RESEARCH

Plastid genome and its phylogenetic implications of Asiatic *Spiraea* (Rosaceae)

Shu-Yan Zhang[1,2,3,4†], Hai-Fei Yan[2,3,4†], Lei Wei[2,3,4], Tong-Jian Liu[2,3,4], Lin Chen[5], Gang Hao[1], Xing Wu[2,3,4*] and Qiao-Ling Zhang[5*]

Abstract

Background *Spiraea* L. is a genus comprising approximately 90 species that are distributed throughout the northern temperate regions. China is recognized as the center of species diversity for this genus, hosting more than 70 species, including 47 endemic species. While *Spiraea* is well-known for its ornamental value, its taxonomic and phylogenetic studies have been insufficient.

Results In this study, we conducted sequencing and assembly of the plastid genomes (plastomes) of 34 Asiatic *Spiraea* accessions (representing 27 Asiatic *Spiraea* species) from China and neighboring regions. The *Spiraea* plastid genome exhibits typical quadripartite structures and encodes 113–114 genes, including 78–79 protein-coding genes (PCGs), 30 tRNA genes, and 4 rRNA genes. Linear regression analysis revealed a significant correlation between genome size and the length of the SC region. By the sliding windows method, we identified several hypervariable hotspots within the *Spiraea* plastome, all of which were localized in the SC regions. Our phylogenomic analysis successfully established a robust phylogenetic framework for *Spiraea*, but it did not support the current defined section boundaries. Additionally, we discovered that the genus underwent diversification after the Early Oligocene (~ 30 Ma), followed by a rapid speciation process during the Pliocene and Pleistocene periods.

Conclusions The plastomes of *Spiraea* provided us invaluable insights into its phylogenetic relationships and evolutionary history. In conjunction with plastome data, further investigations utilizing other genomes, such as the nuclear genome, are urgently needed to enhance our understanding of the evolutionary history of this genus.

Keywords *Spiraea*, Phylogeny, Time divergence, Chloroplast genome, Positive selection

†Shu-Yan Zhang and Hai-Fei Yan contributed equally to this work.

*Correspondence:
Xing Wu
wuxing@scbg.ac.cn
Qiao-Ling Zhang
3905397@qq.com
[1] College of Life Sciences, South China Agricultural University, Guangzhou 510642, China
[2] Key Laboratory of Plant Resources Conservation and Sustainable Utilization, South China Botanical Garden, Chinese Academy of Sciences, Guangzhou 510650, China
[3] South China National Botanical Garden, Guangzhou 510650, China
[4] Guangdong Provincial Key Laboratory of Applied Botany, South China Botanical Garden, Chinese Academy of Sciences, Guangzhou 510650, China
[5] Hangzhou Xixi National Wetland Park Service Center (Hangzhou Xixi National Wetland Park Ecology & Culture Research Center), Hangzhou 310013, China

Background

Spiraea L., a member of the tribe Spiraeeae of Amygdaloideae (Rosaceae), comprises approximately 90 species (http://www.plantsoftheworldonline.org), and is distributed across the northern temperate regions of the world. China is recognized as the center of diversity for this genus, with over 70 species, including 47 endemic species [1]. *Spiraea* is renowned for its ornamental value, particularly its attractive white, pink or purple flowers that bloom profusely in spring or summer, forming heads composed of numerous tiny blooms. Additionally, certain varieties of *Spiraea* with golden foliage, such as *Spiraea×vanhouttei* 'Pink Ice' and 'Gold Fountain', have been developed. To fully exploit their horticultural value, a comprehensive taxonomic and phylogenetic study of *Spiraea* taxa is necessary. However, the phylogeny of *Spiraea* has not been extensively investigated thus far.

The traditional infrageneric classification of *Spiraea* is primarily based on inflorescence types, including panicles, compound corymbs, corymb, umbel, and fascicle, as established by Poyarkova [2], Rehder [3], and Yü and Kuan [4]. For Chinese *Spiraea* taxa, four sections (i.e., sects. *Spiraea*, *Calospira*, *Chamaedryon*, and *Glomerati*) and ten series within the genus have been widely accepted [4]. Previous phylogenetic analyses have revealed issues with traditional classifications based on inflorescence morphology, as the majority of sections are considered paraphyletic or polyphyletic [5–10]. However, most nodes in these phylogenetic trees could not receive strong support values based on limited molecular fragments, highlighting the need for the development of robust molecular markers within *Spiraea*.

The plastid, an essential organelle in plant cells, plays a crucial role in plant growth and development [11]. Compared to nuclear and mitochondrial genomes, the plastid genome possesses unique characteristics, such as uniparental inheritance, absence of recombination, conserved genome structure and gene content, and moderate evolutionary rate. These features make the plastid genome highly valuable for resolving phylogenetic relationships at various taxonomic levels [12]. Furthermore, the assembly of plastid genomes is relatively straightforward due to their high cellular copy numbers and the development of assembly algorithms [13]. As a result, the number of complete plastid genomes has significantly increased in recent years, thanks to the advancement of high-throughput DNA sequencing technologies. Plastid genomes have proven to be invaluable in elucidating the phylogenetic backbone of Rosaceae and uncovering its evolutionary history [14]. The number of plastid genomes within the Rosaceae family has also rapidly increased in recent years (e.g., *Fragaria* [15, 16], *Rubus* [17], and *Rosa* [18, 19]), providing a stronger understanding of phylogenetic

relationships at the species level and the phylogeographical patterns within the family. However, a limited number of plastid genomes of *Spiraea* species have been reported [20–23], and no solid divergence times of the genus have been inferred. Therefore, it is crucial to conduct comparative analyses with more *Spiraea* plastid genomes to better identify the evolutionary pattern of the plastid genome within *Spiraea* and assess its phylogenetic relationships.

In this study, we assembled the complete plastomes for 34 *Spiraea* accessions representing 27 Asiatic species using several available *Spiraea* plastomes from GenBank. Our objectives were to (1) examine the variation in plastome structure and features of *Spiraea*, (2) identify the hypervariable loci across these *Spiraea* plastomes, (3) infer the phylogenomic relationships of *Spiraea*, and (4) determine the divergence time of *Spiraea*. This study will contribute to a better understanding of the evolution of *Spiraea* based on plastome sequences.

Materials and methods

Plant materials, DNA extraction, and sequencing

In this study, we collected a total of 27 Asiatic *Spiraea* species and 34 accessions, representing four sections (i.e., sects. *Calospira*, *Chamaedryon*, *Glomerati* and *Spiraea*) of two subgenera (i.e., subgs. *Protospiraea* and *Metaspiraea*) in *Spiraea*. Among these accessions, 25 were newly collected from Hangzhou Botanical Garden. Voucher specimens were formally identified by Qiao-Ling Zhang and deposited at the Herbarium of South China Botanical Garden (IBSC), Chinese Academy of Sciences (CAS). Silica-gel dried leaves were used for DNA extraction by the cetyltrimethylammonium bromide (CTAB) method [24]. Paired-end (PE) sequencing libraries were constructed and sequenced on an MGISEQ-2000 (MGI, Shenzhen, China) at the Beijing Genomics Institute (BGI, Wuhan, China). Additionally, twelve plastid genomes of *Spiraea* were directly downloaded from GenBank (Accessed on February 5, 2023), which belong to nine *Spiraea* species and three outgroups (i.e., *Malus baccata* (L.) Borkh., *Rosa cymose* Tratt. and *Sibiraea angustata* (Rehd.) Hand.-Mazz.) for phylogenetic inferences. Among all accessions, four *Spiraea* species (*Spiraea japonica* L.F., *S.×vanhouttei* (Birot) Zabel, *S. cantoniensis* Lour., and *S. prunifolia* Siebold & Zucc.) have multiple individuals. The detailed sampling information is summarized in Table S1.

Plastid genome assembly, annotation, and comparison

Plastomes of 25 *Spiraea* accessions were newly assembled using GetOrangelle [25] with the reference plastome of *Spiraea mongolica* (NC_051992). These complete plastid sequences were annotated using the Plastid Genome

Annotator (PGA) [26] and manually checked and adjusted with the aid of Geneious 2019 [27]. The circular physical maps were drawn using OrganellarGenome DRAW (ORDRAW) [28].

The total length of the plastomes, as well as the lengths of the large single copy (LSC), small single copy (SSC), and inverted repeat regions (IR$_A$ and IR$_B$), gene content, and GC content of all *Spiraea* species were analyzed using Geneious. The relationships among the genome size and LSC length, IR length, intron length, and intergenic spacer length of *Spiraea* species were tested using least squares linear regression analysis in R 4.2.2 [29]. The boundary information between the IRs and SC regions was analyzed in Geneious.

