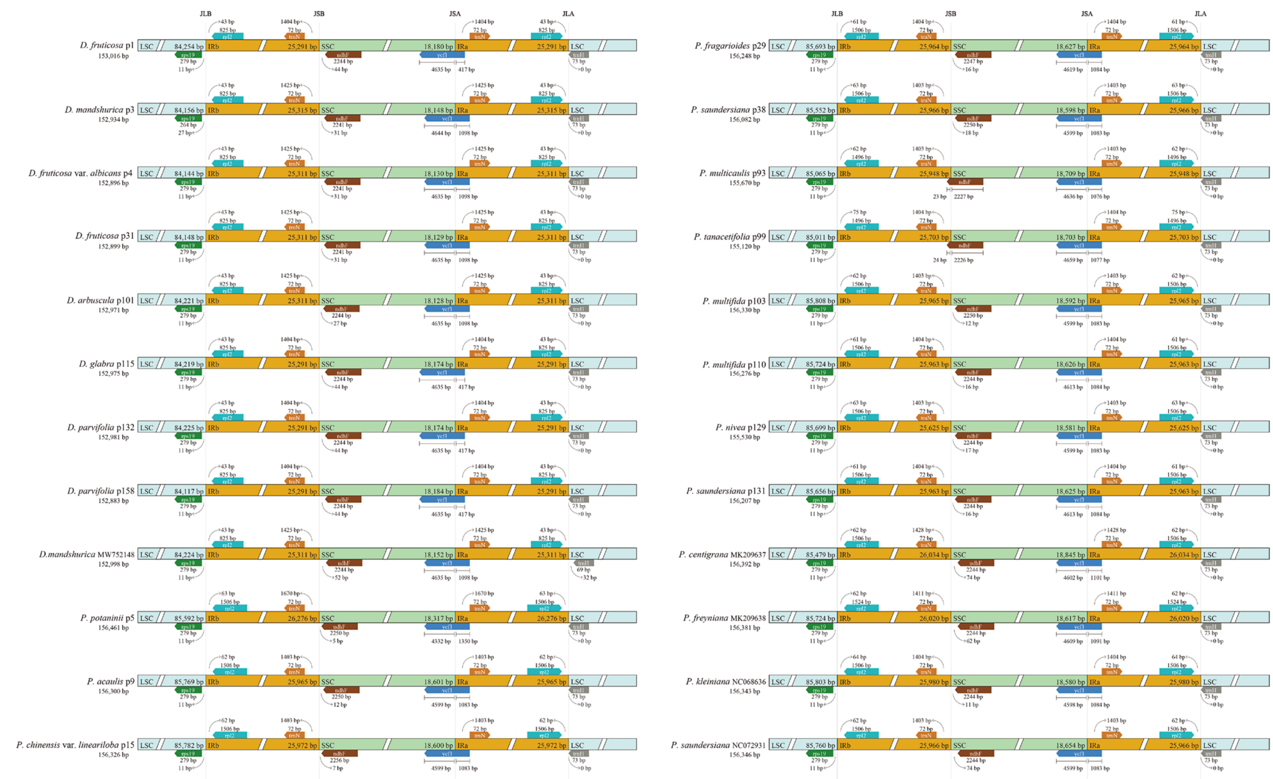

Fig. 2 Comparisons of LSC, SSC and IR region borders among 24 *Potentilla* and *Dasiphora* species plastid genomes. The numbers above indicate the distances between gene termini and regional boundaries, while those below represent their respective sizes. Arrows denote the locations of these distances. This figure is not drawn to scale

molecular markers for future phylogenetic analysis and species identification in *Potentilla* and *Dasiphora*.

Selective pressure analysis

We computed non-synonymous (Ka) to synonymous (Ks) substitution rates for 78 protein-coding genes across 24 plastid genomes of *Potentilla* and *Dasiphora*. A Ka/Ks ratio greater than 1 indicates positive selection, 1 indicates neutral selection, and a ratio less than 1 indicates purifying selection. Overall, the Ka/Ks ratios for most genes were below 1 (Fig. 7), indicating that the plastid genes of *Potentilla* and *Dasiphora* are conserved during evolution and primarily undergo purifying selection, consistent with their functional importance in plastids.

Phylogenetic analysis

In this study, we reconstructed the phylogenetic relationships of 24 *Potentilla* and *Dasiphora* species using maximum likelihood (ML) and Bayesian interference (BI) methods. Our analysis incorporated eight datasets, which included plastid genomes, nuclear ribosomal DNA (nrITS, nrETS), and mitochondrial genes, all processed using PhyloSuite v1.2.3 [24, 65].

In the analysis of the nrDNA data, discrepancies were noted between the topologies derived from the maximum likelihood (ML) and Bayesian interference (BI) analyses; however, the monophyly of *Dasiphora* was consistently supported with the highest confidence (expressed as BI posterior probability/ML bootstrap support: BIPP/MLBS = 1.00/100%) (Fig. S1, S2, S3). Notably, *P. multicaulis* 93 and *P. tanacetifolia* 99 consistently clustered together with very high support across all phylogenetic analyses, distinct from clustering within the *Potentilla* clade (Fig. S1, S2, S3).

In the analysis of mitochondrial gene data, both ML and BI analyses yielded nearly identical topologies. Specifically, *D. fruticosa* p31 formed a distinct clade, while the remaining *Dasiphora* species clustered together with robust support (BIPP/MLBS = 1.00/85%). *P. tanacetifolia* p99 was identified as the sister species to this clade with high support (BIPP/MLBS = 0.91/85%) (Fig. S4). Phylogenetic trees based on mitochondrial genes indicated that neither the *Potentilla* nor the *Dasiphora* are monophyletic, a finding strongly contrasting with the results from nrDNA and plastid genomes.

For the analysis of plastid genome data, the phylogenetic relationships inferred from the CDS (Fig. S5), LSC (Fig. S6), SSC (Fig. S7), IR (Fig. S8), and IGS (Fig. S9) datasets were largely consistent with those derived from the entire plastid genome (Fig. 8). Both the ML and BI methods yielded similar topologies across these datasets. Notably, the relationships based on the complete plastid

