Genome and transcriptome variation in Aquilegia viridiflora underlies the adaptive evolution to the environment in the early stage of speciation

Wei Zhang[1], Hua-Ying Wang[1], Tengjiao Zhang[1], Xiaoxue Fang[1], and Hongxing Xiao[1]

[1]Northeast Normal University

April 05, 2024

Abstract

The elucidation of adaptive evolution mechanisms in the early stage of speciation is important for understanding the origin and evolution of species. Aquilegia viridiflora is in the early stage of speciation, has a wide distribution range, and shows obvious phenotypic variation among its different populations. In this study, to analyze the adaptive evolution mechanism of phenotypic differences in the early stages of speciation, we analyzed the phenotypes, genomes, and transcriptomes of different populations of A. viridiflora. Our results indicated that A. viridiflora originated in northwestern China, and the uplift of the northeastern Qinghai–Tibet Plateau in the late Miocene may have caused its differentiation. Aqcoe5G459400 was a key gene in the early stage of A. viridiflora speciation. Its expression was reduced in environments with large temperature differences between day and night, causing A. viridiflora to develop larger flowers and enhancing its attractiveness to pollinators. Our research revealed the genetic basis of the adaptive evolution of the phenotype in the early stage of speciation and provides new evidence of the pattern of speciation via adaptative radiation in columbine.

INTRODUCTION

Understanding adaptive evolution is a long-term goal of evolutionary biology research (Schluter, 2000). Adaptation is a characteristic that confers some organisms with advantages over others in terms of survival and reproduction (Ronco et al., 2021). There is growing evidence that adaptation can enable species to continue to exist under changing conditions, including climate change (Hoffmann & Sgro, 2011). Temperature and precipitation are considered to be the driving factors that affect the population growth rate and limit the distribution of species (Cahill et al., 2014; Dalgleish, Koons, Hooten, Moffet, & Adler, 2011; Kim & Donohue, 2013). The differences between these environmental factors can cause populations to produce different traits to adapt to the current environment. For example, Csilléry et al.(2020) found that populations *of Abies alba* distributed in different regions present different morphologies and life histories to adapt to the temperature and precipitation regimes of different areas.

Natural selection is the only mechanism that leads to adaptive evolution, so many biologists simply define an adaptation as a feature evolved under natural selection (Futuyma, 2005). In fact, adaptation also emphasizes a process in which members of a population better adapt to certain factors in the environment via changes in characteristics that affect survival and reproduction, and this process is genetically regulated (Nagano et al., 2019). Screening highly differentiated regions in the genomes of populations that have adapted to different environments has become one of the methods for exploring local adaptation-related genes (Bian et al., 2020; Wong et al., 2020). For example, 1,394 highly differentiated SNPs were identified through the study of *Bufo andrewsi* distributed at different altitudes on the Qinghai-Tibet Plateau, indicating that these SNPs may be related to the adaptation of the species to the environments of high-altitude areas (Guo, Lu, Liao, & Merilä, 2016). However, the identification of highly differentiated regions in the genome may produce false positive

results due to the influence of the history of unique populations (Krak et al., 2016). Transcriptome sequencing to identify differentially expressed genes (DEGs) can also be used to identify candidate genes related to adaptation. Steve et al. found that DEGs among *Melanotaenia duboulay* i populations distributed in different regions were related to temperature tolerance. These DEGs played an important role in adapting to increased environmental temperature (S. Smith, Bernatchez, & Beheregaray, 2013). Therefore, the combined analysis of the genome and transcriptome can provide a more comprehensive understanding of the evolutionary history of species and reveal the adaptive evolution mechanisms of species. However, these methods do not combine the main driving force of natural selection, environmental factors, in the detection of genome regions where local adaptation occurs. Accordingly, the environmental association analysis (EAA) strategy was proposed, which can identify genetic mutations related to specific environmental factors and may reveal local adaptation patterns that are not found via the detection of highly differentiated regions (Bierne, Welch, Loire, Bonhomme, & David, 2011; Ritchie, Holzinger, Li, Pendergrass, & Kim, 2015). Bennett *et al.* found 128 SNPs related to adaptation to the environment in the genome of *Aedes aegypti* using the EAA methods to improve the accuracy of disease prediction models (Bennett, McMillan, & Loaiza, 2021).