Divergent hotspot analyses

The plastid genome sequences were aligned using MAFFT v7.308 [30]. The nucleotide polymorphism (*Pi*) of plastid genomes was calculated using the sliding window method with a length of 600 bp and a step size of 200 bp in DnaSP v6 [31]. The sequence heterogeneity within 34 *Spiraea* plastid genomes was visualized using mVISTA with the Shuffle-LAGAN model [32].

Phylogenetic analyses and molecular dating

Phylogenetic analysis was performed on all 34 *Spiraea* plastid genome sequences, with *M. baccata* (MK571561), *R. cymosa* (NC_051550), and *Si. Angustata* (NC_054238) used as outgroups. The 79 protein-coding genes (PCGs) and whole plastid genome sequences (including all coding and non-coding regions) were aligned separately using MAFFT and concatenated into a sequence matrix using AMAS [33]. Maximum likelihood (ML) and Bayesian inference (BI) trees were each inferred using RAxML v8.2.10 [34] and MrBayes v3.2.6 [35]. The ML tree was constructed using RAxML under the GTRGAMMA model with 1,000 bootstrap replicates. BI was performed with MrBayes in the GTR+G+I model on XSEDE (the CIPRES Science Gateway, [36]). The Markov chain Monte Carlo (MCMC) chains were run for 10×10^6 generations, and stationarity was assessed using TRACER v1.7.1 (effective sample size (ESS) > 200) [37]. The first 25% of runs were discarded as burn-in, and the remaining trees were used to construct the consensus tree. Bayesian posterior probabilities (BPPs) were used to estimate support for each branch. The phylogenetic tree file was visualized in FigTree v1.4.3 [38].

To estimate the divergence time of *Spiraea*, we retrieved the other 86 plastid genomes of Rosales taxa from GenBank (Table S1). A total of 78 PCGs extracted from 120 plastid genomes were extracted and aligned separately by MAFFT. Maximum likelihood trees were constructed for each of these 78 PCGs using RAxML,

and molecular clock-like genes (clock-like genes) were screened from the 78 gene trees by using the SortaDate package [39]. The top-ten clock-like genes were selected and aligned using MAFFT, and then concatenated using AMAS for further divergence time estimation.

The concatenated matrix was used to estimate the divergence time using BEAST v2.7 [40]. The yule tree prior was used with an optimized relaxed clock (ORC) and GTR+GAMMA was selected for the molecular clock model. We ran four independent analyses, each consisting of 400 million generations and sampled every 10,000 generations. The results from these analyses were combined using the LogCombiner package. Convergence was assessed using TRACER, based on the effective sample size (ESS) of each parameter with a likelihood > 200, following the method described in [41]). After discarding the first 25% of trees as burn-in, the maximum clade credibility (MCC) tree was generated using the TreeAnnotator package. The mean nodal heights and age estimation with 95% highest posterior densities (HPDs) were visualized using FigTree.

Twelve reliable fossils from Rosaceae were used for calibration, as described by Xiang et al. [42]. The fossil information and calibration settings are shown in Table S2. Notably, *Spiraea*-like fossils (fruits and leaves) have been found in the Early Eocene upland floras of the Okanogan Highlands of northeastern Washington State and British Columbia, Canada [42, 43]. Therefore, the stem age of *Spiraea* was estimated using a standard lognormal prior distribution with a mean of 0.25, a deviation of 0.3, and an offset of 48 Ma, which roughly matched the hypothesized age (49–50 Ma).

To examine the general patterns of diversification through time of *Spiraea*, lineage-through-time (LTT) plots were calculated using APE v3.5 [44]. To account for uncertainty in the dating estimates, the 1,000 random trees from the 4,000 converged trees from BEAST were used to calculate the 95% confidence interval (CI), and all outgroups were removed.

Selection pressure analyses

We excluded PCGs with lengths less than 300 bp. A total of 50 PCGs were chosen and aligned using the MUSCLE algorithm [45]. Stop codons were removed from the aligned sequences using MEGA 11 [46]. The resulting aligned sequences were then converted to PML format using PhyloSuite v1.2.2 [47].

To calculate the non-synonymous substitution rate (d_N), synonymous substitution rate (d_S), and ratio values (ω, d_N/d_S) of each protein-coding gene, we employed the Site Model (M0) using EasycodeML v1.4 [48]. The PCGs were categorized into seven functional groups based on the classification by Wicke et al. [49]: ATP synthase,

Zhang *et al. BMC Plant Biology*   (2024) 24:23

Page 4 of 13

NADPH dehydrogenase, cytochrome b/f complex, photosystem, ribosomal proteins, RNA polymerase, and other genes.

Likelihood ratio tests (LRT) were conducted to determine the presence/absence of selection. The log-likelihood statistic was calculated as $2\Delta L = 2(L_1 - L_0)$. This value was then compared with a χ^2 distribution with degrees of freedom calculated from the difference in the number of parameters between the models. L_0 represents the log-likelihood under the null model, while L_1 represents the likelihood under the alternative model. Model M0 assumes a single value of ω for all sites and lineages, while model M3 includes a discrete distribution of ω values. Model M1 assumes two site classes, and was compared with M2, which allows for three site classes including positive selection. Models M7 and M8 were also compared, with M7 assuming a beta distribution for variation in ω and M8 allowing for positive selection. Positive selection can only be inferred when models M2, M3, or M8 indicate codons with a ω ratio > 1 and the likelihood ratio test for positive selection is significant at $P < 0.05$. Bayesian empirical Bayes (BEB) was used to identify sites potentially under positive selection.

Results

Structural features and genome composition of plastid genomes in *Spiraea*

In this study, we analyzed the plastid genomes of 34 individuals from the *Spiraea* genus, representing 27 different Asiatic species. We obtained nine *Spiraea* plastid genome sequences from GenBank. The complete plastid genomes exhibited the typical quadripartite circular structure, consisting of a large single-copy (LSC) region, a small single-copy (SSC) region, and two inverted repeat (IR) regions (Fig. 1). The length of the genomes ranged from 153,822 bp (*S. japonica*, MZ981784) to 158,637 bp (*S. insularis*), while the LSC region length ranged from 82,226 bp (*S. japonica*) to 86,997 bp (*S. insularis*). The shortest SSC region was 26,333 bp, found in four species: *S. ovalis*, *S. henryi*, *S. nipponica*, and *S. trichocarpa*, while the longest SSC region was 26,492 bp in *S. martini*. The IR region length ranged from 18,700 bp (*S. laeta*) to 18,957 bp (*S. ovalis*). The GC content of the *Spiraea* plastid genome ranged from 36.6% to 36.9%. This information is summarized in Table S3. Linear regression analysis showed a significant correlation between genome size and the length of the SC region ($R^2 = 0.99$, $P < 0.01$) and intergenic spacer (IGS) length ($R^2 = 0.21$, $P < 0.01$), but no correlation with the IR region ($R^2 = 0.03$, $P = 0.31$) or intron length ($R^2 = 0.02$, $P = 0.44$) (Fig. 2).

The *Spiraea* plastid genomes encoded a total of 113–114 genes, including 78–79 protein-coding genes (PCGs), 30 tRNA genes, and 4 rRNA genes (Table S3).

Seventeen genes contained one intron, namely, *atpF*, *clpP*, *petB*, *petD*, *ndhA*, *ndhB*, *rpoC1*, *rpl2*, *rpl16*, *rps12*, *rps16*, *trnA-UGC*, *trnI-GAU*, *trnG-UCC*, *trnK-UUU*, *trnL-UAA*, and *trnV-UAC*. One gene, *ycf3*, contained two introns. Additionally, seventeen genes were located in the IR region, including *ndhB*, *rpl2*, *rpl23*, *rps7*, *rps12*, *rrn4.5*, *rrn5*, *rrn16*, *rrn23*, *trnA-UGC*, *trnI-CAU*, *trnI-GAU*, *trnL-CAA*, *trnN-GUU*, *trnR-ACG*, *trnV-GAC*, and *ycf2* (Table S4). The *psaA* gene was absent in the accession of *S. japonica* var. *japonica* (MZ981784).

Plastid genome variations within *Spiraea*

Our study revealed a significant level of conservation in the genome structure, gene order, and gene content of all *Spiraea* plastid genomes. Figure S1 demonstrates that the inverted repeat (IR) region exhibited relatively stable characteristics compared to the large single copy (LSC) and small single copy (SSC) regions, with strikingly higher GC content and lower nucleotide and gap variations (Fig. 3a). The conservation of all rRNA genes was observed, along with certain protein-coding genes, such as *psbA*, *atpA*, *rpoB*, *psbC*, *psbD*, *rps4*, *rbcL*, *accD*, *petA*, *psbB*, *ycf2*, *ndhD*, and *rps7*.

