BMC Plant Biology

RESEARCH Open Access

Phylogenomics and structural variations of plastid genomes of the columbine genus (Ranunculaceae)

Fang-Min Yu[1,2,3†], Wen-Chuang He[4†], Andrey S. Erst[5†], Huan-Wen Peng[1,2,3], Yuan-Yuan Ling[1,2,3], Tatyana V. Erst[6,7], Michael V. Bardukov[5,8], Wei Wang[1,2,3*] and Kun-Li Xiang[1,2*]

Abstract

Background The columbine genus (*Aquilegia*) contains many species with horticultural and pharmaceutical importance. However, a well-resolved phylogeny for this genus remains lacking owing to recent and rapid radiation. We obtained plastomes of 75 *Aquilegia* species and six species of other genera in Thalictroideae to reconstruct a robust phylogeny for *Aquilegia*. Within the phylogenetic framework, we investigated the evolutionary patterns of structural variations in *Aquilegia* plastomes. We also examined appropriate molecular markers for species identification of this genus.

Results Plastid phylogenomic analyses show that *Aquilegia* is monophyletic and divided into two major clades and eight subclades, largely in agreement with species geographical distributions. *Aquilegia* plastomes have copy number variations (CNVs) of genes at the boundaries of IR/SC regions. An inversion was identified in the large single-copy region (LSC) plastomes of *Aquilegia*, *Semiaquilegia*, and *Urophysa*. The inversion was adjacent to three distinct structures (type I-III) in LSC region, each characterized by presence or absence of different short inverted repeat (sIR) regions. The most recent common ancestor of *Aquilegia* species was type III, and other two types were derived by shortening or loss of sIRs. Ten hypervariable regions were identified in *Aquilegia* plastomes.

Conclusions We present a well-resolved phylogeny for *Aquilegia*. The CNVs of genes were observed frequently at the boundaries of IR/SC regions and sIR regions in the genus, and the inversions and sIR in LSC regions have phylogenetic signal and lead to notable variations in gene content and length of plastomes. The new hypervariable regions can be used for phylogenetic analyses and molecular identification of *Aquilegia*. This study provides new insights into the structural evolution of plastome in angiosperms, and also have important implications for the conservation and utilization of *Aquilegia* species.

Keywords *Aquilegia*, Plastid genome, Structures variation, Phylogeny, Hypervariable region

†Fang-Min Yu, Wen-Chuang He and Andrey S. Erst contributed equally to this work.

*Correspondence:
Wei Wang
wangwei1127@ibcas.ac.cn
Kun-Li Xiang
kunlixiang@ibcas.ac.cn

Full list of author information is available at the end of the article

Background

Plastid genomes (plastomes) of angiosperms are generally recognized to be highly conserved in gene content and structure [1–3]. Typical angiosperm plastomes display a quadripartite circular structure, including two inverted repeat regions (IRs), a large single-copy region (LSC), and a small single-copy region (SSC) [4, 5]. However, many studies have identified structural variations (SVs) in the plastomes of heterotrophic plants, i.e., parasites and mycoheterotrophs have more presence/absence variations (PAVs) in their plastomes causing by the loss of gene functions [6–9]. Increasing of reported plastomes in recent years, lots of SVs of plastomes have been widely reported in angiosperms [10]. The plastome SVs mainly focus on copy number variations (CNVs) resulting from the expansion/contraction of IR regions [11–13]. For example, large-scale IR expansions exceeding several kb results in increased gene copy numbers through duplication events, including *Tetracentron* Oliv. and *Trochodendron* Siebold & Zucc. (Trochodendraceae) [6] and *Annona* L. (Annonaceae) [14]. Conversely, IR losses have been observed in the boundaries of IR/SC regions of *Myosurus* L. and *Ceratocephala* Moench (Ranunculaceae) [15]. However, the evolutionary patterns of the variations within LSC region and/or IR region remain poorly understood.

In particular, SVs in the LSC region of plastomes have been reported and attracted significant attention [16, 17]. Unprecedented SVs have been found in the LSC region of plastomes in Poales, in which one-third of the families have plastomes with inversions exceeding 1 kb in length [17]. A 51 kb inversion spanning *rbcL* to *rpsl6* in the LSC region was reported in the *Lotus japonicus* (Regel) K. Larsen (Fabaceae) plastome [18]. Similar inversions were observed in the LSC region of Ranunculaceae plastomes, including six inversions in Anemoneae [19] and a single 42 kb inversion in *Adonis aestivalis* L. [20]. Although inversions in the LSC region lead to notable variations in the structure, gene content, and length of plastomes, the phylogenetic significance of SVs in the LSC region still needs to be assessed.

Aquilegia L. (Ranunculaceae), commonly known as the columbine genus, comprises *ca.* 80 perennial herbaceous species and is widely distributed in the Northern Hemiphere [21, 22]. Because petals base usually prolong into a spur, flowers of *Aquilegia* are beautiful and unique, making *Aquilegia* species excellent ornamental plants for gardens. Some *Aquilegia* species are pharmaceutically important and have become traditional Chinese medicines [23]. In addition, *Aquilegia* has been a new model for studying plant development, ecology, and evolution [24–27]. Among the genera with more than 15 species in Ranunculaceae, only *Aquilegia* lacks a infrageneric classification system [22]. This is may be because that *Aquilegia* experienced a recent and rapid radiation [28]. Hodges & Arnold (1994) first used molecular data (plastid *atpB-rbcL* and nuclear ITS) to investigate phylogenetic relationships within *Aquilegia* [29], where 14 species of this genus were sampled. In their phylogenetic tree, the monophyletic *Aquilegia* was recognized, but intrageneric relationships were poorly resolved [29]. Based on plastid (*trnK-matK* and *trnS-G*) and nuclear (ITS) sequences, Bastida et al. (2010) sampled 32 *Aquilegia* species and did not resolved inter-species relationships based on maximum parsimony analysis. By sampling 62 recognized species and subspecific entities of *Aquilegia* and 21 plastid DNA regions [30], Fior et al. (2013) divided *Aquilegia* into six geographic groups, but a few species did not belong to any group and the relationships within Europea group were not resolved at all [31]. Recently, Zhang et al. (2021) built a well-supported phylogenetic framework for this genus based on complete plastome sequences, but only included 16 species [32]. To reconstruct a robust phylogenetic framework for *Aquilegia*, more species and plastome sequences of this genus need to be included.

Importantly, Zhai et al. [16] identified an inversion between *accD* and *clpP*1 in *Aquilegia* and its allies (*Semiaquilegia* Makino and *Urophysa* Ulbr.), as compared to other genera in Ranunculaceae. However, only two *Aquilegia* species (*Aquilegia coerulea* E. James and *Aquilegia ecalcarata* Eastw.), as well as one species each of *Semiaquilegia* and *Urophysa*, were sampled by Zhai et al. [16]. It is unknown whether this inversion is of phylogenetic significance and widespread in the plastomes of *Aquilegia* species, especially whether this SV is consistent or variable in *Aquilegia*.

In this study, we sequenced complete plastomes of 75 *Aquilegia* species using the genome-skimming sequencing technique and performed the first comprehensively phylogenetic analysis of *Aquilegia* based on plastome data. Our aims were as follows: (1) to reconstruct a robust phylogeny for *Aquilegia*; (2) to identify SVs of the *Aquilegia* plastomes; (3) to investigate the evolutionary patterns of SVs in *Aquilegia* plastomes; (4) to identify hypervariable regions for efficient and accurate species identification in this genus. The study will enhance the understanding of the evolutionary patterns of plastome SVs in angiosperms, and will facilitate exploration and utilization of *Aquilegia* plants.

Results

Plastome feature

The complete plastomes of 75 *Aquilegia* species and five species from *Dichocarpum* W.T. Wang & P.K. Hsiao, *Semiaquilegia*, *Thalictrum* L., and *Urophysa* in Thalictroideae were assembled and annotated (Table S1). Including the plastome of *Semiaquilegia danxiashanensis*

L. Wu, J.J. Zhou, Q. Zhang & W.S. Deng previously reported, the plastomes of the 81 Thalictroideae species that were sampled in this study have the sizes of the plastomes from 153727 bp in *Dichocarpum* to 164138 bp in *Urophysa* (Table S2). All plastomes exhibited the typical quadripartite structure (Fig. 1), which comprises a pair of IRs (25853–27544 bp) separated by a LSC region (83381–91683 bp) and an SSC region (16702–17598 bp). The plastome sizes of *Aquilegia* (159782–162480 bp), *Semiaquilegia* (160548–162227 bp) and *Urophysa* (164138 bp) were larger than that of other Thalictroideae genera (153727–156258 bp). The plastome synteny of the