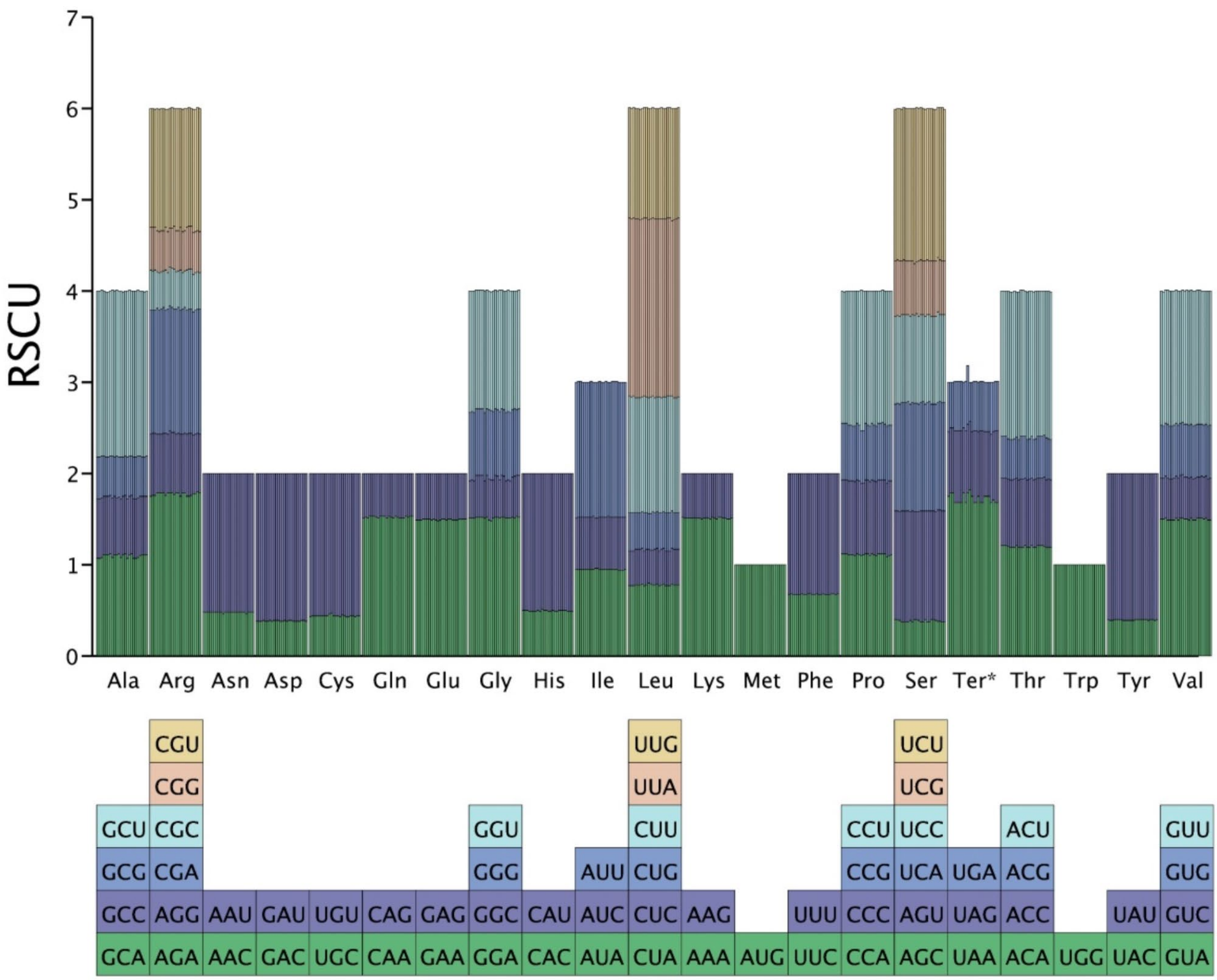

Fig. 3 Codon content of 20 amino acids and stop codons in all protein-coding genes of the 24 plastid genomes of *Potentilla* and *Dasiphora* species. The histogram above each amino acid illustrates codon usage specific to *Potentilla and Dasiphora*. Color in the column graph corresponds to the codons depicted below the figure

genome received the strongest support, with consistent topologies evident in both ML and BI trees (Fig. 8). Within the ingroup, *Potentilla* and *Dasiphora* were bifurcated into two major clades: clade I and clade II, achieving BI posterior probabilities (BIPP) ranging from 0.98 to 1.00 and ML bootstrap support (MLBS) from 61 to 100%. Specifically, clade I (BIPP/MLBS = 1/100%), comprised solely of species from the *Potentilla*, and clade II (BIPP/MLBS = 1/100%) consisted of species from the *Dasiphora* (Fig. 8). Consequently, our findings corroborate recent phylogenetic research on *Potentilla* and *Dasiphora*, demonstrating mutual monophyly between the *Potentilla* and *Dasiphora* [12].

Discussion

Plastids in most angiosperms are typically maternally inherited, exhibit rare recombination, and maintain highly conserved structures, making them invaluable for phylogenetic studies [17, 32, 66, 67]. We sequenced and annotated the plastid genomes of 24 *Potentilla* and *Dasiphora* species. Our analysis revealed that these genomes are structurally and genetically uniform, with sizes ranging from 152,883 bp to 156,461 bp, containing 84 protein-coding sequences, 37 tRNA genes, and 8 rRNA genes (Fig. 1; Table 1), consistent with previous findings [12]. The GC content varied across regions, notably higher in the inverted repeat (IR), likely influenced by the presence of rRNA genes (*rrn*4.5, *rrn*5, *rrn*16, and *rrn*23) [68–70]. Plastid genome sizes showed variability even within species, exemplified by *D. parvifolia* p132 (152,981 bp) and *D. parvifolia* p158 (152,883 bp), as well as *P. multifida* p103 (152,971 bp) and *P. multifida* p110 (156,330 bp) (Table 1), reflecting both IR region expansions/contractions and plant polyploidization events [71].

We observed that the boundary structures of 24 *Potentilla* and *Dasiphora* plastid genomes exhibited high