In the comprehensive analysis of 34 *Spiraea* species' plastid genomes, we observed that the length of the *rps19* gene ranged from 279 to 282 bp, all of which were located in the LSC-IR_B region (Fig. S2). Among them, a segment of 108 bp to 252 bp extended into the IR_B region, resulting in the duplication of the *rps19* gene in the IR_A region. The *ycf1* gene was situated in the SSC-IR_A region, with a length ranging from 1,046 bp to 1,064 bp in the IR_A region. However, the total length of the *ycf1* gene in *S. martini* was only 1,131 bp, with only 82 bp in the SSC region, while *S. veitchii* exhibited the longest length of 4,577 bp. Additionally, a duplication of the *ycf1* pseudogene was observed in the IR_B region. The position of the *ndhF* gene varied, with 22 *Spiraea* accessions harboring it in the SSC region, while it appeared in the IR_B-SSC region in the remaining accessions. The conserved *trnH-GUG* gene, spanning 74 bp to 75 bp, was distributed in the SSC-IR_A region (Fig. S2).

Sliding window analysis demonstrated that nucleotide polymorphisms in *Spiraea* plastid genomes were more conserved in the IR region than in the SC region, and non-coding regions exhibited more pronounced variation than coding regions (Fig. 3a). The nucleotide polymorphism (*Pi*) ranged from 0 to 0.0138, with an average value of 0.0036. Relatively high *Pi* values (> 0.011) were observed in *ndhF-rpl32*, *trnH-GUG-psbA*, *trnG-UCC-atpA*, *petN-psbM*, *trnT-UGU-trnL-UAA*, *rpl16*, *rpl32-trnL-UAG*, and *ycf1*.

Zhang *et al. BMC Plant Biology*   (2024) 24:23
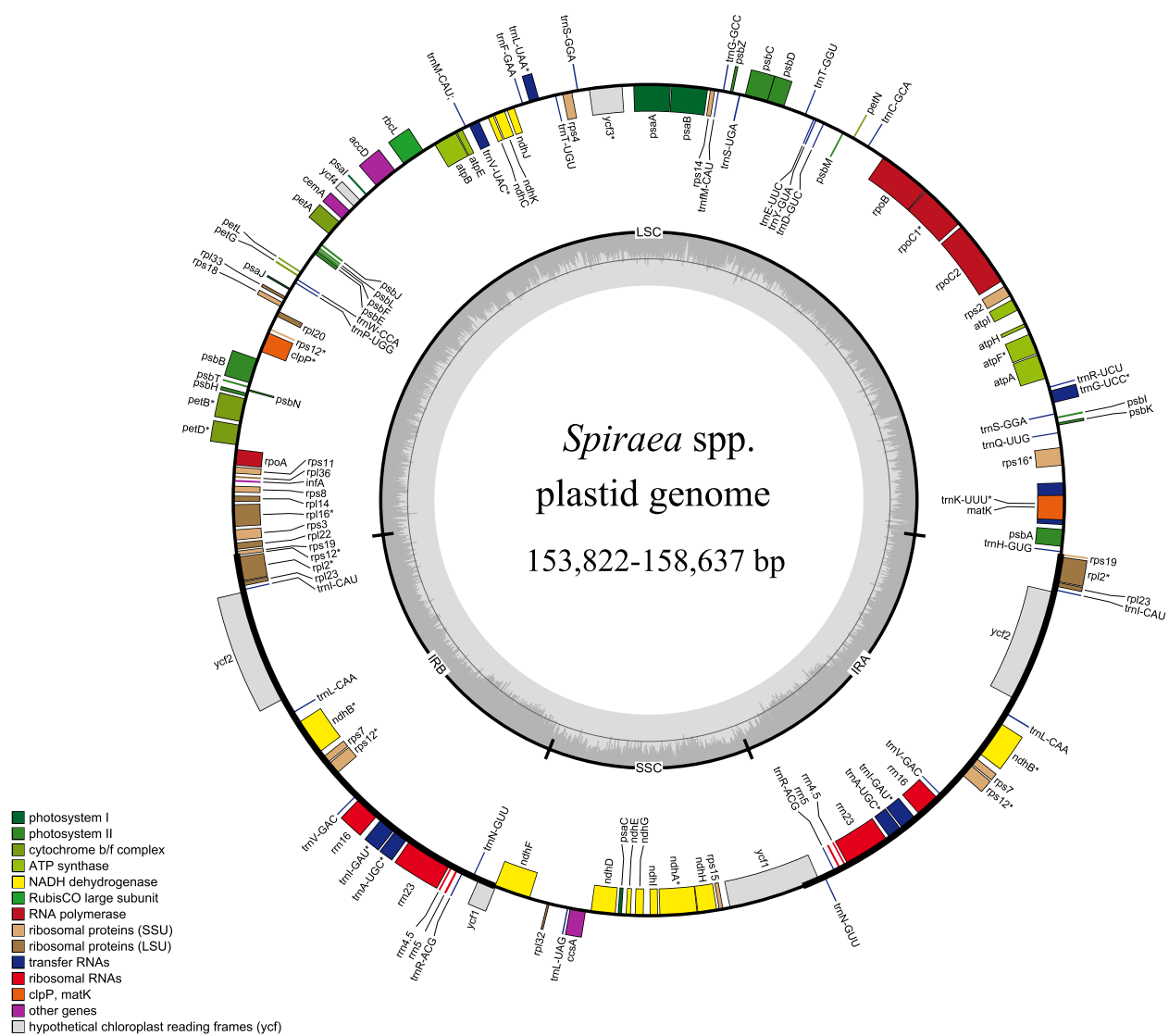
Page 5 of 13

Fig. 1 The plastid genome map of Asiatic *Spiraea* species. The genes belonging to different functional groups are shown in different colors. Genes inside and outside of the external circle are transcribed in clockwise and counterclockwise directions, respectively. The inner circle represents the quadripartite structure, with two copies of the inverted repeat (IR$_A$ and IR$_B$), an LSC, and an SSC region in black with GC content in dark gray and AT content in light gray

Selection pressure analyses

This study investigated the adaptive evolution of *Spiraea* by analyzing the ratio of d_N/d_S. A ratio greater than one indicates positive selection, a ratio less than one indicates purifying selection, and a ratio equal to one indicates neutral evolution. Analysis of 50 PCGs revealed that the average value of d_N was 0.0074, with the highest value observed for *rpl22* (0.0394) and the lowest being 0. The average value of d_S was 0.0449, with the highest value observed for *rps14* (0.1179) and the lowest values observed for *ycf2* (0.0022). The average value of d_N/d_S was 0.2122, with the highest value

observed for *ycf2* (1.7749) and the lowest being 0.0001 (Fig. 3b).

Most genes exhibited a significantly larger d_S compared to d_N, indicating purifying selection. However, the d_N/d_S ratio of the *ycf2* gene (1.7749) was greater than one, suggesting that this gene has undergone significant positive selection ($P < 0.05$; Tables S5-6). When categorizing the protein-coding genes based on their function, it was found that the other type genes had the highest d_N value (0.0127), while photosystem genes had the lowest d_N value (0.0016). In terms of d_S, ribosomal protein genes had the highest value (0.0599), while RNA polymerase

Zhang *et al. BMC Plant Biology*   (2024) 24:23
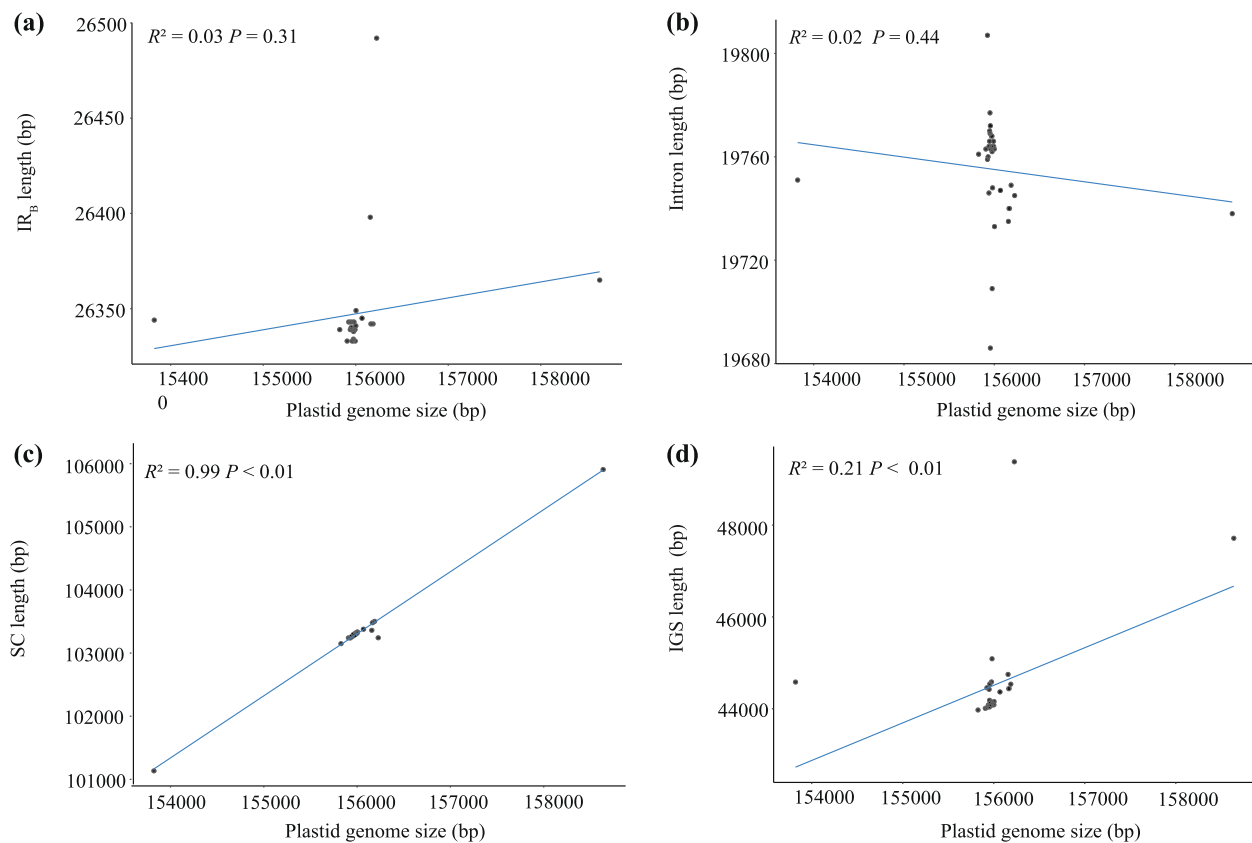
Page 6 of 13

