The genus *Aquilegia* L. (columbine), comprising approximately 70 perennial herb species, belongs to the family Ranunculaceae and is widely distributed in North America and Eurasia (Munz, 1946). Recently, several new species were reported, bringing the number of columbine taxa to about 110 species (Erst et al., 2017, 2020; Luo et al., 2018). Although the morphologies and habitats of columbine species differ, the phylogenetic resolution of this genus at the molecular level is very low, and therefore the genus is considered to be a widespread population complex. The morphological differences of the floral spurs between species of *Aquilegia* are easily observed and attract different pollinators, which has led to the rapid divergence of the columbines to form a large number of species (Hodges and Derieg, 2009). Moreover, natural hybrids among columbine species have also been frequently reported (Taylor, 1967). As a result, *Aquilegia* species have become a model for evolution studies; however, the phylogenetic trees presented in previous studies contain multifurcations, which may be caused by a lack of informative sites (Hodges and Arnold, 1994; Bastida et al., 2010; Fior et al., 2013), complicating subsequent research on the speciation of this genus.

It is therefore very important to construct a relatively clear phylogenetic relationship of these species for future evolutionary studies. Genomic sequencing could compensate for the lack of informative sites in shorter sequences. Notably, the decline in sequencing costs in recent years has made this approach possible for all parts of the plant genome (nuclear, mitochondrial, and chloroplast). Because of the easy interspecific hybridization among *Aquilegia* species, the nuclear genome structure is complex, with a high recombination rate (Filiault et al., 2018). The mitochondrial genomes of the angiosperms are relatively complex; the order of genes differs among species, and only some regions of the genome are conserved (Kubo et al., 2000). In contrast, the monophyletic inheritance of the chloroplast genome sequence is more suitable for the phylogenetic analysis of *Aquilegia* due to its low recombination rate and high level of conservation (Dong et al., 2012; Curci et al., 2015; Downie and Jansen, 2015; Nadachowska-Brzyska et al., 2015). Fior et al. (2013) selected 21 chloroplast genes with rapid evolutionary rates to establish the phylogenetic relationships among *Aquilegia* species. Although the topology of this phylogeny had a lower resolution and...
support for some branches (Fior et al., 2013) than previously constructed trees based on fewer chloroplast sequences (Hodges and Arnold, 1994; Bastida et al., 2010), the resolution and support rate were improved. Hence, the complete chloroplast genome sequence is an ideal molecular marker for inferring the phylogenetic relationships of the *Aquilegia* genus.

The chloroplast genome is a closed-loop structure approximately 115–210 kbp in size, and generally consists of four parts: two inverted repeat regions (IRA and IRB), a large single-copy region (LSC), and a small single-copy region (SSC) (Yurina and Odintsova, 1998; Park et al., 2018). Some plant groups have special chloroplast genome structures, such as species of the genus *Erodium* L’, which lack the IR regions (Guisinger et al., 2010). Because of its stable genomic structure, identical gene content, and conserved sequence (Dong et al., 2012), the chloroplast genome is used as a molecular marker for the inference of phylogenetic relationships (Li et al., 2018; Liu et al., 2018; Lu et al., 2018; Mader et al., 2018; Xie et al., 2018) and adaptative evolution (Dong et al., 2018; Fan et al., 2018). In this study, we assembled and analyzed the chloroplast genomes of 14 columbine species from Asia, Europe, and North America, and constructed a phylogenetic tree of the genus to shed light on radiative speciation in *Aquilegia* and lay a foundation for inferring the evolutionary history of the columbines.

**METHODS**

**Plant materials**

Seeds of *A. amurensis* Kom., *A. ecalcarata* Maxim., *A. oxysepala* Trautv. & C. A. Mey. var. *kansensis* Brühl, *A. parviflora* Ledeb., *A. rockii* Munz, *A. viridiflora* Pall., and *A. yabeana* Kitag. were collected from China (Appendix 1), and all voucher specimens were deposited in the Northeast Normal University Herbarium in Changchun, China (accession numbers NENU_Aq1001–NENU_Aq1007). Seeds were grown in the greenhouse of Northeast Normal University with 12 h of light at 25°C and 12 h of dark at 20°C.

**DNA extraction and sequencing**

Total genomic DNA was extracted from fresh leaves using a modified cetyltrimethylammonium bromide (CTAB) method (Doyle and Doyle, 1987). Genomic library generation and sequencing were used to acquire 2 × 150-bp paired reads generated on the Illumina sequencer (Illumina San Diego, CA, USA). Seeds were grown in the greenhouse of Northeast Normal University with 12 h of light at 25°C and 12 h of dark at 20°C.

**Chloroplast genome assembly and annotation**

To obtain high-quality genome sequences, all reads were filtered as follows: remove reads containing adapters, a content of more than 10% N, or more than 50% low-quality bases (quality value <10). We then used the chloroplast_assembly_protocol pipeline to assemble the chloroplast genome (Sancho et al., 2018). Briefly, DUK (http://duk.sourceforge.net) was used to extract the chloroplast reads, which were filtered using FASTQC version 0.10.1 (Andrew, 2010) and Trimmomatic version 0.32 (Bolger et al., 2014). Next, the pass-filtered reads were de novo assembled using Velvet version 1.2.07 (Zerbino, 2010), SSPACE Basic version 2.0 (Boetzer et al., 2011), and GapFiller version 1.11 (Boetzer and Pirovano, 2012; Nadalin et al., 2012), with annotation performed using the online program DOGMA (Wyman et al., 2004). Finally, the circular genome map of *Aquilegia* was illustrated using the Organellar Genome DRAW tool (Lohse et al., 2013) after manually checking the annotation results.