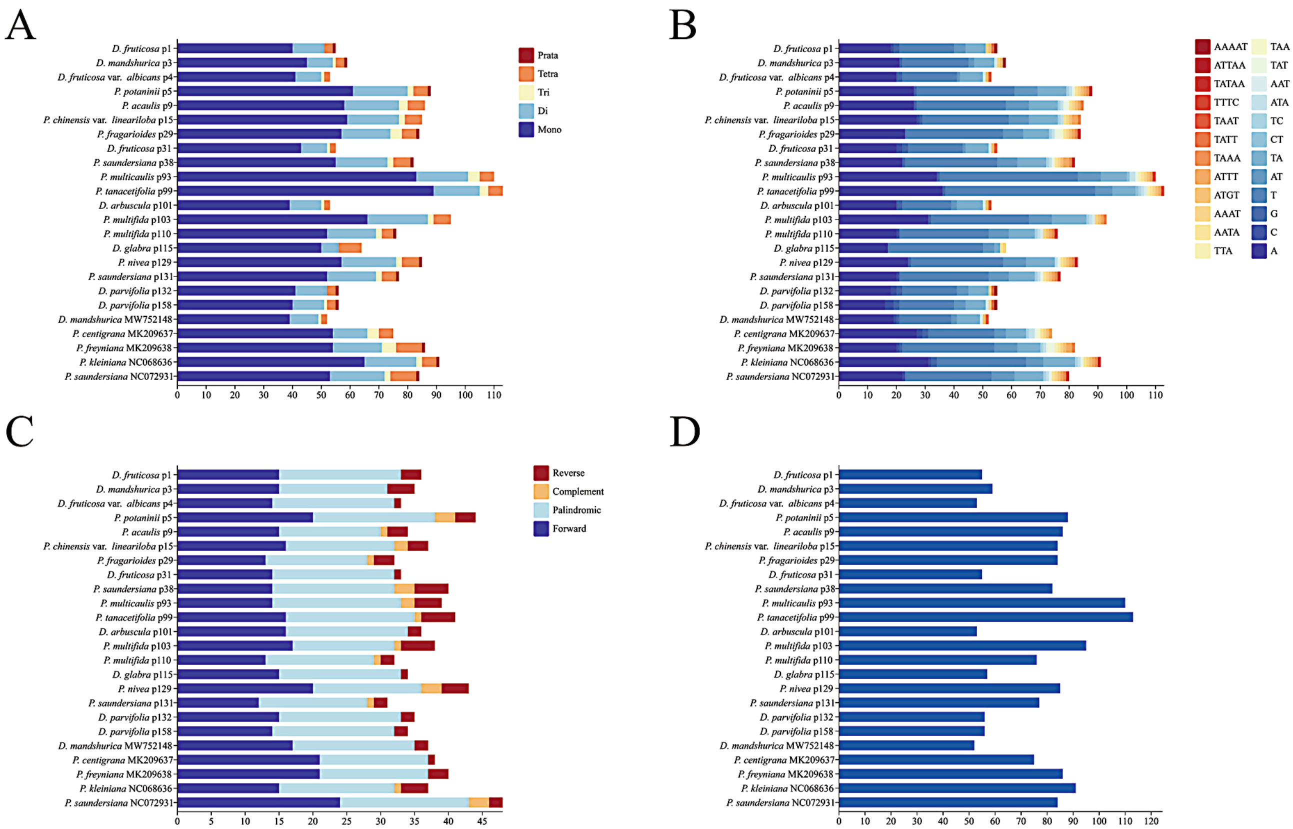

Fig. 4 Analysis of repeated sequences of the 24 *Potentilla* and *Dasiphora* species plastid genomes compared in this study. (**A**): The number of different SSR types; (**B**): The number of different SSR repeat unit types; (**C**): The number of long repeat types; (**D**): The total number of SSR

similarity across their four IR/SC boundaries (Fig. 2). Despite the general stability of these boundary regions, variations were noted in the extent of shifts for the four genes *ycf1*, *rps19*, *ndhF*, and *trnN*, which are located at these boundaries. These shifts likely result from instability in the IR boundaries coupled with varying degrees of regional expansion and contraction. Furthermore, our analysis revealed that the *rpl2* gene in the plastid genomes of all species within the genus v contained one fewer intron than that in *Potentilla*, suggesting an intron deletion event in the *rpl2* gene [72].

SSRs in plastid genomes are notably polymorphic and have been extensively utilized as molecular markers in species identification and phylogenetic analyses [62–64]. We identified 1,820 SSRs across 24 species, with *P. tanacetifolia* p99 displaying the highest frequency of single-nucleotide repeats (Fig. 4). Predominantly, these repetitive sequences were located in the large single copy (LSC) region and consisted mainly of single-nucleotide repeats, corroborating prior studies [73]. Additionally, most SSRs were found in spacer and non-coding regions, with a minor presence in the coding areas. Notably, repetitive sequences within coding regions were primarily found in exons of genes such as *ycf1*, *ClpP*, *rpoA*, *rpoB*, and *rpoC2*. SSRs in non-coding regions often exhibit intraspecific variability in repeat numbers [74], making them effective molecular markers for differentiating species within *Potentilla* and *Dasiphora*.

mVISTA and sliding window analyses of 24 *Potentilla* and *Dasiphora* plastid genomes identified high variability in three protein-coding genes (*rps16*, *matK*, and *ycf1*) and 7 non-coding regions (*rpl32-trnL-UAG*, *rps4-trnT-UGU*, *trnH-GUG-psbA*, *rps16-trnQ-UUG*, *rps15-ycf1*, *trnT-UGU-trnL-UAA*, and *petA-psbJ*) (Figs. 5 and 6). Notably, *rpl32-trnL-UAG* was identified as a specific molecular marker for distinguishing asparagus species [75], while *rps4-trnT-UGU*, *rps16-trnQ-UUG*, and *trnH-GUG-psbA* emerged as potential markers for species identification [76–78]. We also observed that the IR regions were predominantly conserved, whereas non-coding regions exhibited greater variability than coding regions. Studies suggest that non-coding regions evolve more rapidly than coding regions, with their variants providing significant insights for phylogenetic analyses across species [13]. This heightened variability in non-coding regions within the plastid genomes of *Potentilla* and *Dasiphora* underscores their potential in elucidating intergeneric relationships.

Ka/Ks ratios are critical metrics in gene evolution studies, providing insights into substitution rates, selection