Fig. 2 The relationships between the length of different parts of plastome and plastome size within *Spiraea*. **a** Correlation of the IR length and plastome size. **b** Correlation of the intron length and plastome size. **c** Correlation of the SC length and plastome size. **d** Correlation of the IGS length and plastome size

genes had the lowest value (0.0265). Except for the *ycf2* gene whose d_N/d_S value was greater than 1, the rest of the functional genes exhibited values less than 1. Among these groups, other type genes had the highest d_N/d_S value (0.4261), while photosystem genes had the lowest (0.0401) (Fig. 4).

Phylogenetic analyses and divergence time estimates

For this study, a total of 34 *Spiraea* accessions and three outgroups were selected to construct a maximum likelihood (ML) tree and Bayesian inference using the protein-coding genes (PCGs) and whole plastid genome sequences. The ML tree, Bayesian inference, and BEAST MCC tree based on PCGs showed similar phylogenetic topographies, as shown in Fig. 5. Five main clades were identified with high support in the ML tree (Fig. 5a). However, the whole plastid genome sequences only recovered four of these main clades with full support values (Fig. S3). Clade II, identified by PCGs, could not be recovered, and its three branches within the clade were not strongly supported (Fig. S3). Overall, the traditionally defined subgenera and sections of *Spiraea* could not

be supported by any datasets or phylogenetic methods (Fig. 5a-b; Fig. S3; Table S1).

To estimate the divergence date, the top-ten clock-like genes were filtered out using Sortadate. These genes were *accD, atpA, matK, ndhA, ndhD, ndhF, ccsA, rpoB, rpoC2,* and *ycf1*. Twelve reliable fossils from Rosaceae were used. The stem age of *Spiraea* was estimated to be 49.24 Ma (95% HPD: 48.64–50.09 Ma), with diversification occurring since the Early Oligocene (~30 Ma, Fig. S4). Most extant *Spiraea* lineages (29 out of 34) originated from the Pliocene period, and 25 taxa diverged from their most recent common ancestor (MRCA) since the Pleistocene (Fig. 5b). The LTT plot showed a clear shift in lineages since approximately 5 Ma, followed by a rapid accumulation after approximately 2 Ma (Fig. 5c).

Discussion

Plastid genome features and molecular variation patterns of *Spiraea*

The plastid is a vital organelle in plant cells, playing a crucial role in plant growth and development [50]. While significant changes in plastome organization

(a)

(b)

Fig. 3 Basic features of *Spireae* plastid genomes. **a** The nucleotide diversity (*Pi*), GC content, and gap proportion in the plastid genome sequences of *Spiraea* detected by sliding window analysis, using 600 bp windows and a 200 bp step size. **b** Nonsynonymous (d_N) and synonymous (d_S) substitution rates and d_N/d_S for each protein-coding gene within the *Spiraea* plastid genomes

have been observed in various distantly related plant groups, such as conifers, the IR-lacking clade (IRLC) of Fabaceae, and some parasitic flowering plants [51], the plastid genome tends to remain relatively stable across major plant lineages. It typically consists of

a quadripartite structure, with two copies of a large inverted repeat (IR) separated by small (SSC) and large (LSC) single-copy regions [51]. The typical plastome organization is also found in *Spiraea* (Fig. 1; Table S1). In our study, we examined the plastid genome of
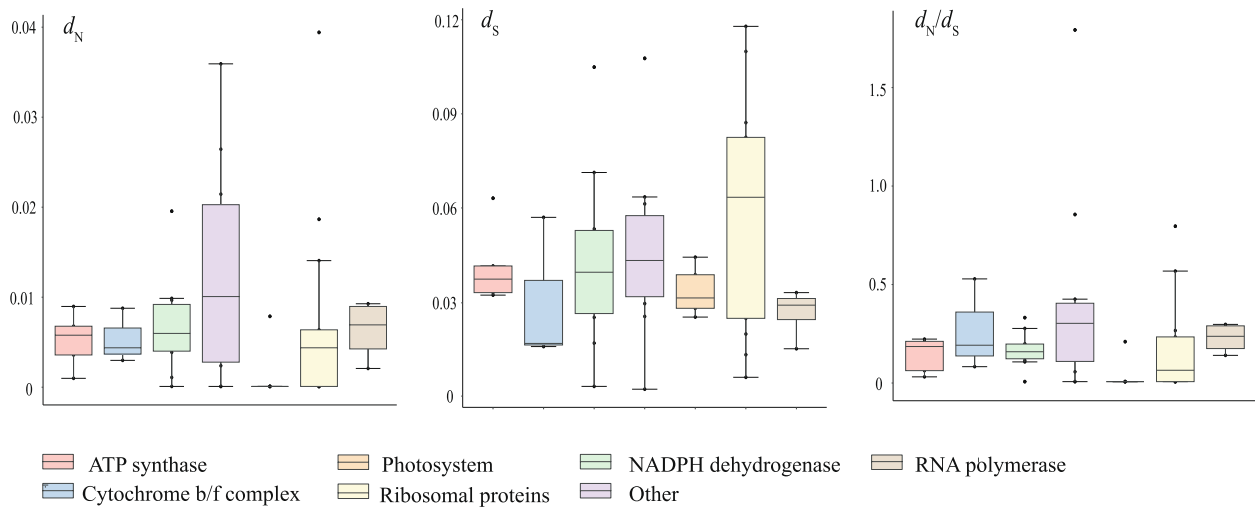
Fig. 4 Boxplot of the values of the nonsynonymous (d_N) and synonymous (d_S) substitution rates and d_N/d_S in each functional gene group
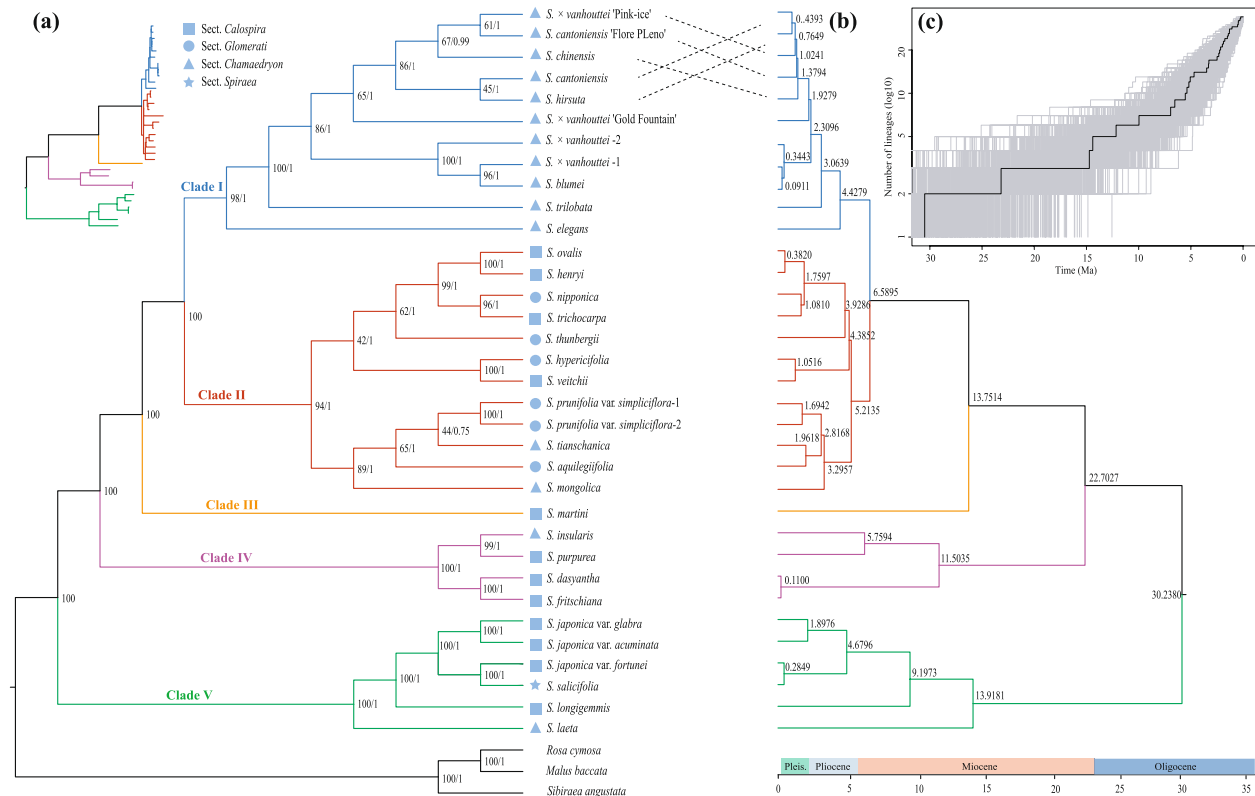