Since plants are rooted in the ground, they must necessarily be subject to selection imposed by the environment. Adaptation to the current environment is very important for plants (Leimu & Fischer, 2008; Suarez-Gonzalez et al., 2016). Therefore, plants are an ideal model for studying adaptive evolution (Flood & Hancock, 2017). To explore the effects of environment-mediated natural selection on the adaptive differentiation of phenotypes, it is necessary to select a plant with a wide distribution range and different phenotypes. *Aquilegia viridiflora* is one such species that shows rich morphological variation. *A. viridiflora* is a perennial herb belonging to Ranunculaceae (Wu, Raven, & Hong, 2001). With the publication of new species, *Aquilegia* currently includes approximately 110 species worldwide and is widely distributed in the northern temperate zone (A. S. Erst et al., 2020; A. S. Erst et al., 2017a; Luo, Erst, Yang, Deng, & Li, 2018; Munz, 1946). Recently, Li *et al* . analyzed the whole genomes of *A. oxysepala* and *A. japonica* as well as putative hybrid thereof and revealed that a high level of genetic divergence was likely an outcome of combined effects of natural selection, genetic drift and divergent sorting of ancestral polymorphisms (M.-R. Li et al., 2019); Lu *et al.* analyzed the genomes and methylomes of 10 *Aquilegia* species distributed worldwide and speculated that their specific genetic and epigenetic structure endows *Aquilegia* species with a high degree of adaptability and promotes the evolution via adaptive radiation of the genus (Lu et al., 2019). Nevertheless, these studies focused on the degree of genomic differentiation among different species within*Aquilegia* . These species are mature species in the late stage of differentiation. In addition to the gene regions that began to differentiate in the early stage of speciation, a variety of factors have led to the obvious differentiation of the genome among species. The phenotypes of different populations of *A. viridiflora* show high diversity, and this species is still in the early stage of speciation. Thus, *A. viridiflora* provides ideal material for studying the key genes that lead to the adaptive evolution of phenotypes in the early stage of speciation and analyzing their molecular genetic mechanisms.

A. viridiflora is mainly distributed in northern China. Northern China has a vast territory and complex topography, extending between 75.14-132.58deg E and 33.24-53.28deg N, in which the east-west span is wide, and the environment changes gradually. The populations of *A. viridiflora* in different regions show obvious differences in phenotypes, which led Andrey S. Erst *et al.* to identify *A. viridiflora* with laminae purple or lilac-blue petals as the new species*A. kamelinii* (A. Erst, Shaulo, & Schmakov, 2013) and *A. viridiflora* with dark purple petals in North China as the new species*A. hebeica* (A. S. Erst et al., 2017b). However, the publication of these new species lacks support from molecular data, and according to field surveys, the phenotypic variation of *A. viridiflora* is continuous (see Figure 1). Therefore, *A.viridiflora* , *A. kamelinii* and *A. hebeica* were considered to be the same species (*A. viridiflora*) in our study. To reveal the genetic basis of the adaptive evolution of *A. viridiflora* , we intended to collect populations covering its main distribution range for genome resequencing and morphological characteristic measurement and analysis. Based on the observed phenotypic variation and the level of population differentiation, some populations were selected for transcriptome sequencing. Then, to explore the early stages of speciation, the adaptive evolution of phenotypes and the correlation between environmental factors, highly differentiated regions in the genome,

and differentially expressed genes were assessed to address the following questions: (1) Which phenotypes of *A. viridiflora* show local adaptation? (2) What are the genes associated with an adaptive phenotype? (3) How do environmental differences affect the genes associated with an adaptive phenotype? Answering these questions may reveal the genetic basis of the adaptive evolution of *A. viridiflora* phenotypes and its contribution to speciation. At the same time, this work can provide a more in-depth reference and theoretical basis for the model of adaptive radiation speciation in*Aquilegia* .