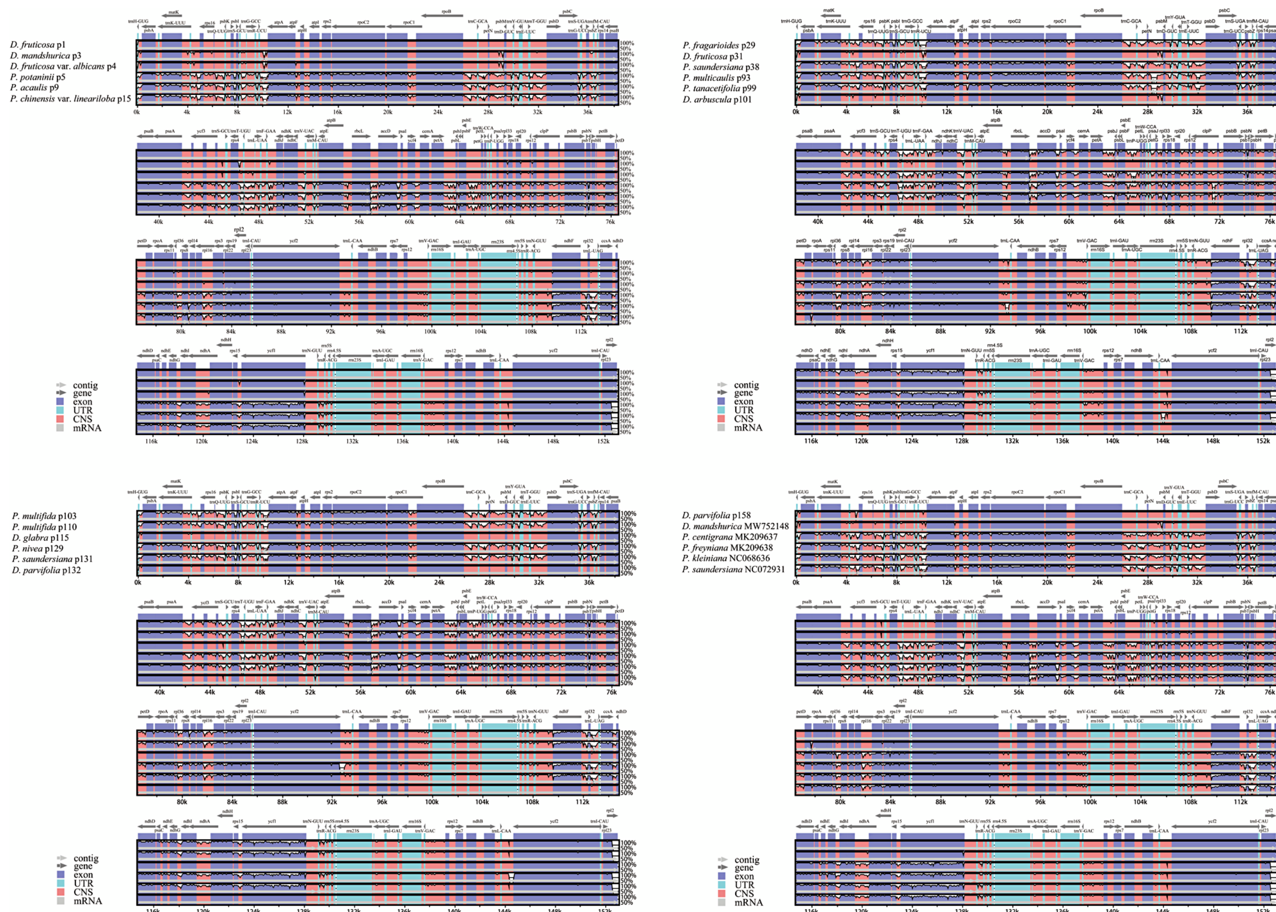

Fig. 5 Nine chloroplast genomes were compared using the *D. fruticosa* p1 annotation as a reference. The vertical scale, ranging from 50–100%, indicates the percentage of sequence identity. Arrows depict the transcriptional direction of each annotated gene

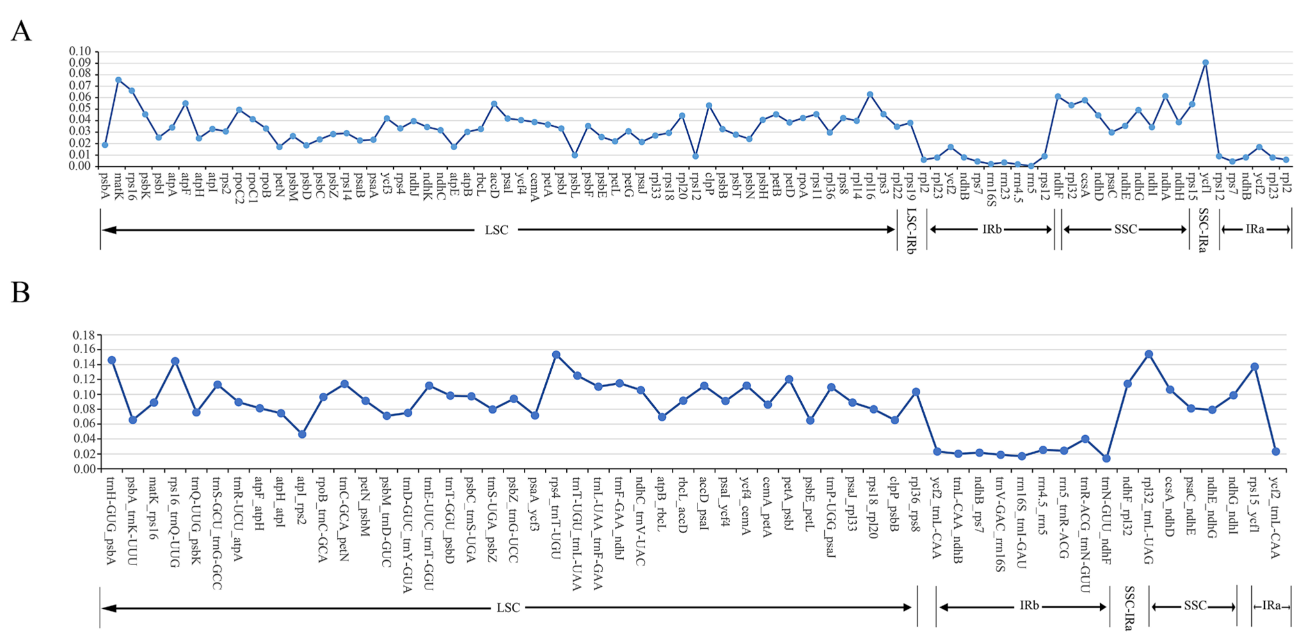

Fig. 6 Sliding window analysis of 24 *Potentilla* and *Dasiphora* species plastid genomes. (**A**): Comparison of nucleotide variability (Pi) within protein-coding regions; (**B**): Analysis of nucleotide variability across non-coding regions

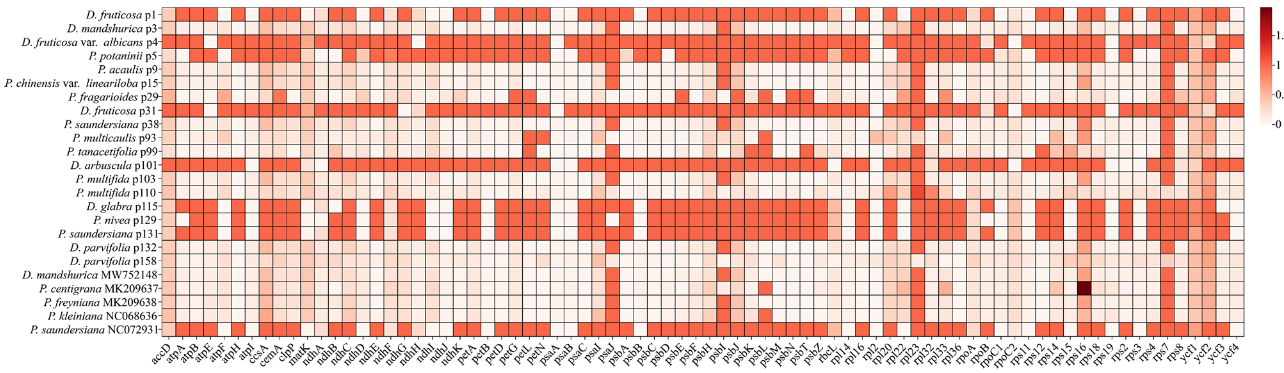

Fig. 7 Heatmap of the Ka/Ks values among 24 *Potentilla* and *Dasiphora* species. Ka/Ks > 1 indicate positive selection, Ka/Ks = 1 indicate neutral selection, and Ka/Ks < 1 indicate purified selection. The darker the color of the square, the larger the value of Ka/Ks, and vice versa