Fig. 5 Phylogeny, divergence time estimate, and LTT plot for *Spiraea*. **a** Cladogram of the maximum likelihood (ML) phylogenetic tree based on 79 protein-coding genes (PCGs). All main clades in both trees (ML and BI) are identical. Node labels are marked with the bootstrap values and Bayesian posterior probabilities (bootstrap value/posterior probability). The upper left shows the ML tree with branch length. **b** Maximum clade-credibility tree (MCC tree) from BEAST. Node labels represent estimated divergence times in millions of years. The confidence intervals are provided in Fig. S4. **c** Lineage-through-time plot for *Spiraea*

Spiraea and identified several variation patterns within these genomes.

First, we observed a positive relationship between the size of SC region and plastid genome size. While expansion/contraction of the IR regions is one major driver of genome size variation [51], this was not evident in *Spiraea* ($R^2 = 0.03$, $P = 0.31$; Fig. 2a). However, there was a significant correlation between the plastid genome size and the length of the SC region ($R^2 = 0.99$, $P < 0.01$; Fig. 2c). The gene content across all *Spiraea* plastid genomes was highly conserved, except for *Spiraea japonica* var. *japonica* (accession no. MZ981784), which had the smallest plastid genome within *Spiraea* and lacked an important photosynthesis-related gene, *psaA*. Additionally, the intergenic spacer (IGS) in the SC region may contribute to the size variation of the SC region in *Spiraea* (Fig. 2d).

Second, we identified several hypervariable hotspots within the *Spiraea* plastome, all of which were localized in the SC regions (Fig. 3a). The pattern of lower IR substitution rates compared to the SC regions has been observed in *Spiraea* plastid genomes (Fig. 3a; Fig. S1), as well as in other plant lineages [51]. This decreased substitution rate in the IR is likely a result of a copy-dependent repair mechanism [52, 53].

Previous studies have used nrDNA ITS sequence data and two plastid DNA loci (*trnL-trnF* and *matK*) to investigate the phylogeny of *Spiraea* [6–9]. However, these methods have shown limited resolution in resolving the phylogeny of *Spiraea* [6–9], indicating low sequence divergence. Our results partially support this hypothesis, as *trnL-trnF* and *matK* exhibit relatively low sequence divergence (each with *Pi* values of 0.0043 and 0.0083) compared to highly variable hotspots (such as hotspots with *Pi* values > 0.011) (Fig. 3a). Therefore, hypervariable loci are needed to increase the phylogenetic resolution in low-level phylogenetic or phylogeographic studies in *Spiraea*.

We conducted an investigation into the hypervariable hotspots across the plastid genomes within the *Spiraea* genus. These findings provide valuable genetic resource information for future phylogenetic and phylogeographic studies. We identified eight top hypervariable loci with *Pi* values larger than 0.011, namely, *ndhF-rpl32*, *trnH-GUG-psbA*, *trnG-UCC-atpA*, *petN-psbM*, *trnT-UGU-trnL-UAA*, *rpl16*, *rpl32-trnL-UAG*, and *ycf1*.

Many of these loci, such as *ndhF-rpl32*, *trnH-GUG-psbA*, *trnT-UGU-trnL-UAA rpl16*, *rpl32ni-trnL-UAG*, and *ycf1*, have been previously identified in studies conducted by Taberlet et al. [54], Shaw et al. [55–57], and Dong et al. [58]. However, it is important to consider the characteristics of these variable regions themselves. Factors such as fragment size and mutation rate (substitution saturation) can have a detrimental effect on phylogenetic analysis. For example, the *trnH-psbA* locus is controversial due to its high mutation rate and significant length variation caused by insertions, deletions, simple sequence repeats, and localized inversions in some plant lineages [59–61]. These factors can impact the accuracy of sequence alignment and subsequently affect the precision of phylogenetic inference and species identification. Within *Spiraea*, the *trnH-psbA* alignment matrix has 89 indels, indicating its unstable size. Additionally, the *ycf1* locus also exhibits a high degree of sequence variability and has been proposed as a candidate barcode for land plants. However, its high mutation rate and absence in certain plant lineages, such as Poaceae [58, 62], can lead to alignment ambiguities similar to those observed in *trnH-psbA* [58, 63].

Third, we investigated the pattern of molecular evolution among plastid genes in *Spiraea*. The evolution of substitution rates among plastid genes has been extensively studied with the recent release of numerous plastid genomes. Generally, non-synonymous rates of photosynthesis-related genes differ significantly from those of housekeeping genes, indicating that photosynthesis-related genes are under stronger functional constraints [64–66]. Our results demonstrated that the photosynthesis-related genes in terms of d_N are highly conserved, providing further support for this hypothesis (Fig. 4).

It is generally assumed that most plant species have low substitution rates of most genes in organelles, and these genes evolve under strong purifying selection. However, certain genes in the plastid have undergone positive selection, particularly those involved in photosynthesis and other metabolic pathways (e.g., [67]). The relationship between the high levels of non-synonymous rate evolution and potential selective factors, such as novel ecological conditions, is still a topic of debate [68]. In our study, we observed d_N/d_S ratios of 1.7749 for the *ycf2* gene, which is likely experiencing strong positive selection ($P < 0.05$; Table S6). Despite its unclear function, *ycf2* is known as the largest chloroplast gene in angiosperms and encodes a motor protein that generates ATP for inner membrane translocation [69] and plant cell survival [70]. Positive selection of the *ycf2* gene has been observed in numerous species (e.g., [71–74]), indicating its adaptive importance in various environments and plant lineages.

Phylogenetic and evolutionary implications of plastid genomes within *Spiraea*

The inheritance of plastid genomes in plants is typically uniparental, with the mother being the primary contributor. This results in a smaller effective population size compared to nuclear genomes. Consequently, the coalescent time of plastid haplotypes is shorter, making them

Zhang *et al. BMC Plant Biology*   (2024) 24:23

Page 10 of 13

ideal for studying genetic variations and tracing plastid evolutionary history [75]. Plastid markers, such as *rbcL*, *matK*, and *trnL-trnF*, have been widely used in phylogenetic studies at different taxonomic levels, including ordinal, familial, tribal, and generic levels [12, 76]. They have also been used in lower-level studies, such as phylogeographic investigations [77, 78] and species identification [79, 80].

Previous studies on the phylogeny of *Spiraea* have utilized nrDNA ITS sequence data and limited plastid DNA loci (*trnL-trnF* and *matK*) [6–9]. However, the resolution of the phylogenetic relationships within *Spiraea* based on ITS and *trnL-trnF* is limited, and none of the sections defined by Rehder [3] and Yü and Lu [81] based on inflorescence morphology are monophyletic [8]. Yu et al. [9] largely supported this hypothesis, except for sect. *Spiraea*. The inclusion of *Spiraea douglasii* in sect. *Spiraea* is questionable and represents a clear difference between the two aforementioned studies [8, 9]. The *Spiraea* species sampled in our study represent all four sections found in China (Table S1; [1, 81]). Our phylogenomic analysis provides the most detailed phylogenetic backbone of *Spiraea* to date and strongly supports the notion that the traditionally defined sections of *Spiraea* are not monophyletic groups [8, 9], despite only including one species (*S. salicifolia*) from sect. *Spiraea* in this study. We also found a complex evolutionary history of *Spiraea*, as there are many shallow relationships within Clades I and II that could not be highly supported (Fig. 5; Fig. S3). Additionally, some taxa with multiple accessions (such as *S. japonica*, *S. cantoniensis*, and *S.×vanhouttei*) are recovered as paraphyletic or polyphyletic groups (Fig. 5; Fig. S3).