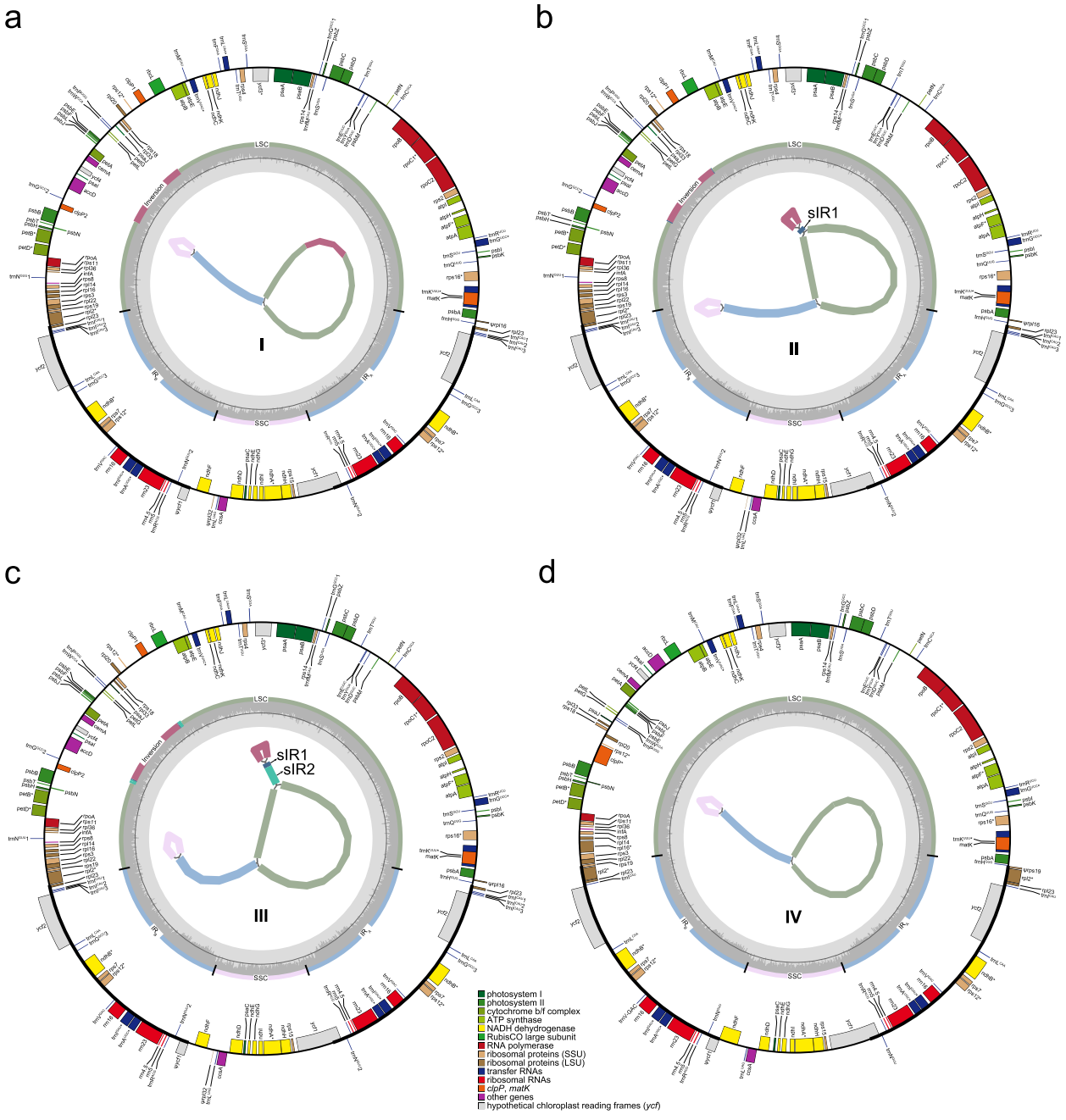

Fig. 1 The four types of plastomes in Thalictroideae. **a** Type I has an inversion without short inverted repeat (sIR), the plastome of *Aquilegia aradanica* Shaulo & Erst. **b** Type II has an inversion with sIR1, the plastome of *Aquilegia alpina* L. **c** Type III has an inversion with sIR1 and sIR2, the plastome of *Aquilegia jonesii* Parry. **d** Type IV does not have an inversion, the plastome of *Dichocarpum dalzielii* (J.R.Drumm. & Hutch.) W.T. Wang & P.K. Hsiao. The genes inside and outside of the circle are transcribed in clockwise and counterclockwise directions, respectively. Genes belonging to different functional groups are shown in different colors. The thick lines indicate the extent of inverted repeats (IR_A and IR_B) that separate the genomes into small single-copy (SSC) and large single-copy (LSC) regions. The pink thick line indicates an inversion in the LSC region

subfamily in terms of the gene arrangement is generally high except for in the LSC region, which shows considerable SVs (Fig. 1). Based on the structural differences in the LSC region, four types (I-IV) of plastome structures were recognized in Thalictroideae. The type I is present in seven *Aquilegia* species, having an *ca.* 12807–12825 bp inversion (INV1) between *accD* and *clpP*1 (Fig. 1a; Table S2). The type II is present in 32 *Aquilegia* species, having a pair of short repeated regions (sIRs1) at both the edges of the 12662–12853 bp inversion, compared to type I (Fig. 1b; Table S2). The type III is present in 36 *Aquilegia* species and in *Urophysa* and *Semiaquilegia*, having longer sIRs (including sIRs1 and sIRs2) at both the edges of the 12592–13121 bp inversion, compared to type II (Fig. 1c; Table S2). The plastome of *S. danxiashanensis* previously reported have a complete sIR1 and an incomplete sIR2, we also identified it as type III although its sIR2 is very short. The type IV is present in *Dichocarpum* and *Thalictrum*, without the inversion and sIR regions (Fig. 1d; Table S2).

To investigate the contraction and expansion events of plastomes of five genera in Thalictroideae, we compared the borders of the IR/SSC regions (Fig. 2a). For the IR_A/SSC region, *ycf1* spanned the junction region, while a pseudogene fragment (*ψycf1*) was located at IR_B side of the IR_B/SSC region in all sampled five genera of Thalictroideae. In the SSC region, *Dichocarpum* lacks *rpl32* gene, whereas other four genera have a pseudogene (*ψrpl32*). For the border of the IR/LSC regions, *Aquilegia* has six copies of $trnI^{CAU}$ in the two IR regions, while other four genera have two copies of $trnI^{CAU}$. Along with the IRs expansion in *Dichocarpum* and *Thalictrum*, the boundary of the IR_B/LSC moved from intergenic region of *rpl2*/*rpl23* to the coding region of *rps19*, leading to a pseudogene fragment (*ψrps19*) at the IR_A region (Fig. 2a).

For the LSC region, two copies of *clpP* (*clpP*1 and *clpP*2) and one copy of $trnG^{GCC}$ ($trnG^{GCC}$2) were found in *Aquilegia*, *Urophysa*, and *Semiaquilegia*. Notably, both two copies of *clpP* are intronless, which are different in one copy of *clpP* with two introns in *Dichocarpum* and *Thalictrum* (Fig. 2b). Additionally, the *clpP*1 gene of *Aquilegia* and *Urophysa* contains an additional tandem repeat region, consisting of 9 and 6 repeats of a 15 bp special motif, respectively.

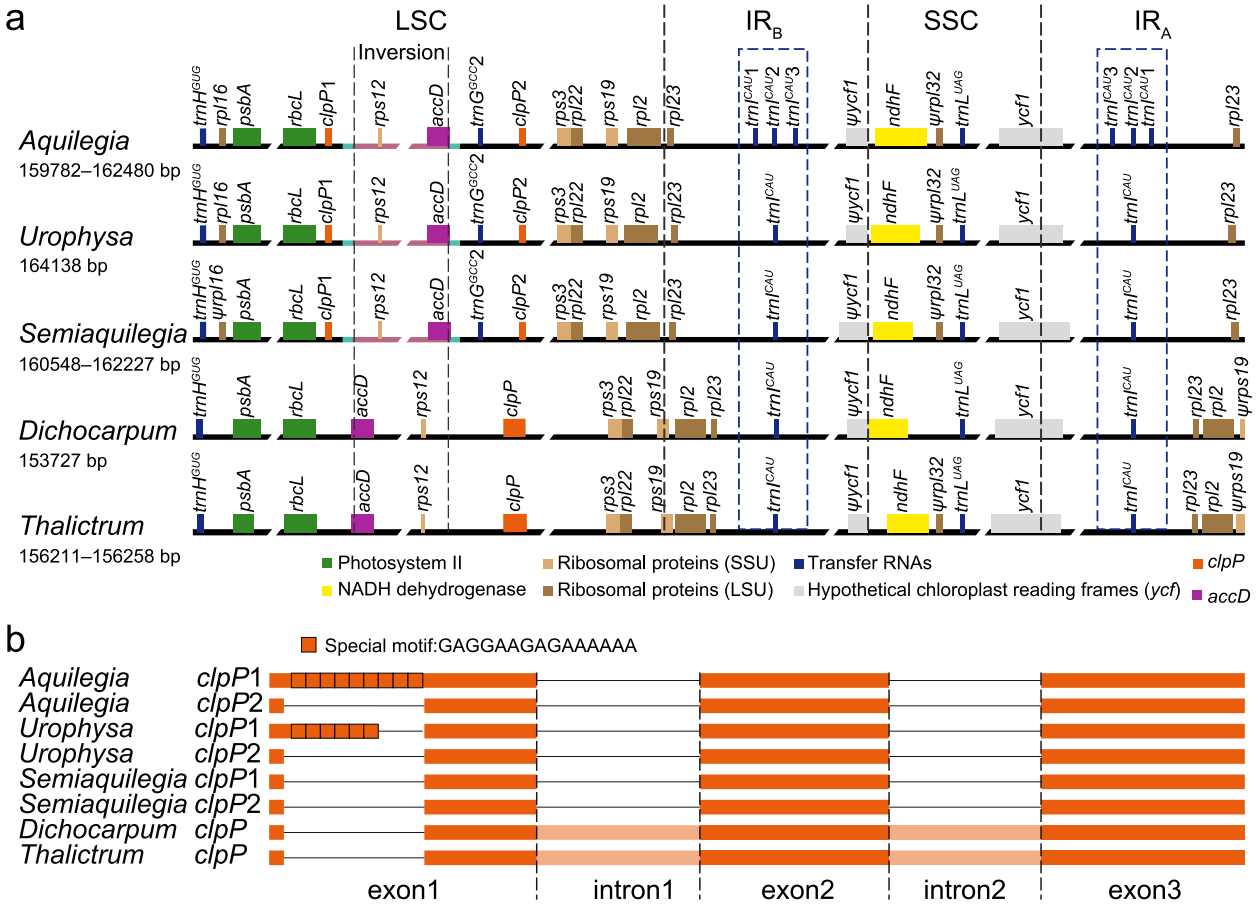

Fig. 2 Comparison of plastome structure in Thalictroideae. **a** The structure variations (SVs) of the large single-copy (LSC), inverted repeats (IRs), and small single-copy (SSC) region boundaries in Thalictroideae. The break in the middle of the bars indicates that other regions and genes are omitted here. The ψ-symbol shows putative pseudogenes. **b** The SVs of the *clpP* gene in Thalictroideae