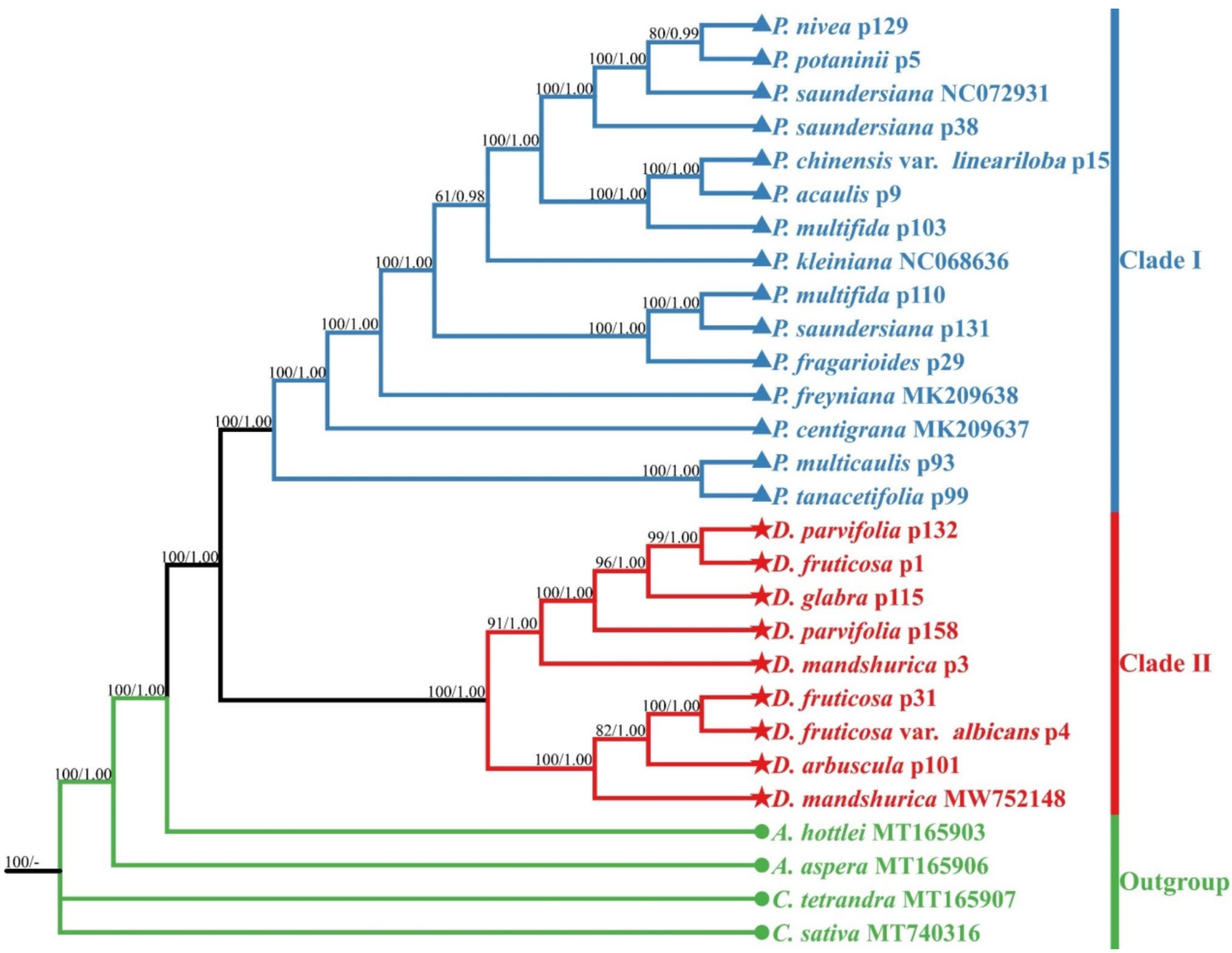

Fig. 8 ML/BI phylogenetic tree of 24 *Potentilla* and *Dasiphora* species based on plastid genomes. The numbers above the branches represent ML bootstrap values (BS)/BI posterior probabilities (PP)

pressures, and advantageous mutations in selected genes [50, 79, 80]. Our analysis revealed that a subset of genes is subject to neutral selection. In contrast, another subset undergoes purifying selection (Fig. 7). Purifying selection diminishes genetic diversity both at directly selected loci and at linked neutral sites, a process believed to significantly influence genomic diversity within natural populations [81, 82]. Based on these findings, we propose that purifying selection is crucial for preserving complex traits in *Potentilla* and *Dasiphora* species.

The taxonomic boundaries between *Potentilla* and *Dasiphora* are currently under dispute, with unclear genus delineations leading to frequent nomenclatural revisions, as numerous *Dasiphora* species have been reclassified under *Potentilla* [1–3, 5–10]. This underscores the urgent need to clarify the taxonomic status of these genera. Although previous studies using ITS and *trnL-F* sequences have not conclusively demonstrated the monophyly of *Potentilla* [35], more recent research has supported its monophyly without fully resolving its phylogenetic relationship with *Dasiphora* [11]. The latest phylogenetic analyses within the Potentilleae tribe offer a more robust framework [12], suggesting that plastid genomes provide superior resolution of angiosperm phylogenies at the generic level [32].

Our study revealed significant inconsistencies between plastid and nuclear phylogenies within the tribes comprising the genera *Potentilla* and *Dasiphora*. We reconstructed the phylogenetic relationships of 24 species from these genera using eight datasets derived from plastid genomes, nrDNA, and mitochondrial genes. The results indicated that phylogenetic trees constructed from nuclear genes (ITS, ETS) and cytoplasmic genes (plastid and mitochondrial) exhibited conflicting topologies. The most notable conflict was observed in the placement of P. *multicaulis* 93 and P. *tanacetifolia* 99, which clustered into a single branch with high support in the nuclear gene phylogeny, resulting in the non-monophyly of Potentilla (Figs. S1, S2 and S3). Among the eight datasets, the phylogenetic tree based on the whole plastid genome exhibited the most robust topology and the highest support values (Fig. 8). These findings align with the most recent phylogenetic framework for Potentilleae [12]. Together with previous studies, our results strongly support the monophyletic evolution of both *Potentilla* and *Dasiphora*.

In contrast, Potentilla was identified as a polyphyletic group with more derived positions based on nuclear gene evidence. Conflicts between nuclear and plastid phylogenies likely result from convergent evolution, incomplete lineage sorting, and reticulation during evolution [83]. Geological and climatic changes have created diverse habitats, promoting local adaptation and ecological differentiation, and have also driven species formation. In alpine plants on the Tibetan Plateau, secondary contact due to range shifts often induces interspecific hybridization or plastid capture, contributing to discordance between nuclear and plastid phylogenies [34, 84, 85]. Likewise, rapid cooling after the late Pliocene has been linked to the radial diversification of Potentilla [86]. Collectively, geographic and climatic changes have played a central role in shaping nuclear-plastid conflicts in the phylogeny of *Potentilla*.

Conflicts between organelle genomic data and nrDNA are common in *Potentilla*. For instance, the positions of the Reptans and Fragarioides branches are reversed in organelle-based versus nrDNA-based phylogenies. These discrepancies suggest hybridization and incomplete lineage sorting may be frequent within the genus. We hypothesize that the topological inconsistencies between nuclear and plastid phylogenetic trees of *Potentilla* and *Dasiphora*, particularly regarding the monophyly of *Potentilla*, are likely driven by interspecific hybridization and incomplete genealogical sorting. To better resolve species boundaries and unravel the evolutionary complexity and cytonuclear discordance in this tribe, future studies incorporating single-copy nuclear genes and expanded taxon sampling will be essential.