**Repeat sequence characterization**

The Perl script MISA (Thiel et al., 2003) was employed to identify the location of simple sequence repeat (SSR) loci in the complete chloroplast genome sequences. The thresholds used to detect the SSRs were 10, 5, 4, 3, 3, and 3 for mono-, di-, tri-, tetra-, penta-, and hexanucleotides, respectively. The recognition results were checked manually, and the redundant results were removed. REPuter (Kurtz et al., 2001) was then used to identify repeat sequences in the chloroplast, including palindromic, forward, reverse, and complementary sequences. The parameters were set as follows: (1) Hamming distance of 3, (2) 90% or greater sequence identity, and (3) a minimum repeat size of 30 bp. The default settings were used for all other parameters.

**Genetic divergence and phylogenetic analysis of *Aquilegia***

The homologous genes were extracted from 14 *Aquilegia* species using a Python script (available on GitHub, see Data Availability Statement), after which these homologous genes were aligned using MAFFT version 7.407 (Katoh and Standley, 2013) with the default settings. Furthermore, the nucleotide diversity (π) of these homologous genes was analyzed using DnaSP version 6.0 (Rozas et al., 2017). To avoid the effect of sequence redundancy when building the phylogenetic trees, we selected the LSC regions, IRB regions, and SSC regions as arrays. In addition, the published chloroplast genome sequences of *A. rockii* (MK573514.1, NC_033341.1), *A. ecalcarata* (NC_041528.1, MK569474.1), and *A. coerulescens* (NC_041527.1, MK569492.1) in GenBank were used. *Semiaquilegia adoxoides* Makino (MH142265.2) was considered as the outgroup (Fior et al., 2013; Zhai et al., 2019). The alignment was aligned using MAFFT version 7.407 and was adjusted manually in CLC Sequence Viewer 8.0 (QIAGEN Digital Insights, Redwood City, California, USA). The maximum likelihood tree was generated using IQ-TREE version 1.6.12 using 1000 bootstrap replicates (Nguyen et al., 2015). Meanwhile, the Bayesian inference trees were produced using MrBayes version 3.2 (Ronquist et al., 2012), based on Markov chain Monte Carlo analyses run for 1,000,000 generations. These trees were sampled every 1000 generations with the first 250 trees discarded in the burn-in period. The program was stopped when the standard deviation was less than 0.01. The final tree was visualized in iTOl (https://itol.embl.de/itol.cgi) (Letunic and Bork, 2006).

**Natural selection analysis**

To identify genes under selection in *Aquilegia*, the genes of the chloroplast genomes were analyzed with the PAML package (Yang, 2007). First, all coding sequences (CDS) of the *Aquilegia* species and other Ranunculaceae species were extracted from the genome sequences.
using a Python script (Appendix 3). Each single-copy sequence was aligned according to its codons using MEGA X (Kumar et al., 2018) and checked manually, and then used as input for CodeML in the PAML package. Moreover, the concatenated alignment was also used to construct phylogenetic relationships among species using IQ-TREE version 1.6.12 (Nguyen et al., 2015). Finally, each CDS alignment was used to calculate the nonsynonymous (dN) and synonymous (dS) substitution rates, along with their ratio (\(\omega = \frac{dN}{dS}\)). \(\omega > 1\) indicates positive selection, \(\omega = 1\) indicates neutral selection, and \(\omega < 1\) indicates negative selection (Yang and Nielsen, 2002). The branch-site model (X. Yang et al., 1998; Z. Yang et al., 1998) was combined with the naive empirical Bayes (NEB) method, and the Bayesian empirical Bayes (BEB) method was used to identify potential positively selected genes using CodeML in the PAML.

**FIGURE 1.** Gene maps of the *Aquilegia viridiflora* chloroplast genome. Genes inside the circle are transcribed clockwise, while genes outside are transcribed counterclockwise (as indicated by arrows). Different colors indicate different functional groups. The dark gray shading within the inner circle corresponds to the GC content and the light gray shading corresponds to the AT content. IRA and IRB, inverted repeat regions; LSC, large single-copy region; ORF, open reading frame; SSC, small single-copy region.
RESULTS

Features of Aquilegia chloroplast genomes

The complete chloroplast genomes of the Aquilegia species from Asia, North America, and Europe displayed a typical quadripartite structure similar to the majority of land plant chloroplast genomes (Fig. 1). The sizes of the complete chloroplast genomes ranged from 157,689 to 161,387 bp. All complete chloroplast genomes were composed of four sections, including an LSC region (86,761–88,076 bp), an SSC region (17,466–18,879 bp), and two IR regions (25,612–28,015 bp). The GC content of the 14 species was similar in both the whole chloroplast genome (38.94%–39.08%) and the corresponding regions (LSC [37.43%–37.71%], SSC [33.30%–33.91%], and IR [43.04%–43.41%]), with the IR regions having the highest GC contents (Table 1). These sequence data are available in GenBank (accession numbers MT919110–MT9191116 and MN809218–MN809224).