Final, we identified 137 predicted functional genes and three pseudogenes for the 75 plastomes of *Aquilegia*. These genes were summarized as 112 unique genes including 78 protein-coding genes, 30 transfer RNA (tRNA) genes, and four ribosomal RNA (rRNA) genes (Table S3). One protein-coding gene (*ycf3*) contains two introns, 15 genes contain a single intron, including nine protein-coding genes (*atpF*, *ndhA*, *ndhB*, *petB*, *petD*, *rpl2*, *rpoC1*, *rps12*, and *rps16*) and six tRNA genes (*trnA*^{UGC}, *trnG*^{UCC}, *trnI*^{GAU}, *trnK*^{UUU}, *trnL*^{UAA}, and *trnV*^{UAC}) in *Aquilegia* (Table S3). No function genes were found in the sIRs (including sIRs1 and sIRs2) across different species in *Aquilegia*. The majority of the protein-coding genes (74/78) of *Aquilegia* contain standard AUG as the initiator codon, three genes (*ndhD*, *rps19*, and *ycf1*) contain GUG instead, and *rpl2* has ACG.

Phylogenetic analyses

Phylogenetic relationships of *Aquilegia* were reconstructed using maximum likelihood (ML) and Bayesian inference (BI) analyses, which generated almost identical topologies based on the complete plastome sequences (Fig. 3). The monophyletic *Aquilegia* was strongly supported [bootstrap support (BS) = 99%, posterior probability (PP) = 1.0], sister to *Urophysa* with weak support (BS = 58%, PP = 0.99). *Aquilegia* includes two major monophyletic groups (I and II) and eight well-supported clades, each defined largely by geography. Clade I-1 consists of one North American species, clade I-2 includes four species from Asia (BS = 100%, PP = 1.0), and clade I-3 comprises most North American species with one nested Italian species (*Aquilegia nugorensis* Arrigoni & E. Nardi) (BS = 100%, PP = 1.0). Clade II-1 consists of five Asian species (BS = 99%, PP = 1.0), clade II-2 includes 15 Asian species with one nested European species (*Aquilegia viscosa* Gouan) (BS = 93%, PP = 1.0), clade II-3 comprises six European species (BS = 68%, PP = 1.0), clade II-4 comprises six European species (BS = 96%, PP = 1.0), and clade II-5 comprises 20 European species (BS = 100%, PP = 1.0). The majority of interspecies relationships in *Aquilegia* are well resolved (Fig. 3).

Seven plastid data subsets, including the CDS of 78 protein-coding genes, the 15 introns, the 117 intergenic spacer (IGS) regions, a combination of 15 introns and 117 IGS regions, the LSC region, the SSC region, and the IR region, were also used to perform phylogenetic analyses using ML and BI method. Phylogenetic analyses of seven plastid data subsets did not generate highly supported topological conflicts compared to the complete plastome dataset, except for the position of clade I (Figs. S1–S7). The 117-IGS and combined IGS and intron subsets recovered clade I-1 as basalmost in clade I with moderate support (BS = 82%, PP = 1.0; BS = 83%, PP = 1.0), the LSC subset recovered clade I-1 as sister to clade I-3 (BS = 79%, PP = 1.0), whereas the SSC subset recovered clade I-1 as sister to clade I-2 (BS = 98%, PP = 1.0).

Evolutionary inference of SVs in LSC region

To infer the evolutionary patterns of SVs of the LSC region, we coded the four types for LSC structures in Thalictroideae: type I with an inversion, type II with an inversion and sIRs1, type III with an inversion, sIRs1, and sIRs2, and type IV without an inversions and sIRs. By comparing the equal-rates (ER), the symmetrical (SYM), and all-rates-different (ARD) likelihood models, we identified ER as best-fit model for our data (Table S4).

Ancestral state reconstruction of the LSC structure is shown in Fig. 4. The most recent common ancestor (MRCA) of *Aquilegia* is type III, i.e., a 12592–13121 bp inversion neighbored with a pair of long sIRs including sIRs1 and sIRs2 (up to 876 bp). Within clade II, the entire sIR structure has lost once in clade II-2 (type III), and the sIR2 structure has lost once for the MRCA of clades II-3, II-4, and II-5 (type II). Interestingly, sIR1 and sIR2 sequences varies greatly across *Aquilegia*, i.e., the length of all sIRs1 is 49 bp, while the lengths of the sIR2 range from 14 to 876 bp (Fig. 4; Table S2).

To further investigate the sequence variation of the sIR regions, we selected 21 representative species of *Aquilegia*, covering all eight clades, all types of structure and all the mutations in the sIR regions (Fig. S8). Apart from the large insertions or deletions that lead to different structures in the sIR region, there also exist 11 small sequence variations in the variable sIR2 of *Aquilegia* (Fig. 5), implying the unusual rich polymorphism of this region.

Identification of hypervariable regions

To identify the potential hypervariable regions in plastomes of *Aquilegia*, we calculated plastome nucleotide diversity for this genus. Nucleotide polymorphism analyses of the genes and intergenic regions consistently show that IR regions are more conserved than LSC and SSC regions (Fig. 6). For genes, nucleotide diversity ranges from 0 (e.g., *trnH*^{GUG}, *trnQ*^{UUG}, and *trnS*^{GCU}) to 0.07967 (*infA*) with an average of 0.00118, whereby ten genes (i.e., *infA*, *trnI*^{CAU}3, *ψrpl16*, *trnS*^{GGA}, *ccsA*, *atpH*, *ycf1*, *trnL*^{CAA}, *psbH*, and *ndhE*) have remarkably high values (π >0.0017; Fig. 6a; Table S5). For the exons of genes, nucleotide diversity ranges from 0 (e.g., *trnK*^{UUU}(*matK*), *trnG*^{UCC}, and *trnL*^{UAA}) to 0.07967 (*infA*) with an average of 0.00107. For the IGS regions, nucleotide diversity ranges from 0 (e.g., *rpoC1–rpoB*, *trnY*^{GUA}–*trnE*^{UUC}, and *psbD–psbC*) to 0.03486 (*rpl2–rpl23*) with an average of 0.00223. Ten of IGS regions have considerably high values (π >0.00706), including *rpl2–rpl23*, *trnL*^{CAA}–*trnG*^{GCC}3, *trnN*^{GUU}1–*infA*, *trnT*^{GGU}–*psbD*, *trnH*^{GUG}–*ψrpl16*, *rpl16–rps3*, *psbT–psbN*, *rpl14–rpl16*, *ccsA–ndhD*, and *rps19–rpl2* (Fig. 6b; Table S6). After removing the short