Conclusions

In this study, we selected 24 species from the *Potentilla* and *Dasiphora* and assembled their complete plastid genomes, which supplemented nrITS, nrETS, and mitochondrial sequences. The plastid genomes across all species showed high conservation, with lengths ranging from 152,883 to 156,461 base pairs and comprising 84 protein-coding genes, 37 tRNA genes, and 8 rRNA genes. Notably, intron deletions were universally observed in the *rpl2* genes of the genus *Dasiphora*. Seven variable regions (*rpl32-trnL-UAG*, *rps4-trnT-UGU*, *trnH-GUG-psbA*, *rps16-trnQ-UUG*, *rps15-ycf1*, *trnT-UGU-trnL-UAA*, and *petA-psbJ*) exhibited high variability and show promise as molecular markers for differentiating between *Potentilla* and *Dasiphora*. Phylogenetic analyses based on nuclear and plastid data supported the monophyly of these genera relative to each other. Our findings suggest that interspecific hybridization or incomplete lineage sorting events may have influenced the evolutionary trajectory of *Potentilla*. Future studies should incorporate nuclear genomic data to investigate the underlying causes of phylogenetic conflicts comprehensively. These insights enhance our understanding of the plastid genomes within *Potentilla* and *Dasiphora* and aid in developing molecular markers to clarify the intricate relationships among species, thereby underpinning further phylogenetic and biogeographic research.

Abbreviations

tRNA	transfer RNA
rRNA	ribosomal RNA
LSC	Large Single-Copy
SSC	Small Single-Copy
IR	Inverted Repeat
GC	Guanine-Cytosine
CDS	Coding Sequences
PCGs	Protein-Coding Genes
IGS	Intergenic Spacer
RSCU	Relative Synonymous Codon Usage
ML	Maximum Likelihood
BI	Bayesian Interference

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12870-025-06186-6 .

Supplementary Material 1

Supplementary Material 2

Acknowledgements

We thanks to Mr. Xiaofeng Chi and Mr. Mingze Xia from the Northwest Institute of Plateau Biology, Chinese Academy of Sciences, for their invaluable assistance with site distribution and sampling.

Author contributions

FQZ and XPL designed this study. XPL, YH, HX, SH, and YN conducted the sampling. JYY and HX analyzed the data. XPL prepared the manuscript. FQZ revised the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This research was partially supported by the Second Tibetan Plateau Scientific Expedition and Research (STEP) program (2019QZKK0502), Qinghai Provincial Science and Technology Major Project (2023-SF-A5), the Biological Resources Programme of Chinese Academy of Sciences (KFJBRP-017–101), Xining Science and Technology Major Project (2023-Z-13), CAS - Qinghai on Sanjiangyuan National Park (LHZX-2021–04), CAS "Light of West China" Program (2024).

Data availability

The datasets generated and analysed during the current study are available in the NCBI (https://www.ncbi.nlm.nih.gov/) repository, OR601516, OR601528, OR601532, OR601533, OR601536, OR601525, OR601527, OR601529, OR601531, OR601537, OR601538, OR601518, OR601519, OR601520, OR601521, OR601522, OR601523, OR601524, OR601526, MW752148, MW752148, MW752148, MW752148, MW752148, MT165903, MT165906, MT740316, MT165907.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 5 September 2024 / Accepted: 31 January 2025

Published online: 10 February 2025

References

1. Dobeš C, Paule J. A comprehensive chloroplast DNA-based phylogeny of the genus *Potentilla* (Rosaceae): implications for its geographic origin, phylogeography and generic circumscription. Mol Phylogenet Evol. 2010;56(1):156–75.
2. Eriksson T, Donoghue MJ, Hibbs MS. Phylogenetic analysis of Potentilla using DNA sequences of nuclear ribosomal internal transcribed spacers (ITS), and implications for the classification of Rosoideae (Rosaceae). Plant Syst Evol. 1998;211(3):155–79.
3. Gehrke B, Bräuchler C, Romoleroux K, Lundberg M, Heubl G, Eriksson T. Molecular phylogenetics of *Alchemilla*, *Aphanes* and *Lachemilla* (Rosaceae) inferred from plastid and nuclear intron and spacer DNA sequences, with comments on generic classification. Mol Phylogenet Evol. 2008;47(3):1030–44.
4. Klackenberg J. The holarctic complex *Potentilla fruticosa* (Rosaceae). Nord J Bot. 1983;3(2):181–91.
5. Eriksson T, Hibbs MS, Yoder AD, Delwiche CF, Donoghue MJ. The phylogeny of Rosoideae (Rosaceae) based on sequences of the internal transcribed spacers (ITS) of nuclear ribosomal DNA and the *trnL/F* region of chloroplast DNA. Int J Plant Sci. 2003;164(2):197–211.
6. Eriksson T, Lundberg M, Töpel M, Östensson P, Smedmark JEE. *Sibbaadia*: A molecular phylogenetic study of a remarkably polyphyletic genus in Rosaceae. Plant Syst Evol. 2015;301(1):171–84.
7. Feng T, Moore MJ, Yan MH, Sun YX, Zhang HJ, Meng AP, et al. Phylogenetic study of the tribe Potentilleae (Rosaceae), with further insight into the disintegration of *Sibbaldia*. J Syst Evol. 2017;55(3):177–91.
8. Morales-Briones DF, Tank DC. Extensive allopolyploidy in the neotropical genus *Lachemilla* (Rosaceae) revealed by PCR-based target enrichment of the nuclear ribosomal DNA cistron and plastid phylogenomics. Am J Bot. 2019;106(3):415–37.
9. Persson NL, Eriksson T, Smedmark JEE. Complex patterns of reticulate evolution in opportunistic weeds (*Potentilla* L., Rosaceae), as revealed by low-copy nuclear markers. BMC Evol Biol. 2020;20(1):38.
10. Persson NL, Toresen I, Andersen HL, Smedmark JEE, Eriksson T. Detecting destabilizing species in the phylogenetic backbone of *Potentilla* (Rosaceae) using low-copy nuclear markers. AoB Plants. 2020;12(3):plaa017.
11. Eriksson T, Persson NL, Smedmark JEE. What is *Potentilla*? A phylogeny-based taxonomy for Potentillinae (Rosaceae). Taxon. 2022;71(3):493–505.
12. Li QQ, Khasbagan, Zhang ZP, Wen J, Yu Y. Plastid phylogenomics of the tribe Potentilleae (Rosaceae). Mol Phylogenet Evol. 2024;190:107961.
13. Shaw J, Lickey EB, Schilling EE, Small RL. Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: the tortoise and the hare III. Am J Bot. 2007;94(3):275–88.
14. Ruhlman TA, Jansen RK. The plastid genomes of flowering plants. In: Maliga P, editor. Chloroplast Biotechnology: methods and protocols. Totowa, NJ: Humana; 2014. pp. 3–38.
15. Daniell H, Lin CS, Yu M, Chang WJ. Chloroplast genomes: diversity, evolution, and applications in genetic engineering. Genome Biol. 2016;17(1):134.
16. Jiang H, Tian J, Yang J, Dong X, Zhong Z, Mwachala G, et al. Comparative and phylogenetic analyses of six Kenya *Polystachya* (Orchidaceae) species based on the complete chloroplast genome sequences. BMC Plant Biol. 2022;22(1):177.
17. Birky CW. Uniparental inheritance of mitochondrial and chloroplast genes: mechanisms and evolution. Proc Natl Acad Sci U S A. 1995;92(25):11331–8.
18. Jansen RK, Raubeson LA, Boore JL, dePamphilis CW, Chumley TW, Haberle RC, et al. Methods for obtaining and analyzing whole chloroplast genome sequences. Methods Enzymol. Volume 395. Academic; 2005. pp. 348–84.
19. Wicke S, Schneeweiss GM, dePamphilis CW, Müller KF, Quandt D. The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. Plant Mol Biol. 2011;76(3):273–97.
20. Raubeson LA, Jansen RK. Chloroplast genomes of plants. CAB Reviews. 2005:45–68.
21. Wang JR, Feng L, Liu S, Pang HB, Qi L, Li J, Sun Y, Qiao WH, Zhang LF, Cheng YL, et al. Inferring the evolutionary mechanism of the chloroplast genome size by comparing whole-chloroplast genome sequences in seed plants. Sci Rep. 2017;7(1):1555.
22. Ravi V, Khurana JP, Tyagi AK, Khurana P. An update on chloroplast genomes. Plant Syst Evol. 2008;271(1):101–22.
23. Xiang CL, Dong HJ, Landrein S, Zhao F, Yu WB, Soltis DE, et al. Revisiting the phylogeny of Dipsacales: new insights from phylogenomic analyses of complete plastomic sequences. J Syst Evol. 2020;58(2):103–17.
24. Zhang D, Tu J, Ding X, Guan W, Gong L, Qiu X, et al. Analysis of the chloroplast genome and phylogenetic evolution of Bidens pilosa. BMC Genomics. 2023;24(1):113.
25. Li Z, Duan B, Zhou Z, Fang H, Yang M, Xia C, et al. Comparative analysis of medicinal plants *Scutellaria baicalensis* and common adulterants based on chloroplast genome sequencing. BMC Genomics. 2024;25(1):39.
26. Straub SCK, Parks M, Weitemier K, Fishbein M, Cronn RC, Liston A. Navigating the tip of the genomic iceberg: next-generation sequencing for plant systematics. Am J Bot. 2012;99(2):349–64.
27. Deng JB, Drew BT, Mavrodiev EV, Gitzendanner MA, Soltis PS, Soltis DE. Phylogeny, divergence times, and historical biogeography of the angiosperm family Saxifragaceae. Mol Phylogenet Evol. 2015;83:86–98.
28. Liu L, Wang Y, He P, Li P, Lee J, Soltis DE, et al. Chloroplast genome analyses and genomic resource development for epilithic sister genera Oresitrophe and Mukdenia (Saxifragaceae), using genome skimming data. BMC Genomics. 2018;19(1):235.