MATERIALS AND METHODS

Sampling

Seeds of *Aquilegia viridiflora* were collected from 20 locations covering its current distribution range and all voucher specimens were identified by Dr Hongxing Xiao and deposited in the Northeast Normal University Herbaria (Figure 1; Table S1). Each of the maternal plants was separated from the others by $>$ 50 m. Furthermore, to determine whether the *A. viridiflora* materials that we collected shared a most recent common ancestor (MRCA), we also collected *A. amurensis* , *A. ecalcarata* , *A. japonica* , *A. oxysepala* var. *kansuensis* , *A. oxysepala* var.*oxysepala* , *A. parviflora* and *A. yabeana* seeds in Asia, as reported by Zhang et al. (W. Zhang, Wang, Dong, Zhang, & Xiao, 2021). In addition, seeds of *Paraquilegia microphylla* were collected as an outgroup. All seeds were grown in the experimental field of Northeast Normal University. Fresh leaves and buds were used to extract DNA and RNA, respectively. These materials were quickly frozen in liquid nitrogen and stored at -80 .

Phenotypic Measurements and Statistics

To ensure the accuracy of the measurements, we planned to measure the following 12 morphological traits of different *A. viridiflora*populations in the experimental field during the full flowering stage. In brief, we randomly selected 10 plants from each population to measure the number of inflorescences and plant height; among these 10 plants, 6 flowers and 6 leaves were randomly selected from each plant, and corolla diameter, pistil length, stamen length, leaf area, leaf perimeter and chlorophyll content were measured and recorded; among these 6 flowers, we randomly selected 3 petals and calyxes to measure petal length (not including spurs), calyx length, spur length, the angle between petals and spurs (referred to simply as angle hereafter) and calyx length.

To reduce the error caused by different measurement batches due to the different florescences included, we used a mixed linear model to evaluate traits in the lme4 (Bates, Machler, Bolker, & Walker, 2014) package in R according to the following regression model equation:

$$Y_i = \mu + \beta_1 + \beta_2 P + \varepsilon_i$$

Where Y_i represents the traits of different populations, $\mu$is the actual measurement, β_1 and β_2 are regression coefficients, A represents the measurement batches, P is the person conducting the measurement, and ε_i is the residual variance. The evaluation results were used for ANOVA and K-means cluster analysis in R.

DNA Sequencing and SNP Calling

Total genomic DNA was extracted with a modified cetyltrimethylammonium bromide (CTAB) method (Doyle & Doyle, 1987). An Illumina Xten at Biomarker Tenchnologies, Inc. (Beijing, China) was used for genomic library generation and sequencing with 2 $\times$ 150 bp paired reads. Furthermore, raw reads of *Semiaquilegia adoxoide* s (SRR437677) were downloaded from the NCBI SRA database (http://www.ncbi.nlm.nih.gov/sra) to be employed as an outgroup. To obtain high-quality genomes, all reads were assessed by FastQC (Andrew, 2010) and filtered as follows: reads with adapters and reads with more than a 10% N content or more than 50% low-quality bases (quality value of less than 10) were removed. Low-quality reads were removed using NGStoolkit (Mulcare, 2004). The clean reads of each sample were mapped to the genome sequence of *A. coerulea* downloaded from the previous study of Filiault et al. (2018) using BWA v.0.7.12 with the default parameters (H. Li & Durbin, 2009). SAMtools v.0.1.18 was used for

sorting reads (H. Li et al., 2009). The HaplotypeCaller, GenotypeGVCFs and CombineGVCFs modules in GATK v.4.1.8.0 were used to produce accurate SNP calls (McKenna et al., 2010). To improve the quality of SNPs, VariantFiltration module in GATK v4.1.8.0 was used for filtration with the following parameters: –filter-name FilterQual –filter-expression "QUAL < 30.0" –filter-name FilterQD –filter-expression "QD < 2.0" –filter-name FilterMQ –filter-expression "MQ < 40.0" –filter-name FilterFS –filter-expression "FS > 60.0" -window 5 -cluster 2. Then, VCFtools v0.1.13 (Danecek et al., 2011) was employed to remove those variants that 1) showed a minor allele frequency (MAF) of 0.02 or less, 2) were not balletic variants, 3) showed a sequencing depth of less than 5, and 4) showed a missing rate exceeding 0.5. The filtered data were used for subsequent analysis. After filtering, each SNP was annotated using SnpEff v.5.0 (Cingolani et al., 2012).