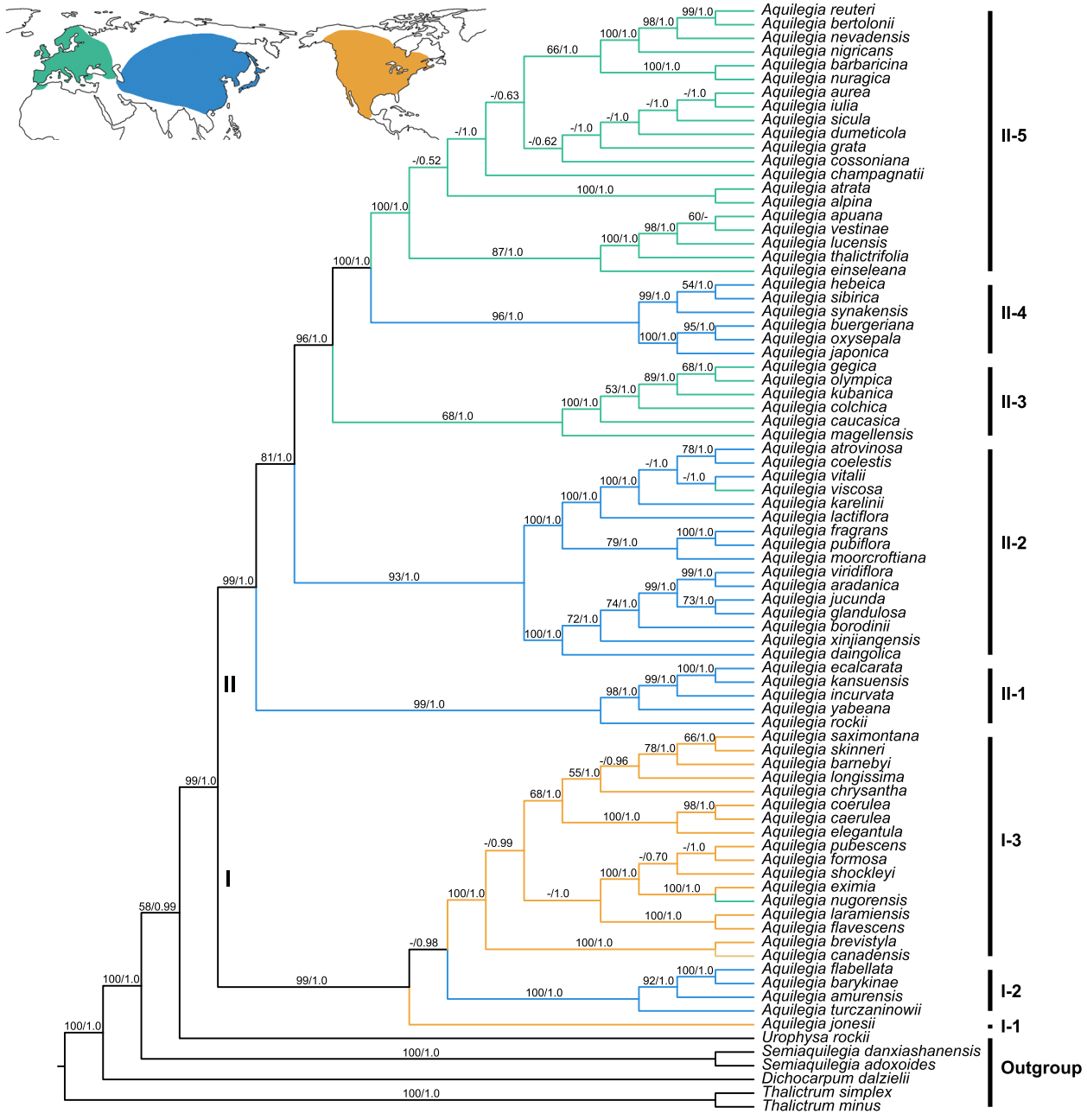

Fig. 3 Phylogenetic relationship of *Aquilegia* obtained from maximum likelihood analysis of the complete plastome dataset. The color of line is referred to the geographical distributions of *Aquilegia* species. Orange, blue and green color indicate North America, Asia and Europe, respectively. Numbers above branches are bootstrap values (BS ≥ 50 %) and posterior probabilities (PP ≥ 0.50), "-" indicates the nodes not supported

hypervariable regions (< 100 bp), nine non-coding regions (*rpl2–rpl23*, *trnL*^{CAA}*–trnG*^{GCC}3, *trnN*^{GUU}1*–infA*, *trnT*^{GGU}*–psbD*, *trnH*^{GUG}*–ψrpl16*, *rpl16–rps3*, *psbT–psbN*, *rpl14–rpl16*, and *ccsA–ndhD*) and one gene (*infA*) of plastomes were identified as hypervariable regions in *Aquilegia*. Phylogenetic analyses based on the ten hypervariable regions resolved the relationships among most species of this genus, and the sequence divergences of the combined ten hypervariable regions are >0 between *Aquilegia* species (Fig. S9).

Discussion

Phylogenetic relationships

All our phylogenetic analyses strongly supported *Aquilegia* as monophyletic (Figs. 3 and S1–S7), in agreement with previous studies (e.g., Bastida et al. [30]; Fior et al. [31]; Ro & McPheron [33]). *Aquilegia* has undergone a recent and rapid radiation [28, 29, 34], the interspecies relationships remain ambiguous using a wide variety of molecular data [29–31]. Based on the complete plastome data, we discover that *Aquilegia* contains two major

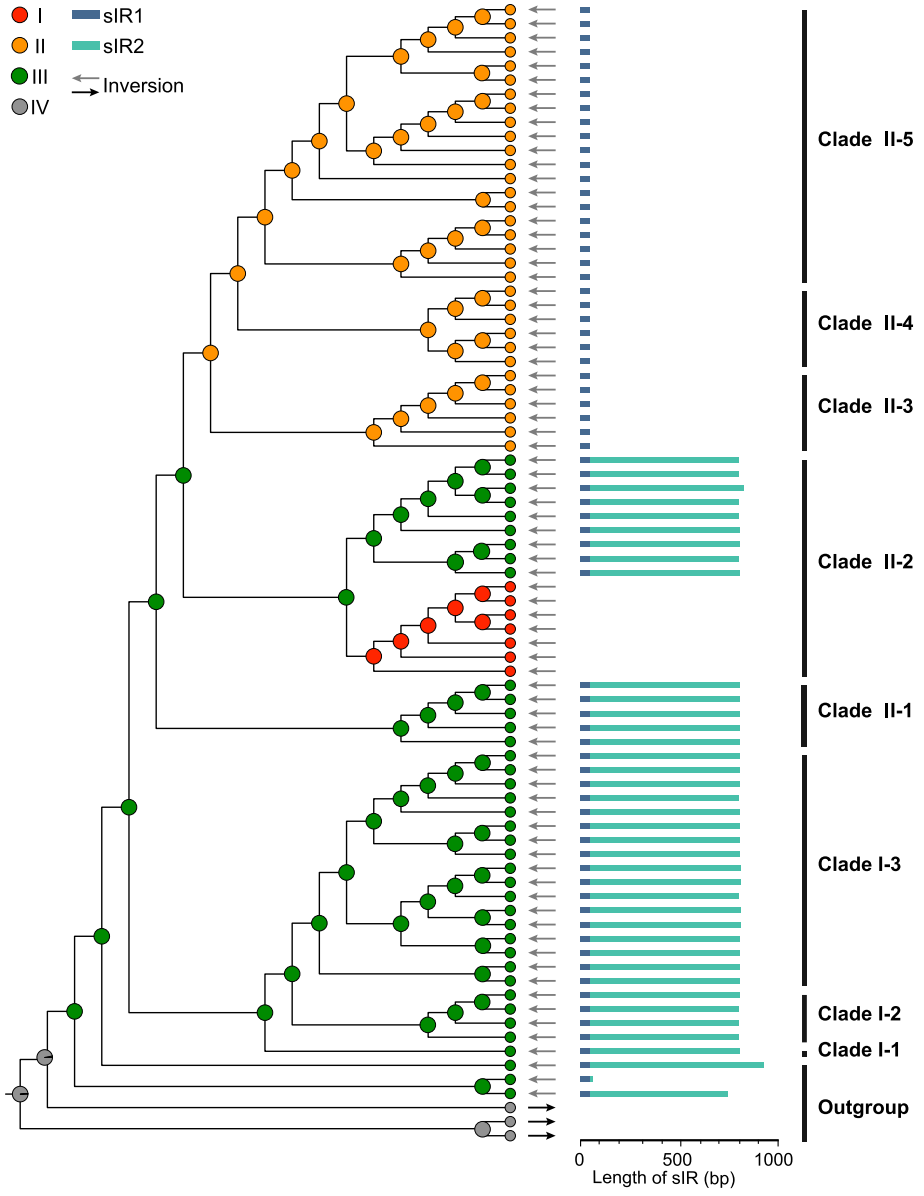

Fig. 4 Evolutionary inference of structure variation in *Aquilegia* plastomes. Colored pie charts show relative probabilities of alternative ancestral character states obtained in phytools. The arrow indicates the direction of inversion. The stacking diagram shows the length of short inverted repeat (sIR), as referred to Table S2

monophyletic groups and eight clades, largely consistent with their geographical distributions (Fig. 3). We provide a well-resolved phylogenetic framework for *Aquilegia* with a more extensive sampling than in any previous study, which is a foundation for the future taxonomic revision for this genus.

The phylogenetic analyses of *Aquilegia* were also conducted based on seven plastid subsets. Compared to other subsets as well as complete plastome dataset, the 78-CDS, 15-intron, and SSC subsets have a relatively weak resolution for the relationships within *Aquilegia* (Figs S1–2, S7). Overall, the phylogeny generated by the complete plastome dataset received higher support than those inferred from seven subsets, suggesting that the complete plastome sequences can resolve well phylogenetic relationships of taxa with recent and rapid radiation, such as *Aquilegia*. The similar cases are also found for *Rubus* L. (Rosaceae) [35] and *Saussurea* DC. (Asteraceae) [36].

Plastome feature

The contraction and expansion in IRs are considered important evolution events, which can change plastome sizes and gene contents [37–39]. Our results show that *Aquilegia*, *Urophysa*, and *Semiaquilegia* have one copy for each of *rpl2* and *rps19* in the LSC regions, while