29. Kane N, Sveinsson S, Dempewolf H, Yang JY, Zhang D, Engels JMM, et al. Ultra-barcoding in cacao (*Theobroma* spp.; Malvaceae) using whole chloroplast genomes and nuclear ribosomal DNA. Am J Bot. 2012;99(2):320–9.
30. Liu L, Zhang C, Wang Y, Dong M, Shang F, Li P. The complete chloroplast genome of *Caryopteris mongholica* and phylogenetic implications in Lamiaceae. Conserv Genet Resour. 2018;10(3):281–5.
31. Xu WQ, Losh J, Chen C, Li P, Wang RH, Zhao YP, et al. Comparative genomics of figworts (*Scrophularia*, Scrophulariaceae), with implications for the evolution of *Scrophularia* and Lamiales. J Syst Evol. 2019;57(1):55–65.
32. Yu JY, Han Y, Xu H, Han S, Li XP, Niu Y, et al. Structural divergence and phylogenetic relationships of *Ajania* (Asteraceae) from plastomes and ETS. BMC Genomics. 2023;24(1):602.
33. Seehausen O. Hybridization and adaptive radiation. Trends Ecol Evol. 2004;19(4):198–207.
34. Duan L, Fu L, Chen HF. Phylogenomic cytonuclear discordance and evolutionary histories of plants and animals. Sci China Life Sci. 2023;66(12):2946–8.
35. Faghir MB, Attar F, Farazmand A, Osaloo SK. Phylogeny of the genus *Potentilla* (Rosaceae) in Iran based on nrDNA ITS and cpDNA *trn*L-F sequences with a focus on leaf and style characters' evolution. Turk J Bot. 2014;38:417–29.
36. Chen S, Zhou Y, Chen Y, Gu J. Fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics. 2018;34(17):i884–90.
37. 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.
38. Shi L, Chen H, Jiang M, Wang L, Wu X, Huang L, et al. CpGAVAS2, an integrated plastome sequence annotator and analyzer. Nucleic Acids Res. 2019;47(W1):W65–73.
39. Qu XJ, Moore MJ, Li DZ, Yi TS. PGA: a software package for rapid, accurate, and flexible batch annotation of plastomes. Plant Methods. 2019;15(1):50.
40. Chan PP, Lin BY, Mak AJ, Lowe TM. tRNAscan-SE 2.0: improved detection and functional classification of transfer RNA genes. Nucleic Acids Res. 2021;49(16):9077–96.
41. Zheng S, Poczai P, Hyvönen J, Tang J, Amiryousefi A. Chloroplot: an online program for the versatile plotting of organelle genomes. Front Genet. 2020;11:576124.
42. Mazumdar P, Binti Othman R, Mebus K, Ramakrishnan N, Harikrishna JA. Codon usage and codon pair patterns in non-grass monocot genomes. Ann Bot. 2017;120(6):893–909.
43. Thiel T, Michalek W, Varshney R, Graner A. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L). Theor Appl Genet. 2003;106(3):411–22.
44. Kurtz S, Choudhuri JV, Ohlebusch E, Schleiermacher C, Stoye J, Giegerich R, REPuter. The manifold applications of repeat analysis on a genomic scale. Nucleic Acids Res. 2001;29(22):4633–42.
45. Amiryousefi A, Hyvönen J, Poczai P, IRscope. An online program to visualize the junction sites of chloroplast genomes. Bioinformatics. 2018;34(17):3030–1.
46. Rozas J, Ferrer-Mata A, Sánchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE, Sánchez-Gracia A. DnaSP 6: DNA sequence polymorphism analysis of large data sets. Mol Biol Evol. 2017;34(12):3299–302.
47. Mayor C, Brudno M, Schwartz JR, Poliakov A, Rubin EM, Frazer KA, et al. VISTA: visualizing global DNA sequence alignments of arbitrary length. Bioinformatics. 2000;16(11):1046–7.
48. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013;30(4):772–80.
49. Zhang Z, Xie P, Guo Y, Zhou W, Liu E, Yu Y. Easy353: a tool to get Angiosperms353 genes for phylogenomic research. Mol Biol Evol. 2022;39(12):msac261.
50. Wang D, Zhang Y, Zhang Z, Zhu J, Yu J. KaKs_Calculator 2.0: a toolkit incorporating gamma-series methods and sliding window strategies. Genomics Proteom Bioinf. 2010;8(1):77–80.
51. 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.
52. Ranwez V, Douzery EJP, Cambon C, Chantret N, Delsuc F. MACSE v2: Toolkit for the alignment of coding sequences accounting for frameshifts and stop codons. Mol Biol Evol. 2018;35(10):2582–4.
53. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods. 2017;14(6):587–9.
54. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol. 2015;32(1):268–74.
55. Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, et al. MrBayes 3.2: efficient bayesian phylogenetic inference and model choice across a large model space. Syst Biol. 2012;61(3):539–42.
56. Asaf S, Khan AL, Khan AR, Waqas M, Kang SM, Khan MA et al. Complete chloroplast genome of *Nicotiana otophora* and its comparison with related species. Front Plant Sci. 2016; 7.
57. Yue F, Cui L, dePamphilis CW, Moret BME, Tang J. Gene rearrangement analysis and ancestral order inference from chloroplast genomes with inverted repeat. BMC Genomics. 2008;9(Suppl 1):S25.
58. Wu L, Nie L, Guo S, Wang Q, Wu Z, Lin Y, et al. Identification of medicinal *Bidens* plants for quality control based on organelle genomes. Front Pharmacol. 2022;13:842131.
59. Gong L, Ding X, Guan W, Zhang D, Zhang J, Bai J, et al. Comparative chloroplast genome analyses of *Amomum*: insights into evolutionary history and species identification. BMC Plant Biol. 2022;22(1):520.
60. Aljuhani WS, Aljohani AY. Complete chloroplast genome of the medicinal plant *Cleome Paradoxa* R.Br. Ex DC: comparative analysis, and phylogenetic relationships among the members of Cleomaceae. Gene. 2022;845:146851.
61. Parvathy ST, Udayasuriyan V, Bhadana V. Codon usage bias. Mol Biol Rep. 2022;49(1):539–65.
62. Yin J, Zhao H, Wu X, Ma Y, Zhang J, Li Y, et al. SSR marker-based analysis for identification and of genetic diversity of non-heading Chinese cabbage varieties. Front Plant Sci. 2023;14:1112748.
63. Yisilam G, Wang CX, Xia MQ, Comes HP, Li P, Li J, et al. Phylogeography and population genetics analyses reveal evolutionary history of the desert resource plant *Lycium Ruthenicum* (Solanaceae). Front Plant Sci. 2022;13:915526.
64. Pei D, Song S, Kang J, Zhang C, Wang J, Dong T, et al. Characterization of simple sequence repeat (SSR) markers mined in whole grape genomes. Genes (Basel). 2023;14(3):663.
65. Xiang CY, Gao F, Jakovlić I, Lei HP, Hu Y, Zhang H, et al. Using PhyloSuite for molecular phylogeny and tree-based analyses. iMeta. 2023;2(1):e87.
66. Palmer JD. Comparative organization of chloroplast genomes. Annu Rev Genet. 1985;19:325–54.
67. Zhang B, Liu X, Xie X, Huan L, Shao Z, Du Z, et al. Genetic evidence for functions of chloroplast CA in *Pyropia yezoensis*: decreased CCM but increased starch accumulation. Adv Biotechnol. 2024;2(2):16.
68. Peng JY, Zhang XS, Zhang DG, Wang Y, Deng T, Huang XH et al. Newly reported chloroplast genome of *Sinosenecio albonervius* Y. Liu & Q. E. Yang and comparative analyses with other *Sinosenecio* species. BMC Genomics. 2022; 23(1): 639.
69. Yan K, Ran J, Bao S, Li Y, Islam R, Zhang N et al. The complete chloroplast genome sequence of *Eupatorium fortunei*: genome organization and comparison with related species. Genes (Basel). 2022; 14(1).
70. Wang L, Wang Y, Yang W, He X, Xu S, Liu X, et al. Network pharmacology and molecular docking analysis on mechanisms of tibetan Hongjingtian (*Rhodiola crenulata*) in the treatment of COVID-19. J Med Microbiol. 2021;70(7):001374.
71. Dodsworth S, Leitch AR, Leitch IJ. Genome size diversity in angiosperms and its influence on gene space. Curr Opin Genet Dev. 2015;35:73–8.
72. Shi YT, Xu L, Shao Z, Gu CH. Loss of *rpl2* intron in Lythraceae. J Zhejiang Sci Technol. 2018;38(2):63–8.
73. Chen C, Xia X, Peng J, Wang D. Comparative analyses of six complete chloroplast genomes from the genus *Cupressus* and *Juniperus* (Cupressaceae). Gene. 2022; 837: 146696.
74. Zhou T, Wang J, Jia Y, Li W, Xu F, Wang X. Comparative chloroplast genome analyses of species in Gentiana section Cruciata (Gentianaceae) and the development of authentication markers. Int J Mol Sci. 2018;19(7):1962.
75. Tian Y, Liu X, Xu Y, Yu B, Wang L, Qu X. Comparative and phylogenetic analysis of *Asparagus meioclados* Levl. and *Asparagus munitus* Wang et S. C. Chen plastomes and utility of plastomes mutational hotspots. Sci Rep. 2023; 13(1): 15622.
76. Mao L, Zou Q, Sun Z, Dong Q, Cao X. Insights into chloroplast genome structure, intraspecific variation, and phylogeny of *Cyclamen* species (Myrsinoideae). Sci Rep. 2023;13(1):87.
77. Loera-Sánchez M, Studer B, Kölliker R. DNA barcode trnh-psba is a promising candidate for efficient identification of forage legumes and grasses. BMC Res Notes. 2020;13(1):35.
78. Kim GB, Lim CE, Kim JS, Kim K, Lee JH, Yu HJ, et al. Comparative chloroplast genome analysis of *Artemisia* (Asteraceae) in East Asia: insights into