There may be two primary reasons why the traditional taxonomic classification and some species boundaries within *Spiraea* cannot be supported by plastid data. First, the phylogenetic relationships inferred from plastid data could not fully capture the evolutionary relationships among taxa, potentially due to the occurrence of pervasive hybridization/introgression events in *Spiraea*. Recent evidence suggests that ancient hybridization/introgression events may have resulted in cytonuclear conflicts [82]. *Spiraea* is a genus known to frequently undergo hybridization, with numerous hybrids arising through horticultural practices [83]. Although wild hybrids are less commonly reported, there have been instances of hybrids occurring across different sections of the genus [83, 84]. These findings indicate that reproductive isolation is not complete among *Spiraea* species, making hybridization a potentially common phenomenon. Consequently, a substantial amount of nuclear genetic data is necessary to further investigate the traditional classification framework and hybridization/introgression events.

Second, more detailed revisions of the infrageneric classification and species boundary delimitation of *Spiraea* are needed. It is evident that the traditional classification system within *Spiraea* is inconsistent with the results obtained from plastid data ([23]; this study), nuclear ribosomal DNA internal transcribed spacer (nrDNA ITS) sequences [6, 7], and their combined analyses [8, 9]. This inconsistency strongly suggests a problem with the traditional infrageneric classification of *Spiraea*. Traditionally, inflorescence morphology has been the main focus in these classification schemes [3, 81], but it is likely that this trait does not accurately reflect the true evolutionary relationships within *Spiraea*. Therefore, an integrative framework for infrageneric classification in the genus is needed, which should incorporate additional diagnostic traits such as seed and palynology evidence [9].

Ancient *Spiraea* taxa are believed to have originated in the Early Eocene based on macrofossil evidence [42, 43]. Our study is the first to identify the divergence sequence of the Asian *Spiraea* taxa using twelve reliable fossil calibrations of Rosaceae. We acknowledge that the crown age inferred in this study may have been underestimated due to our limited sampling. Nevertheless, our divergence time for the genus is much older than the result of *Spiraea* taxa in the Qinghai-Tibetan Plateau (QTP) [6], which can be attributed to the difference in molecular markers, divergence estimate methods, and the species included in the two studies ([6]; this study). We conclude that *Spiraea* diversified after the Early Oligocene (~ 30 Ma), followed by a rapid speciation process during the Pliocene and Pleistocene periods (Fig. 5b-c). The Pliocene–Pleistocene period played a vital role in the speciation process of *Spiraea*, as most extant taxa used in this study originated during this period (Fig. 5b-c), which is supported by fossil records from the Pliocene–Pleistocene period in Asia (data from PBDB: https://paleobiodb.org/#/). The pattern of rapid radiation during this period (especially the Pleistocene) has also been observed in many plant groups in East Asia [85, 86], likely due to the effects of glaciation and climatic oscillations during the Pleistocene. East Asia, being less influenced by Quaternary glaciations, had a lower extinction rate or higher speciation rate during those times, leading to a relatively high net diversification rate [87–90]. Biotic factors, such as hybridization/introgression, possibly related to climate oscillations, have also contributed to the emergence of young species during the Pleistocene period. In theory, climate oscillations during the period may have favored introgressive hybridization/introgression upon secondary contact following initial species divergence [91, 92]. Ancestral polymorphism of plastomes could have been purified through the extent of chloroplast-capture events

Zhang *et al. BMC Plant Biology*   (2024) 24:23

Page 11 of 13

when hybridization/introgression frequently occurred [93], obscuring species boundaries and creating the illusion of young species pairs. Further studies, including nuclear sequences and more samplings, are needed to investigate this hypothesis.

Conclusions

In this study, we sequenced, assembled, and analyzed the plastomes of 34 Asiatic *Spiraea* accessions. The *Spiraea* plastid genome exhibits typical quadripartite structures and encodes 113–114 genes, including 78–79 protein-coding genes (PCGs), 30 tRNA genes, and 4 rRNA genes. There is a significant correlation between genome size and the length of the SC region. Several hypervariable regions within the *Spiraea* plastome were identified, all of which were localized in the SC regions. Our phylogenomic analysis successfully provides a phylogenetic framework for *Spiraea* but does not support its currently defined section boundaries. Additionally, we discovered that the genus underwent diversification after the Early Oligocene (~ 30 Ma), followed by a rapid speciation process during the Pliocene and Pleistocene periods. Although not perfect, the plastomes of *Spiraea* provide invaluable insights into its phylogenetic relationships and evolutionary history. Further investigations utilizing other genomes, such as the nuclear genome, are urgently needed to enhance our understanding of the evolutionary history of the genus.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12870-023-04697-8.

Additional file 1: Fig. S1. Sequence identity plots of plastomes of *Spiraea* by mVISTA. The top line shows the orientation of genes. A cutoff of 70% identity was used for the plots, and the Y-scale represents the percentage identity ranging from 50 to 100%. **Fig. S2.** Comparison of LSC, IRs, and SSC junction positions among *Spiraea* plastomes. **Fig. S3.** Cladogram of the maximum likelihood (ML) phylogenetic tree constructed based on the whole plastid sequences. Node labels represent the bootstrap values (left) and Bayesian posterior probabilities (right), respectively. **Fig. S4.** Maximum clade-credibility tree of Rosaceae from BEAST with 12 fossil calibration points. The bar at each node indicates age with a 95% height posterior distribution. Red stars represent fossil calibration points (see Table S2). **Table S1.** The sampling information in this study. **Table S2.** Fossil calibration points used in this study. **Table S3.** Plastid genome information of *Spiraea*. **Table S4.** List of annotated genes in the plastid genome of *Spiraea*. **Table S5.** The value of nonsynonymous (d_N), synonymous (d_S) substitution rate and d_N/d_S in each functional gene. **Table S6.** The likelihood ratio and positively selected codon site tests in this study.

References

1. Lu LT, Alexander C. Spiraea. In: Wu CY, Raven P, editors. Flora of China (vol 9). St. Louis: Missouri Botanical Garden Press; 2003. p. 47–73.
2. Poyarkova AI. Spiraeoideae. Moscow: Izdatel-stvo Akademii Nauk SSSR; 1939.
3. Rehder A. Manual of cultivated trees and shrubs. Portland: Dioscorides Press; 1940.
4. Yü T, Kuan K. Taxa nova *Rosacearum sinicarum* (1). Acta Phytotax Sin. 1963;8(3):202–34.
5. Huh M. Phylogenetic relationships in the genus *Spiraea* (Rosaceae) inferred from the chloroplast DNA region, *trnL-trnF*. Ame J Plant Sci. 2012;2012(3):559–66.
6. Khan G, Zhang FQ, Gao QB, Fu PC, Xing R, Wang JL, Liu HR, Chen SL. Phylogenetic analyses of *Spiraea* (Rosaceae) distributed in the Qinghai-Tibetan plateau and adjacent regions: Insights from molecular data. Plant Syst Evol. 2016;302(1):11–21.
7. Poliakova TA, Shatokhina AV, Politov DV. Molecular phylogeny of Russian species of the genus *Spiraea* (Rosaceae) according to the nucleotide variability of the its nuclear rDNA region. Russ J Genet. 2022;58(11):1297–305.
8. Potter D, Eriksson T, Evans RC, Oh S, Smedmark JEE, Morgan DR, Kerr M, Robertson KR, Arsenault M, Dickinson TA, et al. Phylogeny and classification of Rosaceae. Plant Syst Evol. 2007;266(1–2):5–43.
9. Yu SX, Gadagkar SR, Potter D, Xu DX, Zhang M, Li ZY. Phylogeny of *Spiraea* (Rosaceae) based on plastid and nuclear molecular data: Implications for morphological character evolution and systematics. Perspect Plant Ecol. 2018;34:109–19.
10. Huh M. Analysis of the phylogenetic relationships in the genus *Spiraea* based on the nuclear ribosomal DNA its region. J Life Sci. 2012;22:285–92.
11. Howe CJ, Barbrook AC, Koumandou VL, Nisbet RER, Symington HA, Wightman TF. Evolution of the chloroplast genome. Philos T Roy Soc B. 2003;358(1429):99–106.
12. Gitzendanner MA, Soltis PS, Yi TS, Li DZ, Soltis DE. Plastome phylogenetics: 30 years of inferences into plant evolution. Adv Bot Res. 2018;85:293–313.