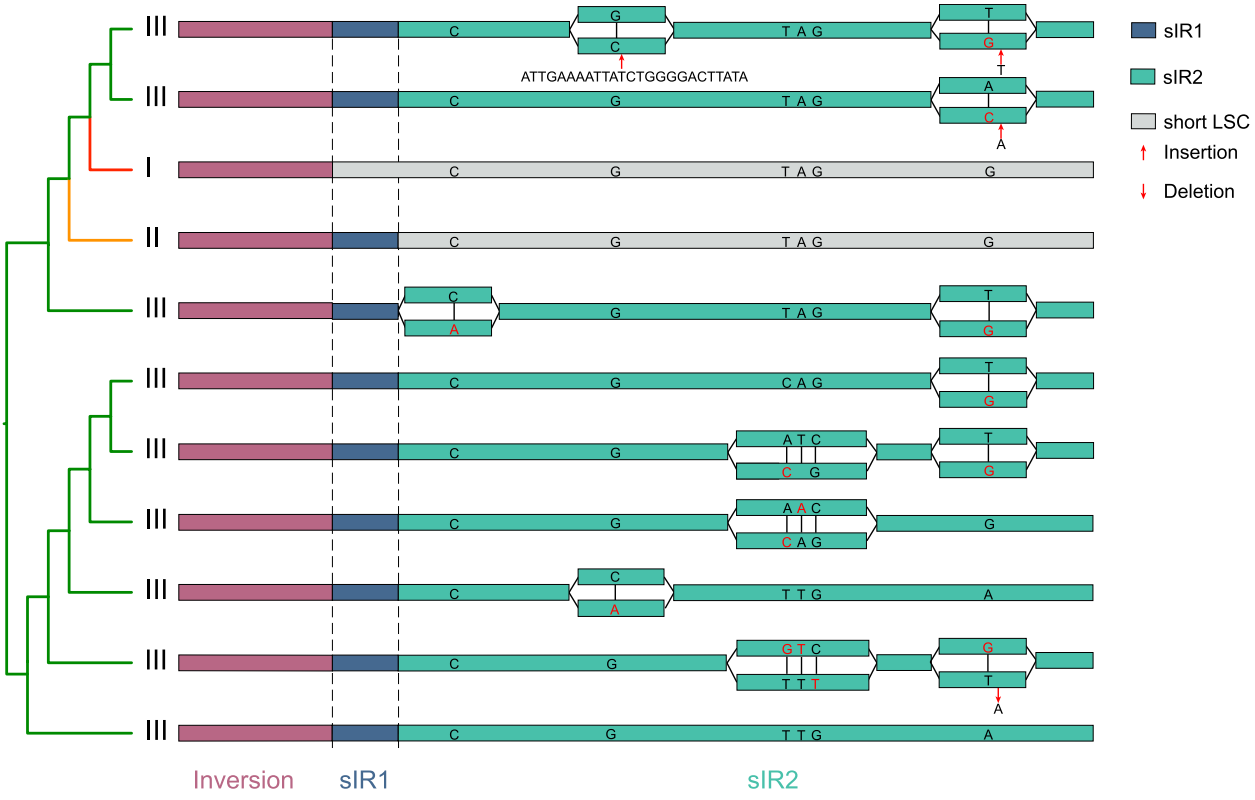

Fig. 5 The structure variations (SVs) in the large single-copy (LSC) region of 11 *Aquilegia* plastomes. The color of branch shows the inferred state of SVs, as referred to Fig. 4. sIR, short inverted repeat

Dichocarpum and *Thalictrum* have two copies for each of *rpl2* and *rps19* in the IRs (Fig. 2a). These CNVs may result from the contraction of IR regions in *Aquilegia*, *Urophysa*, and *Semiaquilegia*. It is usual that CNVs caused by the expansion/contraction of IR regions. For example, one-copy loss of *rpl2* and *rpl23* was found in the IR/LSC region of *Myosurus* (Ranunculaceae) [15] and one-copy loss of *rbcL* was identified in the IR/LSC regions of *Pelargonium transvaalense* R. Knuth (Geraniaceae) [11].

The length of IR regions of *Aquilegia* (25853–26834 bp) is similar to that of *Urophysa* (27544 bp), *Semiaquilegia* (26640–26829 bp), *Dichocarpum* (26448 bp), and *Thalictrum* (26480–26481 bp; Table S2). Especially, we found *Aquilegia* had six copies of $trnI^{CAU}$ in the IR regions near the borders of IR_A/LSC, while the other four genera in Thalictroideae only have two copies (Fig. 2a). The CNVs of $trnI^{CAU}$ are resulted from the small direct repeats of plastomes, instead of expansion/contraction of IR regions. These SVs might have occurred along with the expansion/contraction of IR regions because they are located near the borders of IR/LSC. The small direct repeats of plastomes in IR regions lead to an increase in the length of IRs in *Aquilegia.* Thus, the contraction of IRs in *Aquilegia* do not result in its IR regions being shorter.

In fact, the size of plastomes in *Aquilegia* (159782–162480 bp), as well as *Urophysa* (164138 bp) and *Semiaquilegia* (160548–162227 bp), is bigger than that in *Dichocarpum* (153727 bp) and *Thalictrum* (156211–156258 bp), as they have longer LSC regions (Table S2; Fig. 2). We found that *Aquilegia*, *Urophysa*, and *Semiaquilegia* have 12592–13121 bp inversions in LSC regions, whereas *Dichocarpum* and *Thalictrum* have 12148–12743bp homology region. Additionally, *Aquilegia*, *Semiaquilegia*, and *Urophysa* have two copies of sIR near the inversion, while *Dichocarpum* and *Thalictrum* have one copy of homology region. *Aquilegia*, *Semiaquilegia*, and *Urophysa* have one copy of $trnG^{GCC}$ and two copies of *clpP* gene separately in the LSC regions near the inversion, while *Dichocarpum* and *Thalictrum* have one copy of *clpP* gene. These CNVs of genes in LSC regions occur along with sIR as it is located near the borders of inversion in *Aquilegia*, leading to the increase in the length of the LSC in *Aquilegia*, *Semiaquilegia*, and *Urophysa.*

For the CNV of *clpP* gene, the intronless *clpP* gene in *Aquilegia*, *Semiaquilegia*, and *Urophysa* have been duplicated in the LSC region, while *Thalictrum* and *Dichocarpum* with have only one copy of *clpP* with two introns (Fig. 2b). Zhai et al. [16] inferred that the loss of the second intron likely occurred before the divergence

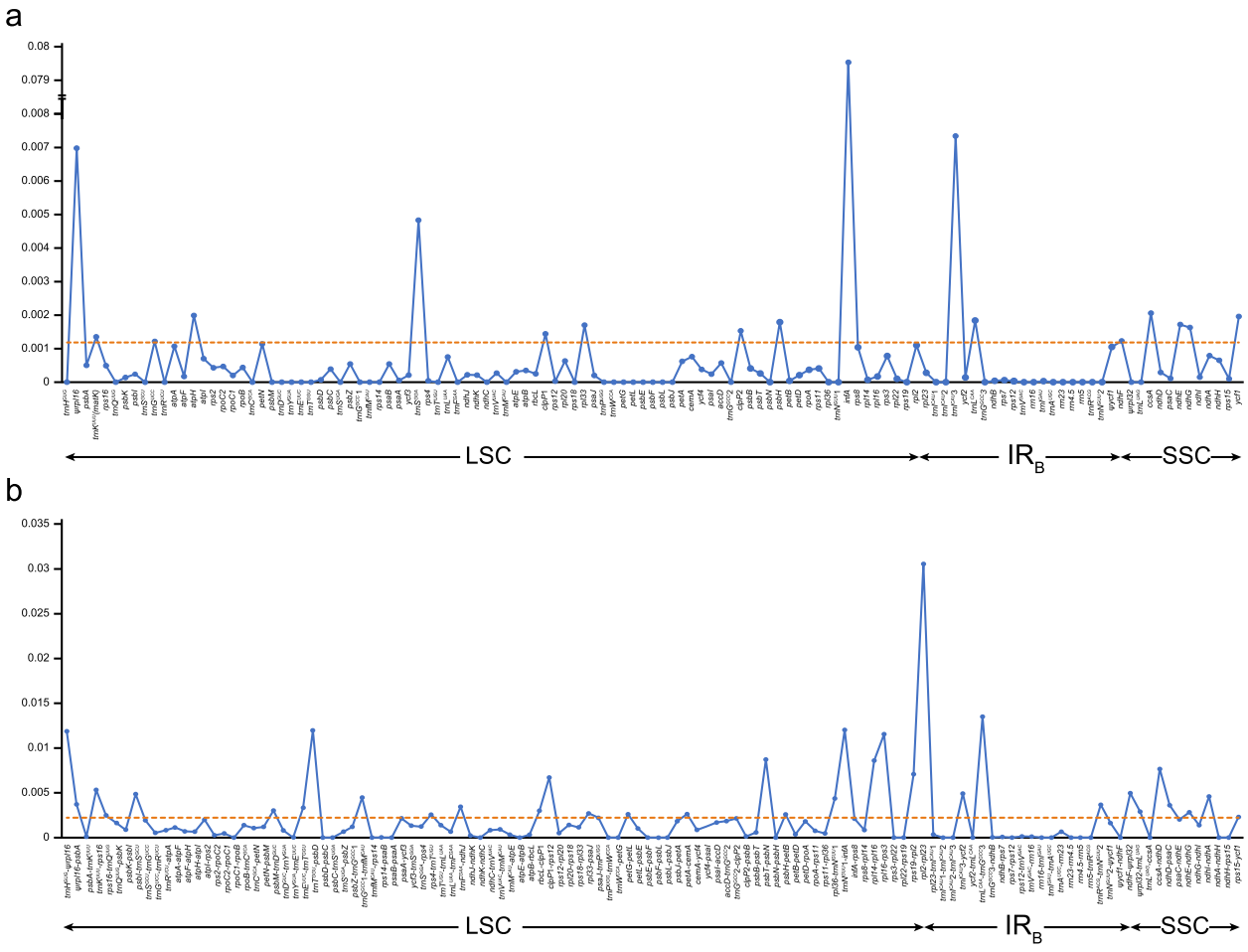

Fig. 6 Comparison of nucleotide variability (Pi) values in *Aquilegia* plastomes. **a** Pi values among genes, **b** Pi values among intergenic spacer regions. The dot line denotes the average value. IR, inverted repeat; SSC, small single-copy; LSC, large single-copy

of *Aquilegia*, *Semiaquilegia*, and *Urophysa*, followed by the loss of the first intron before the gene was duplicated. Additionally, our results indicate that the *clpP*1 gene contained approximately nine and six tandem repeats of a special motif in *Aquilegia* and *Urophysa*, while the other copy are without (Fig. 2b), suggested that CNV play an important role in evolution of plastomes in *Aquilegia*.

In total, *Aquilegia* plastomes have CNVs of genes at the boundaries of IR/SC regions and sIR regions. The small direct repeats in plastomes of *Aquilegia* inhibit contraction and shorten of IRs. The length of *Aquilegia*, *Semiaquilegia*, and *Urophysa* plastome is bigger than that in *Dichocarpum* and *Thalictrum*, influenced on the inversion event occurring in the LSC region.