The chloroplast genomes of the Aquilegia species contained 154 genes (96 protein-coding genes, 48 transfer RNA [tRNA] genes, and eight ribosomal RNA genes). Most of the genes located in the LSC and SSC regions were single copy, while 26 of the genes located in the IR regions were duplicated, including 11 protein-coding genes (rps7, rps12, rps19, rpl2, rpl23, orf42, orf56, ycf2, ycf15, ycf68, and ndhB), 11 tRNA genes (trnI-CAU [x3], trnL-CAA, trnG-UCC, trnV-UAC, trnV-GAC, trnL-AUU, trnN-GUU), and four rRNA genes (rrn4.5, rrn5, rrn16, and rrn23). The LSC region comprises 63 protein-coding genes and 25 tRNA genes, and the SSC region comprises 13 protein-coding genes and a single rRNA gene. Among all the genes, seven protein-coding genes (rpoC1, atpF, rpl2, ycf68, ndhB, ndhF, and ndhA) contained only one intron, while one protein-coding gene (ycf3) contained two introns (Appendix S1).

Repeat analysis

We identified a range of 84–89 repeat sequences in the 14 Aquilegia chloroplast genomes, including 45–51 palindromic repeats and 33–44 forward repeats; reverse and complement repeats were not identified (Fig. 2A). In all species, the palindromic repeats were 56–398 bp in length and the forward repeats were 56–357 bp in length (Fig. 2B, C). The SSR analysis of the 14 species identified (Fig. 2A). In all species, the palindromic repeats were 56–398 bp in length and the forward repeats were 56–357 bp in length (Fig. 2B, C). The SSR analysis of the 14 species identified a range of 69–84 microsatellites of six types; A. chrysantha and A. viridiflora had the lowest and highest numbers of microsatellites, respectively (Fig. 3A). Among all SSRs, the most abundant type was mononucleotide repeats, which accounted for 66.51% of the total SSRs, followed by dinucleotide (13.32%), tetranculotide (7.22%), trinucleotide (5.91%), pentanucleotide (4.32%), and hexanucleotide (2.72%) repeats. AT repeats accounted for a larger proportion of mononucleotide

| Species         | LSC Length (bp) | GC (%) | IRs Length (bp) | GC (%) | SSC Length (bp) | GC (%) | Total Length (bp) | GC (%) |
|-----------------|-----------------|--------|-----------------|--------|-----------------|--------|-------------------|--------|
| A. aurea        | 87,986          | 37.52  | 18,169          | 33.47  | 18,474          | 33.64  | 159,601          | 39.00  |
| A. japonica     | 87,588          | 37.60  | 17,482          | 33.38  | 17,482          | 33.38  | 159,601          | 39.00  |
| A. oxysepala    | 88,053          | 37.44  | 17,466          | 33.64  | 17,466          | 33.64  | 159,601          | 39.00  |
| A. sibirica     | 87,655          | 37.63  | 18,638          | 33.64  | 18,638          | 33.64  | 159,289          | 39.05  |
| A. kansuensis   | 87,651          | 37.43  | 18,474          | 33.47  | 18,474          | 33.47  | 159,131          | 38.96  |
| A. yabeana      | 88,030          | 37.60  | 17,466          | 33.64  | 17,466          | 33.64  | 159,601          | 39.00  |
| A. amurensis    | 87,375          | 37.63  | 18,724          | 33.50  | 18,724          | 33.50  | 159,667          | 38.96  |
| A. parviflora   | 87,371          | 37.52  | 17,466          | 33.64  | 17,466          | 33.64  | 159,667          | 38.96  |
| A. chrysantha   | 87,662          | 37.44  | 17,466          | 33.64  | 17,466          | 33.64  | 159,667          | 38.96  |
| A. formosa      | 87,865          | 37.52  | 18,724          | 33.50  | 18,724          | 33.50  | 159,667          | 38.96  |
| A. rockii       | 87,375          | 37.63  | 18,724          | 33.50  | 18,724          | 33.50  | 159,667          | 38.96  |
| A. viridiflora  | 88,076          | 37.61  | 18,662          | 33.64  | 18,662          | 33.64  | 159,667          | 38.96  |
| A. amurensis    | 87,655          | 37.63  | 18,724          | 33.50  | 18,724          | 33.50  | 159,667          | 38.96  |
| A. parviflora   | 87,371          | 37.52  | 17,466          | 33.64  | 17,466          | 33.64  | 159,667          | 38.96  |
| A. ecalcarata   | 87,865          | 37.52  | 18,724          | 33.50  | 18,724          | 33.50  | 159,667          | 38.96  |
| A. sibirica     | 87,651          | 37.43  | 18,474          | 33.47  | 18,474          | 33.47  | 159,131          | 38.96  |

Note: IRs = inverted repeat regions; LSC = large single-copy region; NCBI = National Center for Biotechnology Information; SSC = small single-copy region.

Raw data were downloaded from NCBI.
repeats (92.95%) than GC repeats (7.05%). Similarly, the AT content (90.15%) accounted for a larger proportion than the GC content (9.85%) in dinucleotides (Fig. 3B, Appendix S2). Not surprisingly, all SSRs were detected in noncoding regions of the Aquilegia chloroplast genome.

Sequence divergence and phylogeny of Aquilegia

The \( \pi \) value was used to evaluate sequence divergence in Aquilegia chloroplast genomes. In genic regions, the range of variation in \( \pi \) was 0–0.00511, with a mean of 0.00061; \( \pi \) of the LSC region (0–0.00511, with a mean of 0.00055) was higher than in other regions (0–0.00453 in the IR regions, with a mean of 0.00041; 0–0.00252 in the SSC region, with a mean of 0.0013). Overall, these results demonstrated that the sequence divergence in Aquilegia chloroplast genomes was small, but some regions showed high genetic diversity, such as rpoC2, trnS-GGA, and trnL-CAA (\( \pi > 0.004 \)) (Fig. 4, Appendix S3).