Zhang *et al. BMC Plant Biology*   (2024) 24:23

Page 12 of 13

13. Bock R. Structure, function, and inheritance of plastid genomes. In: Bock R, editor. Cell and molecular biology of plastids. Berlin, Heidelberg: Springer Berlin Heidelberg; 2007. p. 29–63.

14. Zhang SD, Jin JJ, Chen SY, Chase MW, Soltis DE, Li HT, Yang JB, Li DZ, Yi TS. Diversification of Rosaceae since the Late Cretaceous based on plastid phylogenomics. New Phytol. 2017;214(3):1355–67.

15. Li C, Cai C, Tao Y, Sun Z, Jiang M, Chen L, Li J. Variation and evolution of the whole chloroplast genomes of *Fragaria* spp. (Rosaceae). Front Plant Sci. 2021;12:754209.

16. Sun J, Sun R, Liu HB, Chang LL, Li ST, Zhao MZ, Shennan C, Lei JJ, Dong J, Zhong CAF, et al. Complete chloroplast genome sequencing of ten wild *Fragaria* species in china provides evidence for phylogenetic evolution of *Fragaria*. Genomics. 2021;113(3):1170–9.

17. Yu JJ, Fu J, Fang YP, Xiang J, Dong HJ. Complete chloroplast genomes of *Rubus* species (Rosaceae) and comparative analysis within the genus. BMC Genomics. 2022;23:32.

18. Gao C, Li T, Zhao X, Wu C, Zhang Q, Zhao X, Wu M, Lian Y, Li Z. Comparative analysis of the chloroplast genomes of *Rosa* species and RNA editing analysis. BMC Plant Biol. 2023;23:318.

19. Zhang C, Li SQ, Xie HH, Liu JQ, Gao XF. Comparative plastid genome analyses of *Rosa*: insights into the phylogeny and gene divergence. Tree Genet Genomes. 2022;18:20.

20. Choi J, Kim W, Park JY, Kang JS, Yang TJ. The complete plastid genome and nuclear ribosomal transcription unit sequences of *Spiraea prunifolia* f. *simpliciflora* (Rosaceae). Korean J Plant Taxon. 2023;53(1):32–7.

21. Shen W, Lin J, Lin H. The chloroplast genome of *Spiraea thunbergii* (Rosaceae). Mitochondrial DNA Part B. 2022;7(10):1879–81.

22. Yang J, Kang GH, Pak JH, Kim SC. Characterization and comparison of two complete plastomes of Rosaceae species (*Potentilla dickinsii* var. *glabrata* and *Spiraea insularis*) endemic to Ulleung island, Korea. Int J Mol Sci. 2020;21(14):4933.

23. Zhang SD, Yan K, Ling LZ. Characterization and phylogenetic analyses of ten complete plastomes of *Spiraea* species. BMC Genomics. 2023;24(1):137.

24. Doyle JJ, Doyle JL. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem Bull. 1987;19:11–5.

25. Jin JJ, Yu WB, Yang JB, Song Y, dePamphilis CW, Yi TS, Li DZ. Getorganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. Genome Biol. 2020;21:241.

26. Qu XJ, Moore MJ, Li DZ, Yi TS. PGA: a software package for rapid, accurate, and flexible batch annotation of plastomes. Plant Methods. 2019;15:50.

27. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, et al. Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics. 2012;28(12):1647–9.

28. Lohse M, Drechsel O, Bock R. Organellargenomedraw (OGDRAW): a tool for the easy generation of high-quality custom graphical maps of plastid and mitochondrial genomes. Curr Genet. 2007;52(5–6):267–74.

29. R Core Team. R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing; 2022.

30. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013;30(4):772–80.

31. Rozas J, Ferrer-Mata A, Sanchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE, Sanchez-Gracia A. DnaSP 6: DNA sequence polymorphism analysis of large data sets. Mol Biol Evol. 2017;34(12):3299–302.

32. Frazer KA, Pachter L, Poliakov A, Rubin EM, Dubchak I. VISTA: computational tools for comparative genomics. Nucleic Acids Res. 2004;32:W273–9.

33. Borowiec ML. AMAS: A fast tool for alignment manipulation and computing of summary statistics. PeerJ. 2016;4:e1660.

34. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014;30(9):1312–3.

35. Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP. Mrbayes 3.2: Efficient bayesian phylogenetic inference and model choice across a large model space. Syst Biol. 2012;61(3):539–42.

36. Miller MA, Pfeiffer W, Schwartz T. Creating the cipres science gateway for inference of large phylogenetic trees. In: 2010 gateway computing environments workshop (GCE). New Orleans: IEEE; 2010. p. 1–8.

37. Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. Posterior summarization in bayesian phylogenetics using TRACER 1.7. Syst Biol. 2018;67(5):901–4.

38. Rambaut A. Figtree v1.4.3. 2009. [[WWW document] URL http://tree.bio.ed.ac.uk/software/figtree]. Accessed 5 Feb 2023.

39. Smith SA, Brown JW, Walker JF. So many genes, so little time: a practical approach to divergence-time estimation in the genomic era. PLoS ONE. 2018;13(5):e0197433.

40. Bouckaert R, Vaughan TG, Barido-Sottani J, Duchêne S, Fourment M, Gavryushkina A, Heled J, Jones G, Kühnert D, De Maio N, et al. BEAST 2.5: an advanced software platform for bayesian evolutionary analysis. PLOS Comput Biol. 2019;15:e1006650.

41. Drummond AJ, Ho SYW, Phillips MJ, Rambaut A. Relaxed phylogenetics and dating with confidence. PLOS Biol. 2006;4:e88.

42. Xiang Y, Huang CH, Hu Y, Wen J, Li S, Yi T, Chen H, Xiang J, Ma H. Evolution of rosaceae fruit types based on nuclear phylogeny in the context of geological times and genome duplication. Mol Biol Evol. 2017;34(2):262–81.

43. DeVore ML, Pigg KB. A brief review of the fossil history of the family Rosaceae with a focus on the Eocene Okanogan highlands of eastern Washington state, USA, and British Columbia. Canada Plant Syst Evol. 2007;266(1–2):45–57.

44. Paradis E, Claude J, Strimmer K. APE: analyses of phylogenetics and evolution in R language. Bioinformatics. 2004;20(2):289–90.

45. Edgar RC. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics. 2004;5:113.

46. Tamura K, Stecher G, Kumar S. MEGA11 molecular evolutionary genetics analysis version 11. Mol Biol Evol. 2021;38(7):3022–7.

47. Zhang D, Gao FL, Jakovlic I, Zou H, Zhang J, Li WX, Wang GT. Phylosuite: an integrated and scalable desktop platform for streamlined molecular sequence data management and evolutionary phylogenetics studies. Mol Ecol Resour. 2020;20(1):348–55.

48. Gao FL, Chen CJ, Arab DA, Du ZG, He YH, Ho SYW. EasycodeML: a visual tool for analysis of selection using codeml. Ecol Evol. 2019;9(7):3891–8.

49. Wicke S, Schneeweiss GM, dePamphilis CW, Muller KF, Quandt D. The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. Plant Mol Biol. 2011;76(3–5):273–97.

50. Gould SB, Waller RF, McFadden GI. Plastid evolution. Annu Rev Plant Biol. 2008;59(1):491–517.

51. Jansen RK, Ruhlman TA. Plastid genomes of seed plants. In: Bock R, Knoop V, editors. Genomics of chloroplasts and mitochondria. Dordrecht: Springer Netherlands; 2012. p. 103–26.

52. Perry AS, Wolfe KH. Nucleotide substitution rates in legume chloroplast DNA depend on the presence of the inverted repeat. J Mol Evol. 2002;55(5):501–8.

53. Zhu AD, Guo WH, Gupta S, Fan WS, Mower JP. Evolutionary dynamics of the plastid inverted repeat: the effects of expansion, contraction, and loss on substitution rates. New Phytol. 2016;209(4):1747–56.

54. Taberlet P, Gielly L, Pautou G, Bouvet J. Universal primers for amplification of three non-coding regions of chloroplast DNA. Plant Mol Biol. 1991;17(5):1105–9.

55. Shaw J, Lickey EB, Beck JT, Farmer SB, Liu WS, Miller J, Siripun KC, Winder CT, Schilling EE, Small RL. The tortoise and the hare II: relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. Am J Bot. 2005;92(1):142–66.

56. 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.

57. Shaw J, Shafer HL, Leonard OR, Kovach MJ, Schorr M, Morris AB. Chloroplast DNA sequence utility for the lowest phylogenetic and phylogeographic inferences in angiosperms: the tortoise and the hare IV. Am J Bot. 2014;101(11):1987–2004.