Structure evolution of *Aquilegia* plastomes

All plastomes in *Aquilegia*, *Semiaquilegia*, and *Urophysa* contain an inversion in the LSC region between *accD*–*clpP*1 (Fig. 2a). Inversion in the LSC region of the plastome has also been found in other taxa. For example, *Adonis sutchuenensis* Franch. in Ranunculaceae has an inversion between *rps16* and $trnT^{UGU}$ [16]. Plastome inversion has also been reported in the LSC region of *Lotus japonicus* between *rbcL*–*rps16* [18], and multiple inversions have been found in the plastomes of Poales [17, 40].

Here, we found three types of LSC structures in *Aquilegia* plastomes, which were likely generated by a inversion event of and exhibited phylogenetic signals. The ancestral structure of *Aquilegia* is type III, having a pair of longer sIRs (including sIRs1 and sIRs2) at the edge of the inversion in *Aquilegia* (Fig. 4). The sIRs are shortened and lost at the edge of the inversion in Type II and Type I, respectively (Fig. 1c), showing 11 small variations in the plastomes of *Aquilegia* (Fig. 5). As the inversion could be probably induced by the neighbored sIR region, the large or small variations in the sIR regions (Figs. 1 and 4) could contribute to the fixation of this inversion in the plastomes of the *Aquilegia*. Extensive repeats in the plastomes were found to be positively correlated with the observed inversions and rearrangements in Poales [17]. We further found that the inversions in LSC regions were

induced by repeats and possibly fixed by lose of repeats in *Aquilegia*.

Inversions in the LSC regions also lead to notable variations in gene content and length of plastomes. *Aquilegia*, *Semiaquilegia*, and *Urophysa* had one extra copy of *trnG*^{GCC} and two copies of intronless *clpP* gene in the LSC regions near the inversion, while *Thalictrum* and *Dichocarpum* had only one copy of *clpP* gene with two introns in similar region. Additionally, although *Aquilegia*, *Semiaquilegia*, and *Urophysa* have longer LSC regions owing to the inversions, their total plastid lengths are thus longer than those of *Dichocarpum* and *Thalictrum* (Fig. 2a). Thus, the inversions may contribute to the increased copy number of genes in near the inversions and have a strong impact on the length of plastome.

In all, the LSC structure with both the inversion and the longer sIR was estimated as the ancestral state, which have evolved independently into other structures with loss of different sIR regions. The plastid SVs have phylogenetic signals in *Aquilegia*. The loss of distinct sIR regions led to notable variations in the structure, gene content, and length of the plastomes.

Hypervariable regions in plastomes of *Aquilegia*

Highly polymorphic regions in plastomes are widely utilized as molecular markers for phylogenetic reconstruction [41–43] and species identification [44–46]. Phylogenetic analyses of *Aquilegia* based on nuclear ITS sequences and one plastid region (*atpB–rbcL*; Table S6) [29] and ITS and two plastid regions (*trnK–matK*, *trnS–G*; Table S6) [30] have failed to effectively distinguish infrageneric relationships due to low nucleotide variability among different species. More plastid fragments, such as *rpl32–trnL*^{UAG}, *trnK*^{UUU}*–rps16*, *psbM–trnD*^{GUC}, *psaI–accD*, *ndhJ–trnF*, *rps16–trnQ*^{UUG}, *trnD*^{GUC}*–trnT*^{GGU}, *trnT*^{GGU}*–psbD*, *trnC*^{GCA}*–rpoB*, *trnC*^{GCA}*–ycf6*, *atpH–atpI*, *psbE–petL*, *ycf6–psbM*, *trnS*^{GUC}*–trnG*^{UUC}, *ndhC–trnV*^{UAC}, *trnS*^{UGA}*–trnG*^{GCC}, *psbB–psbH*, *rps4–trnF*^{GAA}, *rps12–rpl20*, *matK* and intron of *ndhA*, have also been applied to analyze intrageneric relationships of *Aquilegia* and result in highly supported phylogenetic trees (Table S5–S6) [31]. Additionally, candidate molecular markers of *trnT*^{GGU}*–psbD* and *psbT–psbN* were identified based on 62 species of the genus by Fior et al. [31]. Therefore, additional data source beyond the commonly used markers are needed for phylogenetic analysis.

In the present study, the whole genome and non-coding regions manifested higher sequence divergence than protein-coding genes of *Aquilegia* plastomes (Figs. 3 and S1-4), which was consistent with most higher plants [47, 48]. After removing the short hypervariable regions (< 100 bp), nine non-coding regions (*rpl2–rpl23*, *trnL*^{CAA}*–trnG*^{GCC}3, *trnN*^{GUU}1*–infA*, *trnT*^{GGU}*–psbD*, *trnH*^{GUG}*–ψrpl16*, *rpl16–rps3*, *psbT–psbN*, *rpl14–rpl16*, and *ccsA–ndhD*) and one gene (*infA*) of plastomes were identified as hypervariable regions. Based on the ten identified hypervariable regions, the sisters of most *Aquilegia* species can be recognized and all sequence divergences between *Aquilegia* species are >0 (Fig. S9), suggesting that the ten regions can serve as potential molecular markers for *Aquilegia* and have implications for molecular taxonomy, phylogeography, and genetic diversity assessment of this genus.

Conclusions

In this study, we present a well-supported phylogenetic framework for *Aquilegia* based on plastome data. We found the CNVs of genes are observed frequently for the contraction of IR and sIR regions, such as *rpl2*, *rps19*, *trnG*^{GCC}, *trnI*^{CAU}, and *clpP*, indicating its important role to the plastome evolution of *Aquilegia*. In particular, we found a common inversion in plastomes of *Aquilegia* and its allies. The inversion was divided into three distinct structure types (I-III) in the LSC region, each characterized by different short inverted repeat (sIR) regions. The type III with both the inversion and the longer sIR was inferred as the ancestral structure, which evolved into other structures with loss of different sIR regions. The inversions and sIR have phylogenetic signal and lead to notable variations in gene content and length of plastomes. We also identified potential molecular markers for species identification of this genus. This study provides new insights into the structural evolution pattern of plastome for one of the most diversified plant taxa in angiosperms.

Materials and methods

Sample sequencing, data assembly and annotation

We sampled 75 species of *Aquilegia*. Based on Ling et al. [49], we selected six species of *Dichocarpum*, *Semiaquilegia*, *Thalictrum* and *Urophysa* in Thalictroideae as outgroups. Supplementary Table S1 documents the geographic origins and depository herbaria for all newly acquired samples. Their formal identification was undertaken by Wei Wang and Andrey S. Erst. No specific permissions or licenses were required for our collections and experiments. Total genomic DNA was extracted from about 1.5 g of silica geldried leaves of living plants or herbarium specimens using a modified cetyltrimethylammonium bromide (CTAB) method [50] and purified using the Tiangen Isolation/Extraction/Purification Kit (Tiangen Biotech (Beijing) Co., Ltd.). Short insert of 300–500 bp libraries were prepared for sequencing on the Illumina HiSeq X-Ten platform. Approximately 10 GB of raw sequencing data was generated for each sample. To obtain high-quality genome sequences, raw reads were processed using Fastp v0.19.7 [51] to remove adapter sequences and low-quality reads.

The complete plastomes were assembled from the genome-skimming sequencing data using GetOrganelle v1.7.6.1 [52] with the default parameters. The obtained plastid contigs were further visualized and scaffolded to complete circular genomes in Bandege v0.8.1 [53]. Highly accurate annotation of organelle genomes was performed by using the Organellar Genome GeSeq tool [54], with subsequent manual correction in Geneious v9.0.2 [55]. Four plastomes from *Aquilegia barnebyi* Munz (GenBank accession No. NC_053820), *Aquilegia kansuensis* (Brühl) Erst (NC_058528), *Aquilegia rockii* Munz (NC_046738), and *Aquilegia yangii* Y. Luo & L. Li (NC_058529) were used as reference sequences. The circular plastomes were visualized by using OGDRAW v1.3.1 [56], with subsequent manual editing.

Plastome variation analysis

To assess the expansion/contraction of the IR regions, we compared the boundaries between the IR/SC and their adjacent genes for the whole plastomes of five genera in Thalictroideae by IRscope [57], and then used the results to further manually modify the annotations.

Phylogenetic analyses

For the purpose of reconstructing the phylogenetic relationships, multiple sequence alignments of 81 complete plastomes were implemented using MAFFT v7 [58] with standard parameters, and further adjusted manually in Geneious v8.0.4 [55]. Phylogenetic analyses of *Aquilegia* were performed using ML and BI methods in RAxML v8.2.11 [59] and MrBayes 3.2.7 [60], respectively. The nucleotide substitution model was determined by the Akaike Information Criterion (AIC) via jModeltest 2.1.4 [61]. RAxML was conducted with the GTR + Γ substitution model, and the fast bootstrap option with 1,000 replicates. For BI analyses, two independent runs, each consisting of four Markov Chain Monte Carlo (MCMC) chains, were conducted with one tree sampled for every 1000 generations over 50 million generations. We assessed the convergence of the parameters of the two individual runs in Tracer 1.7.2 [62]. A majority rule (>50%) consensus tree was generated after discarding the first 25% of sampled trees as burn-in. Because molecular evolutionary rates of the different regions in plastome are diverse, analyses of phylogenetic relationships also were performed based on seven plastid subsets, including concatenation of 78 protein-coding genes, concatenation of 15 introns, concatenation of 117 IGS regions, concatenation of 15 introns and 117 IGS regions, a LSC region, an SSC region, and an IR region. Following the suggestion by Wang et al. (2014) [63], the thresholds BS ≥ 70% and PP ≥ 0.95 were used as an indication of significant incongruence between eight datasets.

Ancestral structural variation reconstruction

We coded the different types of LSC structures as character states. We applied the fit-Discrete function in the R package 'geiger' to decide which likelihood models (ER, SYM, or ARD) should be used. Simulated stochastic character maps were obtained using the *make.simmap* function in the R package 'phytools' under the optimal likelihood model [64, 65].