To reveal the phylogeny of Aquilegia, aligned chloroplast genome sequences were used to construct phylogenetic trees using both maximum likelihood and Bayesian analyses. The two resulting trees showed identical topologies, and the bootstrap values and posterior probabilities were very high for each lineage. The Aquilegia species were divided into two clades: one clade contained A. aurea and A. vulgaris from Europe and A. sibirica, A. oxysepala var. oxysepala, A. japonica, A. ecalcarata, A. rockii, A. viridiflora, A. yabeana and A. oxysepala var. kansuensis from Asia; the other clade contained A. formosa, A. chrysantha, and A. coerulea from North America and A. amurensis and A. parviflora from Asia. All the topologies supported A. japonica and A. oxysepala as sister clades, and A. sibirica shared a common ancestor with them. Interestingly, the A. ecalcarata sequence assembled by us clustered with A. rockii, while the A. ecalcarata sequence downloaded from GenBank was grouped with A. yabeana and A. oxysepala var. kansuensis. In addition, A. viridiflora formed a single clade with A. ecalcarata and A. rockii. Although A. oxysepala var. oxysepala and A. oxysepala var. kansuensis are considered varieties of the same species, they were found in two different clades. Similarly, A. japonica and A. amurensis, which are treated as a single species by the Flora of China (Li, 2007), were also found in two different clades (Fig. 5).

Positive selection analysis

Positive selection tests were performed on 54 CDS from Aquilegia and their related species using the PAML package. No significant selection was found to act on the chloroplast genes of Aquilegia (\( P > 0.05 \)), but three genes with a higher posterior probability were
detected using the BEB and NEB methods (atpB, petG, and rpl36). Therefore, atpB, petG, and rpl36 were considered to be genes potentially under positive selection (Table 2).

DISCUSSION

The structure of Aquilegia chloroplast genomes

In this study, we assembled and annotated the complete chloroplast genomes of 14 Aquilegia species, including 10 species from Asia, two from Europe, and two from North America. Based on these chloroplast genome sequences, we calculated polymorphism and inferred the phylogenetic relationships within Aquilegia.

The structure and gene order of chloroplast genomes are highly conserved in the angiosperms (Choi et al., 2016). In our study, the chloroplast genomes of 14 Aquilegia species showed a typical quadripartite structure (Fig. 1), and the gene composition and gene order were similar in each species. The expansion or contraction of IR regions plays an important role in the length of the chloroplast genome (Raubeson et al., 2007; Wang et al., 2008; Yang et al., 2010). In the Aquilegia chloroplast genomes, the total length of the complete sequence was directly proportional to the length of the IR region (Table 1). Insertion/deletion polymorphisms (indels) in these sequences resulted in variations in the length of the Aquilegia chloroplast genome, which is a common phenomenon found in Camellia L. (Huang et al., 2014), Quercus L. (Yin et al., 2018), Amaranthus L. (Chaney et al., 2016), and the other angiosperms (Jiang et al., 2017). Compared with the other two regions, the GC content was the highest in the IR regions in Aquilegia. This effect may be caused by the presence of more rDNA in the IR regions, which has a higher GC content (approximately 50%) (Xie et al., 2018).

Both long repetitive sequences and SSRs with high copy-number diversity are valuable and useful molecular markers in studies of plant population genetics, phylogenetic reconstruction, and plant evolution at the intraspecific level (Wu et al., 2015; Ivanova et al., 2017). Here, long repeat sequences and SSRs of different lengths were found in each species (Figs. 2, 3), indicating that they can both be used as molecular markers for research on Aquilegia. Among these regions, the SSC region had the highest nucleotide polymorphism level, followed by the LSC region; the IR regions had the lowest nucleotide polymorphism level, indicating that the IR regions were most conserved. This result is likely due to the high conservation of the rDNA in the IR regions (Hershkovitz and Zimmer, 1996). The nucleotide polymorphisms of chloroplast genes in Aquilegia were smaller than those of other genera, such as Populus L. (Gao et al., 2019), Camellia (Li et al., 2019a), and Anguinum Fourr. (Jin et al., 2019); however, some variable genes were identified, including rpoC2, trnS-GGA, and trnL-CAA (Fig. 4). These regions with high
levels of polymorphism are also a good resource for studying the phylogeny and population genetics of *Aquilegia*, especially rpoC2, which has the highest levels of polymorphism (Walker et al., 2019).

### The phylogeny of *Aquilegia* based on chloroplast genomes

Biogeographic and phylogenetic analyses have indicated that *Aquilegia* had a common ancestor from eastern Asia, and later adaptive radiations took place independently in North America and Western Europe (Bastida et al., 2010; Fior et al., 2013). *Aquilegia amurensis* is restricted to the northern Greater Khingan Mountains, while *A. parviflora* is distributed in the northern Greater Khingan Mountains and Siberia. Despite this, we found these species were phylogenetically close to *Aquilegia* species from North America, whereas the remaining Asian species were phylogenetically close to *Aquilegia* species from Europe.