58. Dong WP, Xu C, Li CH, Sun JH, Zuo YJ, Shi S, Cheng T, Guo JJ, Zhou SL. *Ycf1*, the most promising plastid DNA barcode of land plants. Sci Rep. 2015;5:8348.

59. Kress WJ, Erickson DL. A two-locus global DNA barcode for land plants: the coding *rbcL* gene complements the non-coding *trnH-psbA* spacer region. PLoS ONE. 2007;2(6):e508.

60. Pang XH, Liu C, Shi LC, Liu R, Liang D, Li H, Cherny SS, Chen SL. Utility of the *trnH-psbA* intergenic spacer region and its combinations as plant DNA barcodes: a meta-analysis. PLoS ONE. 2012;7(11):e48833.

Zhang *et al. BMC Plant Biology*   (2024) 24:23

Page 13 of 13

61. Whitlock BA, Hale AM, Groff PA. Intraspecific inversions pose a challenge for the trnh-psba plant DNA barcode. PLoS ONE. 2010;5(7):e11533.

62. Goremykin VV, Holland B, Hirsch-Ernst KI, Hellwig FH. Analysis of acorus calamus chloroplast genome and its phylogenetic implications. Mol Biol Evol. 2005;22(9):1813–22.

63. Jansen RK, Cai Z, Raubeson LA, Daniell H, Depamphilis CW, Leebens-Mack J, Muller KF, Guisinger-Bellian M, Haberle RC, Hansen AK, et al. Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns. P Natl Acad Sci USA. 2007;104(49):19369–74.

64. Wicke S, Schneeweiss GM. Next-generation organellar genomics: potentials and pitfalls of high-throughput technologies for molecular evolutionary studies and plant systematics. In: Hörandl E, Appelhans M, editors. Next-generation sequencing in plant systematics. 676. Oberreifenberg: Koeltz Botanical Books; 2015.

65. Wu CS, Wang TJ, Wu CW, Wang YN, Chaw SM. Plastome evolution in the sole hemiparasitic genus *Laurel dodder* (Cassytha) and insights into the plastid phylogenomics of Lauraceae. Genome Biol Evol. 2017;9(10):2604–14.

66. Yan XK, Liu TJ, Yuan X, Xu Y, Yan HF, Hao G. Chloroplast genomes and comparative analyses among thirteen taxa within Myrsinaceae s.str. clade (Myrsinoideae, Primulaceae). Int J Mol Sci. 2019;20:4534.

67. Silva SR, Miranda VFO, Michael TP, Płachno BJ, Matos RG, Adamec L, Pond SLK, Lucaci AG, Pinheiro DG, Varani AM. The phylogenomics and evolutionary dynamics of the organellar genomes in *Carnivorous utricularia* and *Genlisea* species (Lentibulariaceae). Mol Phylogenet Evol. 2023;181:107711.

68. Ye WQ, Yap ZY, Li P, Comes HP, Qiu YX. Plastome organization, genome-based phylogeny and evolution of plastid genes in podophylloideae (berberidaceae). Mol Phylogenet Evol. 2018;127:978–87.

69. Kikuchi S, Asakura Y, Imai M, Nakahira Y, Kotani Y, Hashiguchi Y, Nakai Y, Takafuji K, Bedard J, Hirabayashi-Ishioka Y, et al. A *ycf2*-Ftshi heteromeric AAA-Atpase complex is required for chloroplast protein import. Plant Cell. 2018;30(11):2677–703.

70. Drescher A, Ruf S, Calsa T, Carrer H, Bock R. The two largest chloroplast genome-encoded open reading frames of higher plants are essential genes. Plant J. 2000;22(2):97–104.

71. Chen J, Zang Y, Shang S, Liang S, Zhu ML, Wang Y, Tang XX. Comparative chloroplast genomes of zosteraceae species provide adaptive evolution insights into seagrass. Front Plant Sci. 2021;12:741152.

72. Chen YQ, Zhong H, Zhu YT, Huang YZ, Wu SS, Liu ZJ, Lan SR, Zhai JW. Plastome structure and adaptive evolution of Calanthe s.l. species. PeerJ. 2020;8:e10051.

73. Fan WB, Wu Y, Yang J, Shahzad K, Li ZH. Comparative chloroplast genomics of Dipsacales species: Insights into sequence variation, adaptive evolution, and phylogenetic relationships. Front Plant Sci. 2018;9:689.

74. Yang Q, Fu GF, Wu ZQ, Li L, Zhao JL, Li QJ. Chloroplast genome evolution in four montane Zingiberaceae taxa in China. Front Plant Sci. 2022;12:774482.

75. Petit RJ, Vendramin GG. Plant phylogeography based on organelle genes: an introduction. In: Weiss S, Ferrand N, editors. Phylogeography of southern European refugia: evolutionary perspectives on the origins and conservation of European biodiversity. Dordrecht: Springer; 2007.

76. Soltis D, Soltis P, Endress P, Chase MW, Manchester S, Judd W, Majure L, Mavrodiev EJC. Phylogeny and evolution of the angiosperms: revised and. updated. Chicago: The University of Chicago Press; 2018.

77. Schaal BA, Hayworth DA, Olsen KM, Rauscher JT, Smith WA. Phylogeographic studies in plants: problems and prospects. Mol Ecol. 1998;7(4):465–74.

78. Soltis DE, Morris AB, McLachlan JS, Manos PS, Soltis PS. Comparative phylogeography of unglaciated eastern North America. Mol Ecol. 2006;15(14):4261–93.

79. Hollingsworth PM, Forrest LL, Spouge JL, Hajibabaei M, Ratnasingham S, van der Bank M, Chase MW, Cowan RS, Erickson DL, Fazekas AJ, et al. A DNA barcode for land plants. P Natl Acad Sci USA. 2009;106(31):12794–7.

80. Hollingsworth PM, Li DZ, van der Bank M, Twyford AD. Telling plant species apart with DNA: from barcodes to genomes. Philos T R Soc B. 2016;371:20150338.

81. Yü TT, Lu LD. Spiraea. In: Yü TT, editor. Flora reipublicae Popularis Sinicae, vol. 36. Beijing: Science Press; 1974. p. 1–66.

82. Rose JP, Toledo CAP, Lemmon EM, Lemmon AR, Sytsma KJ. Out of sight, out of mind: Widespread nuclear and plastid-nuclear discordance in the flowering plant genus *Polemonium* (Polemoniaceae) suggests widespread historical gene flow despite limited nuclear signal. Syst Biol. 2021;70(1):162–80.

83. Zaveska Drabkova L, Pospiskova M, Businsky R. Phylogeny and infrageneric delimitation in *Spiraea* (Rosaceae) inferred from AFLP markers and a comparison with morphology. Bot J Linn Soc. 2017;185(4):525–41.

84. Businsky R. Transitive inflorescence types in *Spiraea* (Rosaceae-Spiraeoideae) undermine the fundamental classification concept of the genus. Phyton. 2015;55(1):69–120.

85. Xing YW, Ree RH. Uplift-driven diversification in the Hengduan Mountains, a temperate biodiversity hotspot. P Natl Acad Sci USA. 2017;114(17):E3444–51.

86. Zhao N, Park S, Zhang YQ, Nie ZL, Ge XJ, Kim S, Yan HF. Fingerprints of climatic changes through the late Cenozoic in southern Asian flora: Magnolia section *Michelia* (Magnoliaceae). Ann Bot. 2022;130(1):41–52.

87. Adams J. Species richness: patterns in the diversity of life. Heidelberg: Springer Berlin; 2009.

88. Axelrod D, Al-Shehbaz I, Raven P. History of the modern flora of China. In: Zhang AL, Wu SG, editors. Floristic characteristics and diversity of East Asian plants. Beijing: China: Higher Education Press; 1998. p. 43–55.

89. Qian H, Ricklefs RE. Large-scale processes and the Asian bias in species diversity of temperate plants. Nature. 2000;407(6801):180–2.

90. Yan HF, Zhang CY, Anderberg AA, Hao G, Ge XJ, Wiens JJ. What explains high plant richness in East Asia? Time and diversification in the tribe Lysimachieae (primulaceae). New Phytol. 2018;219(1):436–48.

91. Taylor SA, Larson EL, Harrison RG. Hybrid zones: Windows on climate change. Trends Ecol Evol. 2015;30(7):398–406.

92. Walter RP, Roy D, Hussey NE, Stelbrink B, Kovacs KM, Lydersen C, McMeans BC, Svavarsson J, Kessel ST, Porsmoguer SB, et al. Origins of the greenland shark (*Somniosus microcephalus*): Impacts of ice-olation and introgression. Ecol Evol. 2017;7(19):8113–25.

93. Rieseberg LH, Soltis DE. Phylogenetic consequences of cytoplasmic gene flow in plants. Evol Trend Plant. 1991;5(1):65–84.

Publisher's Note