In addition, we randomly selected 21 species covering all clades and all types of LSC region SVs in these clades for comparative genomic analyses. The sequences of inversion and the sIRs in LSC region were further extracted in Bandege v0.8.1 [53] for three SV types in *Aquilegia*. To found the conserved synteny, we performed sequence alignment using MAFFT v7 [57], and adjusted manually in Geneious v9.0.2 [55].

Identifying hypervariable regions

To identifying hypervariable regions, the sequence alignment of *Aquilegia* plastomes without outgroups was subjected to a sliding window analysis in DNAsp v6.12.03 [66]. We evaluated the nucleotide diversity (Pi) of all genes, genes without introns, and IGS regions. To evaluate the effectiveness of the ten hypervariable regions identified in our study, we performed phylogenetic analyses for the 75 sampled species of *Aquilegia* using ML and BI methods, with *Urophysa rockii* Ulbr. and *Semiaquilegia adoxoides* (DC.) Makino as outgroups.

Abbreviations

IR	Inverted repeat
LSC	Large single-copy
SSC	Small single-copy
SC	single-copy
SV	Structure variation
PAV	Presence/absence variation
CNV	Copy number variation
ca.	Circa
ITS	Internal transcribed spacer
sIR	Short repeated region
tRNA	Transfer RNA
rRNA	Ribosomal RNA
ML	Maximum likelihood
BI	Bayesian inference
BS	Bootstrap support
PP	Posterior probability
IGS	Intergenic spacer
ER	Equal-rates likelihood model
SYM	Symmetrical likelihood model
ARD	All-rates-different likelihood model
MRCA	Most recent common ancestor
CTAB	Cetyltrimethylammonium bromide
AIC	Akaike Information Criterion
MCMC	Markov Chain Monte Carlo

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12870-025-07093-6.

Additional file 1:Fig. S1. Phylogenetic relationships in of ***Aquilegia*** obtained from an maximum likelihood analysis of the 78-protein-coding gene

dataset. Fig. S2. Phylogenetic relationships in of ***Aquilegia*** obtained from an maximum likelihood analysis of the 15-introns dataset. Fig. S3. Phylogenetic relationships in of ***Aquilegia*** obtained from an maximum likelihood analysis of the concatenated dataset of 117 intergenic spacer regions. Fig. S4. Phylogenetic relationships in of ***Aquilegia*** obtained from an maximum likelihood analysis of the concatenated dataset of 15 introns and 117 intergenic spacer regions. Fig. S5. Phylogenetic relationships in of ***Aquilegia*** obtained from an maximum likelihood analysis of the large single-copy region dataset. Fig. S6. Phylogenetic relationships in of ***Aquilegia*** obtained from an maximum likelihood analysis of the inverted repeat region dataset. Fig. S7. Phylogenetic relationships in of ***Aquilegia*** obtained from an maximum likelihood analysis of the small single-copy region dataset. Fig. S8. Comparison of structure variations (SVs) in the large single-copy region of 21 representative species of ***Aquilegia.*** Fig. S9. Phylogenetic relationships in ***Aquilegia*** obtained an maximum likelihood (ML) analysis of the concatenated dataset of the ten identified hypervariable regions.

Additional file 2: Table S1. Species, voucher, distribution, and GenBank accession numbers for the samples used in this study. Table S2. Summary of types of structure variations and plastome lengths of Thalictroideae. Table S3. Summary of characteristics of genes in plastomes of Thalictroideae. Table S4. Comparison of three transition models in ancestral structure variation reconstruction. Table S5. Summary of characteristics of 121 genes in plastome sequences in ***Aquilegia***. Table S6. Summary of characteristics of 120 intergenic spacer (IGS) regions in plastome sequences in ***Aquilegia***.

Authors' contributions

WW and KLX planned and designed the research; HWP, ASE, TVE, MVB and WW collected samples; ASE identified the species; FMY and KLX performed the experiments; FMY, WCH, HWP, YYL, ASE and KLX analyzed the data; FMY, WCH, WW, ASE and KLX wrote and revised the manuscript. All authors read and approved the manuscript.

Funding

This research was partially funded by the National Key Research and Development Program of China (2024YFF1701404 and 2023YFF0805800), the National Natural Science Foundation of China (32361133549, 32170210, and 32400182), the Beijing Natural Science Foundation (5244047), Chinese Academy of Sciences President's International Fellowship Initiative (2025PVA0116), Russian Science Foundation (24-44-00027) and the state assignments performed by the CSBG SB RAS (AAAA-A21-121011290024-05) to ASE and MVB, state assignments performed by the ICG SB RAS, FWNR-2025-0020, state assignments performed by the SFSCA RAS, № 075-00563-25-02.

Data availability

All sequences in this study are available in the National Center for Biotechnology Information (NCBI) (https://www.ncbi.nlm.nih.gov/nuccore/), with GenBank accession numbers (PV546610–PV546614; PV564728–PV564802) shown in Table S1.

Declarations

Ethics approval and consent to participate

We complied with all relevant institutional, national and international guide lines with permissions from Institute of Botany, Chinese Academy of Sciences. No materials from animal or human were used in this research.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

[1]State Key Laboratory of Plant Diversity and Specialty Crops, Institute of Botany, Chinese Academy of Sciences, Beijing 100093, China
[2]China National Botanical Garden, Beijing 100093, China
[3]University of Chinese Academy of Sciences, Beijing 100049, China
[4]Shenzhen Branch, Guangdong Laboratory of Lingnan Modern Agriculture, Genome Analysis Laboratory of the Ministry of Agriculture and Rural Affairs, Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen 518120, China
[5]Central Siberian Botanical Garden, Siberian Branch of Russian Academy of Sciences, Zolotodolinskaya str. 101, Novosibirsk 630090, Russia
[6]Institute of Cytology and Genetics, Siberian Branch of Russian Academy of Sciences, Novosibirsk 630090, Russia
[7]Siberian Federal Scientific Centre of Agrobiotechnologies of the Russian Academy of Sciences (SFSCA RAS), Krasnoobsk 630501, Russia
[8]Novosibirsk State University, Novosibirsk 630090, Russia

Received: 19 May 2025 / Accepted: 18 July 2025
Published online: 20 August 2025

References

1. Ruhlman TA, Jansen RK. The plastid genomes of flowering plants. In: Maliga P, editor. Chloroplast biotechnology: Methods and protocols. Totowa: Humana Press; 2014.
2. Henriquez CL, Abdullah, Ahmed I, Carlsen MM, Zuluaga A, Croat TB, et al. Molecular evolution of chloroplast genomes in Monsteroideae (Araceae). Planta. 2020;251:72.
3. Zhao F, Li B, Drew BT, Chen YP, Wang Q, Yu WB, et al. Leveraging plastomes for comparative analysis and phylogenomic inference within Scutellarioideae (Lamiaceae). PLoS One. 2020;15:e0232602.
4. Green BR. Chloroplast genomes of photosynthetic eukaryotes. Plant J. 2011;66:34–44.
5. 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;76:273–97.
6. Sun YX, Moore MJ, Meng AP, Soltis PS, Soltis DE, Li JQ, et al. Complete plastid genome sequencing of Trochodendraceae reveals a significant expansion of the inverted repeat and suggests a Paleogene divergence between the two extant species. PLoS One. 2013;8: e60429.
7. Graham SW, Lam VKY, Merckx VSFT. Plastomes on the edge: The evolutionary breakdown of mycoheterotroph plastid genomes. New Phytol. 2017;214:48–55.
8. Peng HW, Lian L, Zhang J, Erst AS, Wang W. Phylogenomics, plastome degradation and mycoheterotrophy evolution of Neottieae (Orchidaceae), with emphasis on the systematic position and Loess Plateau-Changbai Mountains disjunction of *Diplandrorchis*. BMC Plant Biol. 2022;22:507.
9. Edlund M, Anderson BM, Su HJ, Robison T, Caraballo-Ortiz MA, Der JP, et al. Plastome evolution in Santalales involves relaxed selection prior to loss of *ndh* genes and major boundary shifts of the inverted repeat. Ann Bot. 2025;135:515–30.
10. Li CJ, Xie XT, Liu HX, Wang RN, Li DZ. Plastome evolution in the East Asian *lobelias* (Lobelioideae) using phylogenomic and comparative analyses. Front Plant Sci. 2023;14:1144406.
11. Weng M, Ruhlman TA, Jansen RK. Expansion of inverted repeat does not decrease substitution rates in *Pelargonium* plastid genomes. New Phytol. 2017;214:842–51.
12. Xiang KL, Mao W, Peng HW, Erst AS, Yang YX, He WC, et al. Organization, phylogenetic marker exploitation, and gene evolution in the plastome of *Thalictrum* (Ranunculaceae). Front Plant Sci. 2022;13: 897843.
13. Martin G, Baurens FC, Cardi C, Aury JM, D'Hont A. The complete chloroplast genome of Banana (*Musa acuminata*, Zingiberales): Insight into plastid monocotyledon evolution. PLoS One. 2013;8:e67350.
14. Blazier JC, Ruhlman TA, Weng ML, Rehman SK, Sabir JSM, Jansen RK. Divergence of RNA polymerase α subunits in angiosperm plastid genomes is mediated by genomic rearrangement. Sci Rep. 2016;6:24595.
15. Long J, He WC, Peng HW, Erst AS, Wang W, Xiang KL. Comparative plastome analysis of the sister genera *Ceratocephala* and *Myosurus* (Ranunculaceae) reveals signals of adaptive evolution to arid and aquatic environments. BMC Plant Biol. 2024;24: 202.
16. Zhai W, Duan X, Zhang R, Guo C, Li L, Xu G, et al. Chloroplast genomic data provide new and robust insights into the phylogeny and evolution of the Ranunculaceae. Mol Phylogenet Evol. 2019;135:12–21.
17. Wu H, Li DZ, Ma PF. Unprecedented variation pattern of plastid genomes and the potential role in adaptive evolution in Poales. BMC Biol. 2024;22:97.