The phylogeny based on the chloroplast genome was not completely consistent with that of the study by Fior et al. (2013). In our study, *A. oxysepala* var. *oxysepala*, *A. japonica*, and *A. sibirica* fell within a single clade; however, Filiault et al. (2018) had concluded that *A. oxysepala* var. *oxysepala* was located at the base of the phylogenetic tree, and *A. japonica* and *A. sibirica* shared a most recent common ancestor (MRCA). Li et al. (2014) used a combination of morphological characteristics, habitat type, and nuclear and chloroplast phylogenies (Bastida et al., 2010; Fior et al., 2013; Li et al., 2014) of these three species to propose that *A. sibirica* diverged first from the MRCA, and *A. oxysepala* var. *oxysepala* and *A. japonica* then differentiated into new species (Li et al., 2019b) containing...
more individuals. Our results also support the research of Li et al. (2019b). In addition, the position of *A. viridiflora* in this study was inconsistent with the phylogeny based on chloroplast genes by Fior et al. (2013) and the phylogeny by Lu et al. (2019). The inconsistency may be caused by incomplete lineage sorting and introgression in species undergoing rapid adaptive radiation (Meyer et al., 2017; Cai et al., 2020); therefore, the taxonomic status of *A. viridiflora* is worthy of further study. In addition, according to the *Flora of China* (Li,

| Gene name | lnL0 | lnL1 | df | P | BEB | NEB |
|-----------|-----|-----|----|---|-----|-----|
| *psbM*    | −293.77830 | −292.80279 | 1 | 0.08124 | NA | NA |
| *psbL*    | −281.27366 | −281.27366 | 1 | 0.5 | NA | NA |
| *ccsA*    | −6608.03557 | −6608.03529 | 1 | 0.5 | NA | NA |
| *psaC*    | −927.93380 | −927.93380 | 1 | 0.14717 | NA | NA |
| *psaB*    | −7228.74873 | −7228.74875 | 1 | 0.49748 | NA | NA |
| *rpl33*   | −957.68785 | −957.68785 | 1 | 0.5 | NA | NA |
| *psbF*    | −281.15093 | −281.15093 | 1 | 0.5 | NA | NA |
| *psaL*    | −404.70806 | −404.70806 | 1 | 0.5 | NA | NA |
| *atpD*    | −2841.63959 | −2841.63959 | 1 | 0.49944 | NA | NA |
| *atpH*    | −756.68830 | −756.68830 | 1 | 0.5 | NA | NA |
| *rps19*   | −1282.74781 | −1282.74822 | 1 | 0.49831 | NA | NA |
| *atpE*    | −1748.33822 | −1748.33822 | 1 | 0.09889 | NA | NA |
| Note: A = alanine (amino acid); BEB = Bayesian empirical Bayes; NA = not available; NEB = naive empirical Bayes.

Genes under positive selection.

*P* > 99%.
Adaptive evolution of *Aquilegia*

Synonymous and nonsynonymous nucleotide substitution patterns play an important role in adaptive evolution. In *Aquilegia*, no significant positive selection was detected for the majority of genes, with only three genes (*petG*, *rpl36*, and *atpB*) showing possible positive selection; these may have played an important role in adaptive evolution in *Aquilegia*. Based on annotation information from the UniProtKB database (https://www.uniprot.org), in Arabidopsis thaliana (L.) Heynh., *petG* controls the components of the cytochrome b6-f complex subunit 5, which mediates electron transfer between photosystem II (PSII) and PSI, cyclic electron flow around PSI, and state transitions (Sato et al., 1999; Kandlbinder et al., 2004); the *rpl36* gene encodes the 50S ribosomal protein L36, which serves as a structural component of the ribosome (Sato et al., 1999; Koia et al., 2013); and the *atpB* gene controls the ATP synthase subunit beta, which produces ATP from ADP in the presence of a proton gradient across the membrane (Sato et al., 1999; Friso et al., 2004). Previous studies showed that *rpl36* was under positive selection in the Araceae and Sophora tonkinensis Gagnep. (Fan et al., 2020; Henriquez et al., 2020), while *atpB* was under positive selection in Urophysa Ulbr. and the Liliaceae (sensu lato) (Xie et al., 2018; She et al., 2020). These genes are highly correlated with physiological processes such as photosynthesis and disease resistance; thus, their positive selection may assist *Aquilegia* species in rapid adaptation to various environments and enable their wide global distribution.

ACKNOWLEDGMENTS

The authors thank the reviewers for their valuable suggestions and comments and acknowledge Mingzhou Sun and Xiaoxue Fang (Key Laboratory of Molecular Epigenetics, Northeast Normal University, Ministry of Education) for their help in collection of material. This research was supported by the National Natural Science Foundation of China (32070244).

AUTHOR CONTRIBUTIONS

X.H. and W.H. designed the study and evaluated the results; W.H. and Z.W. collected the materials; Z.W., D.J., and Z.T. participated in the data analysis; Z.W. and W.H. prepared the manuscript; and all authors read and approved the final manuscript.

DATA AVAILABILITY

Raw sequence data is available from the National Center for Biotechnology Information Sequence Read Archive under the accession number PRJNA666554. Chloroplast sequence data are available in GenBank (accession numbers MT919110–MT919116 and MN809218–MN809224). The Python script used for extraction of homologous genes is available on GitHub (https://github.com/ zhangw348/NENU_plant-systems-and-evolution).

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

APPENDIX S1. List of genes encoded by the *Aquilegia* chloroplast genomes.

APPENDIX S2. Number of each type of simple sequence repeat in *Aquilegia* species.

APPENDIX S3. The nucleotide diversity of all genes of *Aquilegia*.

LITERATURE CITED

Andrews, S. 2010. FastQC: A quality control tool for high throughput sequence data. Website http://www.bioinformatics.babraham.ac.uk/projects/fastqc/ [accessed 19 January 2021].

Bastida, J. M., J. M. Alcántara, P. J. Rey, P. Vargas, and C. M. Herrera. 2010. Extended phylogeny of *Aquilegia*: The biogeographical and ecological patterns of two simultaneous but contrasting radiations. *Plant Systematics and Evolution* 284: 171–185.