evolutionary divergence and phylogenomic implications. BMC Genomics. 2020;21(1):415.
79. Liu F, Movahedi A, Yang W, Xu L, Xie J, Zhang Y. The complete chloroplast genome and characteristics analysis of *Callistemon rigidus* R.Br. Mol Biol Rep. 2020;47(7):5013–24.
80. Zhang Z. KaKs_Calculator 3.0: calculating selective pressure on coding and non-coding sequences. Genomics Proteom Bioinf. 2022;20(3):536–40.
81. Brunet TDP, Doolittle WF, Bielawski JP. The role of purifying selection in the origin and maintenance of complex function. Stud Hist Philos Sci Part A. 2021;87:125–35.
82. Cvijović I, Good BH, Desai MM. The effect of strong purifying selection on genetic diversity. Genetics. 2018;209(4):1235–78.
83. Cristina Acosta M, Premoli AC. Evidence of chloroplast capture in South American *Nothofagus* (subgenus *Nothofagus*, Nothofagaceae). Mol Phylogenet Evol. 2010; 54(1): 235–242.
84. Qin HT, Möller M, Milne R, Luo YH, Zhu GF, Li DZ, et al. Multiple paternally inherited chloroplast capture events associated with *Taxus* speciation in the Hengduan Mountains. Mol Phylogenet Evol. 2023;189:107915.
85. Liu LM, Du XY, Guo C, Li DZ. Resolving robust phylogenetic relationships of core Brassicaceae using genome skimming data. J Syst Evol. 2021;59(3):442–53.
86. Xue T, Feng T, Liang Y, Yang X, Qin F, Yu J et al. Radiating diversification and niche conservatism jointly shape the inverse latitudinal diversity gradient of *Potentilla* L. (Rosaceae). BMC Plant Biol. 2024; 24(1): 443.

Publisher's note