18. Kato T, Kaneko T, Sato S, Nakamura Y, Tabata S. Complete structure of the chloroplast genome of a legume, *Lotus japonicus*. DNA Res. 2000;7:323–30.
19. Park S, Park S. Large-scale phylogenomics reveals ancient introgression in Asian *Hepatica* and new insights into the origin of the insular endemic *Hepatica maxima*. Sci Rep. 2020;10:16288.
20. Johansson JT. There large inversions in the chloroplast genomes and one loss of the chloroplast gene *rpsl6* suggest an early evolutionary split in the genus *Adonis* (Ranunculaceae). Plant Syst Evol. 1999;218:133–43.
21. Munz PA. *Aquilegia* : The cultivated and wild columbines. Ithaca: Gentes Herbarium; 1946. p. 7.
22. Tamura M. *Aquilegia*. In: In Die Naturlichen Pflanzenfamilien, vol. 17a. Berlin: Duncker and Humblot; 1995. p. 223–497.
23. Xiao PG. Aquilegia. In: Flora Reipublicae Popularis Sinicae, vol. 27. Beijing: Science Press; 1979. p. 491–501.
24. Filiault DL, Ballerini ES, Mandáková T, Aköz G, Derieg NJ, Schmutz J, et al. The *aquilegia* genome provides insight into adaptive radiation and reveals an extraordinarily polymorphic chromosome with a unique history. eLife. 2018;7:e36426.
25. Kramer EM. *Aquilegia* : A new model for plant development, ecology, and evolution. Annu Rev Plant Biol. 2009;60:261–77.
26. Shi T, Chen J. A reappraisal of the phylogenetic placement of the *Aquilegia* whole-genome duplication. Genome Biol. 2020;21:295.
27. Xie J, Zhao H, Li K, Zhang R, Jiang Y, Wang M, et al. A chromosome-scale reference genome of *Aquilegia oxysepala* var. *kansuensis*. Hortic Res. 2020;7: 113.
28. Hodges SA, Arnold ML. Spurring plant diversification: Are floral nectar spurs a key innovation? Proc R Soc Lond B Biol Sci. 1995;262:343–8.
29. Hodges SA, Arnold ML. Columbines: A geographically widespread species flock. Proc Natl Acad Sci U S A. 1994;91:5129–32.
30. Bastida JM, Alcántara JM, Rey PJ, Vargas P, Herrera CM. Extended phylogeny of *Aquilegia*: The biogeographical and ecological patterns of two simultaneous but contrasting radiations. Plant Syst Evol. 2010;284:171–85.
31. Fior S, Li M, Oxelman B, Viola R, Hodges SA, Ometto L, et al. Spatiotemporal reconstruction of the *Aquilegia* rapid radiation through next-generation sequencing of rapidly evolving cpDNA regions. New Phytol. 2013;198:579–92.
32. Zhang W, Wang H, Dong J, Zhang T, Xiao H. Comparative chloroplast genomes and phylogenetic analysis of *Aquilegia*. Appl Plant Sci. 2021;9: e11412.
33. Ro KE, McPheron BA. Molecular phylogeny of the *Aquilegia* group (Ranunculaceae) based on internal transcribed spacers and 5.8S nuclear ribosomal DNA. Biochem Syst Ecol. 1997;25:445–61.
34. Whittall JB, Hodges SA. Pollinator shifts drive increasingly long nectar spurs in columbine flowers. Nature. 2007;447:706–9.
35. Xu Y, Li Y, Chen Y, Wang L, Xue B, Zhang X, et al. Comparative analysis of complete chloroplast genomes of in China: Hypervariable regions and phylogenetic relationships. Genes. 2024;15:716.
36. Xu LS, Herrando-Moraira S, Susanna A, Galbany-Casals M, Chen YS. Phylogeny, origin and dispersal of *Saussurea* (Asteraceae) based on chloroplast genome data. Mol Phylogenet Evol. 2019;141: 106613.
37. Menezes APA, Resende-Moreira LC, Buzatti RSO, Nazareno AG, Carlsen M, Lobo FP, et al. Chloroplast genomes of species (Malpighiaceae): Comparative analysis and screening of high divergence sequences. Sci Rep. 2018;8:2210.
38. Abdullah, Mehmood F, Shahzadi I, Waseem S, Mirza B, Ahmed I, et al. Chloroplast genome of *Hibiscus rosa-sinensis* (Malvaceae): Comparative analyses and identification of mutational hotspots. Genomics. 2020;112:581–91.
39. Kim KJ, Lee HL. Complete chloroplast genome sequences from Korean ginseng (*Panax schinseng* Nees) and comparative analysis of sequence evolution among 17 vascular plants. DNA Res. 2004;11:247–61.
40. Michelangeli FA, Davis JI, Stevenson DWm. Phylogenetic relationships among Poaceae and related families as inferred from morphology, inversions in the plastid genome, and sequence data from the mitochondrial and plastid genomes. Am J Bot. 2003;90:93–106.
41. Li X, Li Y, Zang M, Li M, Fang Y. Complete chloroplast genome sequence and phylogenetic analysis of. Int J Mol Sci. 2018;19:2443. https://doi.org/10.3390/ijms19082443.
42. Liu W, Kong H, Zhou J, Fritsch P, Hao G, Gong W. Complete chloroplast genome of *Cercis chuniana* (Fabaceae) with structural and genetic comparison to six species in Caesalpinioideae. Int J Mol Sci. 2018;19:1286.
43. Sun J, Wang Y, Liu Y, Xu C, Yuan Q, Guo L, et al. Evolutionary and phylogenetic aspects of the chloroplast genome of *Chaenomeles* species. Sci Rep. 2020;10:11466.
44. Burke SV, Ungerer MC, Duvall MR. Investigation of mitochondrial-derived plastome sequences in the *Paspalum* lineage (Panicoideae; Poaceae). BMC Plant Biol. 2018;18:152.
45. CBOL Plant Working Group1, Hollingsworth PM, Forrest LL, Spouge JL, Hajibabaei M, Ratnasingham S, et al. A DNA barcode for land plants. Proc Natl Acad Sci. 2009;106:12794–7.
46. Li X, Yang Y, Henry RJ, Rossetto M, Wang Y, Chen S. Plant DNA barcoding: From gene to genome. Biol Rev. 2015;90:157–66.
47. 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;101:1987–2004.
48. Zhang Y, Du L, Liu A, Chen J, Wu L, Hu W, et al. The complete chloroplast genome sequences of five species: Lights into phylogenetic and taxonomic analyses. Front Plant Sci. 2016;7:306.
49. Ling YY, Peng HW, Lian L, Erst AS, Xiang KL, Wang W. Out of and in East Asia: Phylogeny, biogeography and diversification of Thalictroideae (Ranunculaceae) in the Northern Hemisphere. Ann Bot. 2024;134:1251–62.
50. Doyle JJ, Doyle JL. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem Bull. 1987;19:11–5.
51. Chen S, Zhou Y, Chen Y, Gu J. Fastp: An ultra-fast all-in-one FASTQ preprocessor. Bioinformatics. 2018;34:i884-90.
52. 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;21:241.
53. Wick RR, Schultz MB, Zobel J, Holt KE. Bandage: Interactive visualization of *de novo* genome assemblies. Bioinformatics. 2015;31:3350–2.
54. Tillich M, Lehwark P, Pellizzer T, Ulbricht-Jones ES, Fischer A, Bock R, et al. GeSeq – Versatile and accurate annotation of organelle genomes. Nucleic Acids Res. 2017;45:W6-11.
55. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, et al. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics. 2012;28:1647–9.
56. Greiner S, Lehwark P, Bock R. Organellar Genome DRAW (OGDRAW) version 1.3.1: Expanded toolkit for the graphical visualization of organellar genomes. Nucleic Acids Res. 2019;47:W59-64.
57. Amiryousefi A, Hyvönen J, Poczai P. IRscope: An online program to visualize the junction sites of chloroplast genomes. Bioinformatics. 2018;34:3030–1.
58. Katoh K, Toh H. Parallelization of the MAFFT multiple sequence alignment program. Bioinformatics. 2010;26:1899–900.
59. Stamatakis A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014;30:1312–3.
60 Ronquist F, Teslenko M, Mark PVD, Ayres DL, Darling A, Hohna S, et al. MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. Syst Biol. 2012;61:539–42.
61. Posada D. jModelTest: Phylogenetic model averaging. Mol Biol Evol. 2008;25:1253–6.
62. Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. Syst Biol. 2018;67:901–4.
63. Wang W, Li H, Chen Z. Analysis of plastid and nuclear DNA data in plant phylogenetics-evaluation and improvement. Sci China Life Sci. 2014;57:280–6.
64. Pennell MW, Eastman JM, Slater GJ, Brown JW, Uyeda JC, FitzJohn RG, et al. Geiger v2.0: An expanded suite of methods for fitting macroevolutionary models to phylogenetic trees. Bioinformatics. 2014;30:2216–8.
65. Revell LJ. Phytools: an R package for phylogenetic comparative biology (and other things). Methods Ecol Evol. 2012;3:217–23.
66. Rozas J, Ferrer-Mata A, Sánchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE, et al. DnaSP 6: DNA sequence polymorphism analysis of large data sets. Mol Biol Evol. 2017;34:3299–302.

Publisher's Note