Boetzer, M., C.-V. Henkel, H. J. Jansen, D. Butler, and W. Pirovano. 2011. Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics* 27(4): 578–579.

Boetzer, M., and W. Pirovano. 2012. Toward almost closed genomes with GapFiller. *Genome Biology* 13(6): R56.

Bolger, A. M., M. Lohse, and B. Usadel. 2014. Trimmmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* 30(15): 2114–2120.

Cai, L., Z. Xi, E. M. Lemmon, A. R. Lemmon, A. Mast, C. E. Buddenhagen, L. Liu, and C. C. Davis. 2020. The perfect storm: Gene tree estimation error, incomplete lineage sorting, and ancient gene flow explain the most recalcitrant ancient angiosperm clade, Malpighiales. *Systematic Biology* https://doi.org/10.1093/sysbio/syaa083.

Chaney, L., R. Mangelson, T. Ramaraj, E. N. Jellen, and P. J. Maughan. 2016. The complete chloroplast genome sequences for four *Amaranthus* species (*Amaranthaceae*). *Applications in Plant Sciences* 4(9): 1600063.

Choi, K. S., M. G. Chung, and S. Park. 2016. The complete chloroplast genome sequences of three Veronicaceae species (*Plantaginaceae*): Comparative analysis and highly divergent regions. *Frontiers in Plant Science* 7: 355.

Curci, P. L., D. De Paola, D. Danzi, G. G. Vendramin, and G. Sonnante. 2015. Complete chloroplast genome of the multifunctional crop globe artichoke and comparison with other Asteraceae. *PLoS ONE* 10(3): e0120589.

Dong, W., J. Liu, J. Yu, L. Wang, and S. Zhou. 2012. Highly variable chloroplast markers for evaluating plant phylogeny at low taxonomic levels and for DNA barcoding. *PLoS ONE* 7(4): e35071.

Dong, W.-L., R.-N. Wang, N.-Y. Zhang, W.-B. Fan, M.-F. Fang, and Z.-H. Li. 2018. Molecular evolution of chloroplast genomes of orchid species: Insights into phylogenetic relationship and adaptive evolution. *International Journal of Molecular Sciences* 19(3): 716.

Downie, S. R., and R. K. Jansen. 2015. A comparative analysis of whole plastid genomes from the Apiales: Expansion and contraction of the inverted repeat,
mitochondrial to plastid transfer of DNA, and identification of highly divergent noncoding regions. Systematic Botany 40(1): 336–351.

Doyle, J. J., and I. L. Doyle. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochemical Bulletin 19: 11–15.

Erst, A. S., W. Wang, S.-X. Yu, K. Xiang, J. Wang, D. N. Shaulo, V. S. Smirnov, et al. 2017. Two new species and four new records of *Aquilegia* (Ranunculaceae) from China. Phytotaxa 316(2): 121–137.

Erst, A. S., C. A. Pendry, T. V. Erst, H. Ikeda, K. Xiang, and W. Wang. 2020. Two new taxon and one new record of *Aquilegia* (Ranunculaceae) from India and Pakistan. Phytotaxa 439(2): 108–118.

Fan, W.-B., Y. Wu, J. Yang, K. Shahzad, and Z.-H. Li. 2018. Comparative chloroplast genomics of Dipsacales species: Insights into sequence variation, adaptive evolution, and phylogenetic relationships. Frontiers in Plant Science 9: 689.

Fan, W., D. Tang, W. Kunhua, Q. Fang, L. Li, L. Yang, Z. Yanxia, et al. 2020. The complete chloroplast genome sequence of the medicinal plant *Sophora tonkinensis*. Scientific Reports 10(1): 12473.

Filialau, D. L., E. S. Ballerini, T. Mandakova, G. Aköz, N. J. Derieg, J. Schmutz, J. Jenkins, et al. 2018. The *Aquilegia* genome provides insight into adaptive radiation and reveals an extraordinarily polymorphic chromosome with a unique history. elife 7: e36426.

Fior, S., M. Li, B. Oxelman, R. Viola, S. A. Hodges, L. Ometto, and C. Varotto. 2013. Spatiotemporal reconstruction of the *Aquilegia* rapid radiation through next-generation sequencing of rapidly evolving cp DNA regions. New Phytologist 198(2): 579–592.

Friso, G., L. Giacomelli, A. J. Ytterberg, J.-B. Peltier, A. Rudella, Q. Sun, and K. J. Van Wijk. 2004. In-depth analysis of the thylakoid membrane proteome of *Arabidopsis thaliana* chloroplasts: New proteins, new functions, and a plastid proteome database. The Plant Cell 16(2): 478–499.

Gao, K., J. Li, W. U. Khan, T. Zhao, X. Yang, X. Yang, B. Guo, and X. An. 2019. Comparative genomic and phylogenetic analyses of *Populus* section *Leuce* using complete chloroplast genome sequences. Tree Genetics & Genomes 15(3): 32.

Guisinger, M. M., J. V. Kuehl, J. L. Boore, and R. K. Jansen. 2010. Extreme reconfiguration of plastid genomes in the angiosperm family Geraniaceae: Rearrangements, repeats, and codon usage. Molecular Biology and Evolution 28(1): 583–600.

Henriquez, C. L., F. Mehmoood, I. Shahzadi, Z. Ali, M. T. Waheed, T. B. Croat, P. Fior, S., M. Li, B. Oxelman, R. Viola, S. A. Hodges, L. Ometto, and C. Varotto. 2013. Spatiotemporal reconstruction of the *Aquilegia* rapid radiation through next-generation sequencing of rapidly evolving cp DNA regions. New Phytologist 198(2): 579–592.

Huang, H., C. Shi, Y. Liu, S.-Y. Mao, and L.-Z. Gao. 2014. Thirteen plastid genome analyses and contribution to the understanding of chloroplast phylogeny and adaptive evolution in subgenus *Anguinum*. Russian Journal of Genetics 55(7): 872–884.

Kandindiner, A., I. Finkemeier, D. Wormuth, M. Hanitzsch, and K. J. Dietz. 2004. The antioxidant status of photosynthesizing leaves under nutrient deficiency: Redox regulation, gene expression and antioxidant activity in *Arabidopsis thaliana*. Physiologia Plantarum 120(1): 63–73.

Katoh, K., and D. M. Standley. 2013. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. Molecular Biology and Evolution 30(4): 772–780.

Koia, J., R. Moyle, C. Hendry, L. Lim, and J. R. Botella. 2013. Pineapple translation factor SU11 and ribosomal protein L36 promoters drive constitutive transgene expression patterns in *Arabidopsis thaliana*. Plant Molecular Biology 81(4–5): 327–336.

Kubo, T., S. Nishizawa, A. Sugawara, N. Itochda, A. Estiati, and T. Mikami. 2000. The complete nucleotide sequence of the mitochondrial genome of sugar beet (*Beta vulgaris* L.) reveals a novel gene for RNaseA (GCA). Nucleic Acids Research 28(13): 2571–2576.

Kumar, S., G. Stecher, M. Li, C. Knyaz, and K. Tamura. 2018. MEGA X: Molecular evolutionary genetics analysis across computing platforms. Molecular Biology and Evolution 35(6): 1547–1549.

Kurtz, S., J. V. Choudhuri, E. Ohlebusch, C. Schleiermacher, J. Stoye, and R. Giegerich. 2001. REPfinder: The method applications of repeat analysis on a genomic scale. Nucleic Acids Research 29(22): 4633–4642.

Letunic, I., and P. Bork. 2006. Interactive Tree of Life (ITOIL): An online tool for phylogenetic tree display and annotation. Bioinformatics 23(1): 127–128.

Li, J. 2007. Flora of China. Harvard Papers in Botany 13(2): 301–303.

Li, L. F., H. Y. Wang, D. Pang, Y. Liu, B. Liu, and H. X. Xiao. 2014. Phenotypic and genetic evidence for ecological speciation of *Aquilegia japonica* and *A. oxysepala*. New Phytologist 204(4): 1028–1040.

Li, X., Y. Li, M. Zhang, M. Li, and Y. Fang. 2018. Complete chloroplast genome sequence and phylogenetic analysis of *Quercus acutissima*. International Journal of Molecular Sciences 19(8): 2443.

Li, W., C. Zhang, X. Guo, Q. Liu, and K. Wang. 2019a. Complete chloroplast genome of *Camellia japonica* genome structures, comparative and phylogenetic analysis. PLoS ONE 14(5): e0216645.

Li, M.-R., H.-Y. Wang, N. Ding, T. Liu, Y.-C. Huang, H.-X. Xiao, B. Liu, and L.-F. Li. 2019b. Rapid divergence followed by adaptation to contrasting ecological niches of two closely related cambium species *Aquilegia japonica* and *A. oxysepala*. Genome Biology and Evolution 11(3): 919–930.

Liu, W., H. Kong, J. Zhou, P. Fritsch, G. Hao, and W. Gong. 2018. Complete chloroplast genome of *Ceris chuniana* (Fabaceae) with structural and genetic comparison to six species in Caesalpinioideae. International Journal of Molecular Sciences 19(5): 1286.

Lohse, M., O. Drechsel, S. Kahlau, and R. Bock. 2013. OrganellarGenome DRAW—A suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets. Nucleic Acids Research 41(W1): W575–W581.

Lu, Q., W. Ye, R. Lu, W. Wu, and Y. Qiu. 2018. Phylogenomic and comparative analyses of complete plastomes of *Cromia* and *Stemono* (Stemonaceae). International Journal of Molecular Sciences 19(8): 2383.

Lu, T.-M., R.-Li, N. Ding, Z.-H. Wang, L.-Z. Lan, X. Gao, and L.-F. Li. 2019. Genetic and epigenetic signatures associated with the rapid radiation of *Aquilegia* species. bioRxiv: 782821 [Preprint] [published 7 December 2019]. Available at https://doi.org/10.1101/782821 [accessed 19 January 2021].

Luo, Y., A. S. Erst, C.-X. Yang, J.-P. Deng, and L. Li. 2018. *Aquilegia yangi* (Ranunculaceae), a new species from western China. Phytotaxa 348(4): 289–296.

Mader, M., B. Pakull, C. Blanc-Jolivet, M. Paulini-Drewes, Z. Bouda, B. Degen, I. Small, and B. Kersten. 2018. Complete chloroplast genome sequences of four Meliaceae species and comparative analyses. International Journal of Molecular Sciences 19(3): 701.

Meyer, B. S., M. Matschiner, and W. Salzburger. 2017. Disentangling incomplete lineage sorting and introgression to refine species-tree estimates for Lake Tanganyika cichlid fishes. Systematic Biology 66(4): 531–550.

Munz, P. A. 1946. *Aquilegia*: The cultivated and the wild columbines. Gentes Herbarum 7: 1–50.
| Species                 | Latitude (°N) | Longitude (°E) | Distribution region | Size (Gbp) | Raw reads | Chloroplast reads | Depth    | Voucher specimen |
|------------------------|---------------|----------------|---------------------|------------|-----------|-------------------|----------|------------------|
| *A. viridiflora*       | 40.954        | 111.672        | Asia                | 13         | 18,729,599 | 9,936,532         | 4774x     | NENU_Aq1001      |
| *A. oxysepala var. kansuensis* | 31.815        | 109.009        | Asia                | 11         | 16,161,175 | 3,273,451         | 1519x     | NENU_Aq1002      |
| *A. ecalcarata*        | 37.160        | 102.223        | Asia                | 11         | 16,159,854 | 3,721,439         | 1875x     | NENU_Aq1003      |
| *A. parviflora*        | 50.422        | 121.476        | Asia                | 9.6        | 14,222,775 | 3,179,153         | 1517x     | NENU_Aq1004      |
| *A. amurensis*         | 52.672        | 123.870        | Asia                | 9.9        | 14,758,620 | 6,110,285         | 2874x     | NENU_Aq1005      |
| *A. rockii*            | 33.9125       | 121.1041       | Asia                | 11         | 15,537,263 | 3,696,958         | 1664x     | NENU_Aq1006      |
| *A. yabeana*           | 31.815        | 112.041        | Asia                | 13         | 18,296,460 | 3,276,927         | 1523x     | NENU_Aq1007      |

APPENDIX 1. *Aquilegia* sampling information.
### APPENDIX 2. Information about the *Aquilegia* sequence data previously published by Filiault et al. (2018) and downloaded from the National Center for Biotechnology Information Sequence Read Archive (SRA).

| Species                  | SRA no. | Size (Gbp) | Chloroplast reads | Depth         | Distribution region |
|--------------------------|---------|------------|-------------------|---------------|---------------------|
| *A. aurea*               | SRR405095 | 25.9      | 15,526,578        | 8520×         | Europe              |
| *A. vulgaris*            | SRR404349 | 27.5      | 48,865,464        | 26,870×       | Europe              |
| *A. sibirica*            | SRR405090 | 25.2      | 28,912,821        | 16,384×       | Asia                |
| *A. formosa*             | SRR408554 | 28.4      | 11,593,572        | 7209×         | North America       |
| *A. chrysantha*          | SRR408559 | 26.8      | 11,964,708        | 7209×         | North America       |
| *A. japonica*            | SRR413499 | 26.6      | 28,881,079        | 16,384×       | Asia                |
| *A. oxysepala var.*      | SRR413921 | 28.0      | 41,390,034        | 24,248×       | Asia                |

*aSequencing was performed on the Illumina platform.*

### APPENDIX 3. Chloroplast genome sequences downloaded from GenBank.

| Species                          | GenBank accession no. |
|----------------------------------|-----------------------|
| *Aconitum brachypodum*           | NC_041579.1           |
| *Actaea vaginata*                | MK253451.1            |
| *Adonis coerulea*                | MK253469.1            |
| *Anemolema glaucifolium*         | MH205609.1            |
| *Anemone raddeana*               | NC_041526.1           |
| *Anemonopsis macrophylla*        | NC_041527.1           |
| *Aquilegia coerulea*             | NC_041528.1           |
| *Aquilegia coerulea*             | MK69474.1             |
| *Aquilegia ecalcarata*           | NC_041529.1           |
| *Aquilegia ecalcarata*           | MK69475.1             |
| *Aquilegia rockii*               | NC_046738.1           |
| *Aquilegia rockii*               | MK573514.1            |
| *Asteropyrum cavalieri*           | NC_041530.1           |
| *Beesia calthifolia*             | NC_041531.1           |
| *Calathodes oxyacarpa*           | NC_041475.1           |
| *Callianthemum taipaicum*        | NC_041476.1           |
| *Callinthus palustris*           | MK253465.1            |
| *Ceratoccephala falcata*         | MK253464.1            |
| *Clematis terniflora*            | KJ956785.1            |
| *Consolida ajacis*               | NC_041534.1           |
| *Coepis chinensis*               | MK69483.1             |
| *Delphinium anthriscifolium*     | MK253461.1            |
| *Dichocarpum daizelii*           | MK253459.1            |
| *Enemion raddeanum*              | NC_041535.1           |
| *Eranthis stellata*              | NC_041536.1           |
| *Glaucomaum palmatum*            | MK69492.1             |
| *Gymnocodium gymnantrum*         | NC_033341.1           |
| *Halerpestes sarmentosa*         | MK253457.1            |
| *Helleborus thibetanus*          | NC_041540.1           |
| *Hydrastis canadensis*           | MK69495.1             |
| *Ipsypium manshuricum*           | NC_041541.1           |
| *Leptopyrum humanoideae*         | NC_041542.1           |
| *Megaleranianthus saniculifolia* | FJ597983.1            |
| *Naravelia pilulifera*           | NC_039542.1           |
| *Nigella damascena*              | NC_041537.1           |
| *Oxygraphis glacialis*           | NC_041538.1           |
| *Paraquilegia anemonoides*        | NC_041479.1           |
| *Pulsatilla chinensis*           | MK69491.1             |
| *Ranunculus macranthus*          | DG359689.1            |
| *Semiaquilegia adoxoides*        | MK69498.1             |
| *Staphisagria macrosporuma*      | MN648404.1            |
| *Thalictrum thalictroides*       | NC_039433.1           |
| *Trollius chinensis*             | NC_031849.1           |
| *Trollius ranunculoides*         | MK253447.1            |
| *Urophysa rockii*                | MK69502.1             |