Population Genetic Structure and Demographic History

We estimated the phylogenetic relationships of other *Aquilegia*species with *A. viridiflora* to determine whether the *A. viridiflora* materials used in our study shared an MRCA by using IQ-TREE multicore version 1.6.12 (Nguyen, Schmidt, Von Haeseler, & Minh, 2015) and MEGA X (Kumar, Stecher, Li, Knyaz, & Tamura, 2018) with 1000 bootstrap replicates. Both the ML tree and NJ tree indicated that 20 populations of *A. viridiflora* shared an MRCA, therefore, IQ-TREE was used to evaluate the phylogenetic relationships between different populations of *A. viridiflora* . To infer population genetic structure, principal component analysis (PCA) was performed by using EIGENSOFT v.6.1.4 (Price et al., 2006). ADMIXTURE v.1.3.0 (Alexander, Novembre, & Lange, 2009) was used to investigate the maximum likelihood of the ancestry of each individual or population. We set the K values from 1 to 10 with 10 replicates for each K value and examined the optimum K value according to the lowest value of the error rate. By combining the above results, the group modes of the 20 populations were determined (For the convenience of the following description, the grouping results of 20 populations are written here: NE, EL, CN and NW respectively).

The Genrealized Phylogenetic Coalescent Sampler (G-PhoCS) (Gronau, Hubisz, Gulko, Danko, & Siepel, 2011) was used to infer the demographic history of the *A. viridiflora* , including population divergence times, ancestral population size and migration rates based on neutral loci. The neutral loci were obtained according to the method of Wang et al. (2016), and the parameters were automatically set by setting the 'find-finetunes' attribute to 'TRUE'. Because G-PhoCS often shows limitations in recognizing complex migration models, we performed inferrences under different gene flows between each group based on the allele frequency by applying the f4-statistic using the qpDstat module in AdmixTools 7.0 (Patterson et al., 2012). Then, we set up a scenario involving gene flow by combining the results for the f4-statistic. Each Markov chain was run for 2, 000, 000 generations while sampling parameter values every 20th iteration. To obtain stable and reliable results, we independently ran the analysis five times. The burn-in and convergence of each run were determined by the boa (B. J. Smith, 2007) module in R packages. A neutral mutation rate of $10^{-8}$and a generation time of 1 year were used to calculate the effective population size, divergence times and migration rates (M.-R. Li et al., 2019).

Nucleotide Variation Pattern and Selected Analysis

Nucleotide diversity (π) and Tajima's D were calculated for the four groups with a 100 kb nonoverlapping sliding window using VCFtools (Danecek et al., 2011), respectively. In addition, the fixation index (Fst) between each of the four groups were also calculated by VCFtools (Danecek et al., 2011) using SNPs. PopLDdecay (C. Zhang, Dong, Xu, He, & Yang, 2019) software can be applied to compute linkage disequilibrium (LD) decay among different groups and chromosomes separately. To avoid the influence of LD on subsequent association and selection analyses, we employed PLINK v.1.9 (Purcell et al., 2007) to filter SNPs with following parameters: –indep-pairwise 50 10 0.2. The filtered data with 151,577 SNPs were phased using Beagle v.3.3.2 (Browning & Browning, 2007). To explore the effect of local adaptation on 12 quantitative traits, we used the single-phenotype Qst-Fst test with the R package 'QstFstComp' (Gilbert & Whitlock, 2015). If Qst > Fst, it means that the differentiation of traits is the major effect of divergent selection and shows local adaptation. We used the half-sib dam model and 10000 resampling steps for each QstFstComp analysis. With the aim of identifying candidate loci under natural selection from 151,577 SNPs, we used

BayeScan v.2.1 (Foll & Gaggiotti, 2008) software with the default parameters, and PGDSpider (Lischer & Excoffier, 2012) was used to produce an input file for BayeScan. SNPs with a q value lower than 0.05 were considered to be potentially selected loci.

Environmental Association Analysis

Nineteen current bioclimatic variables, with a spatial resolution of 30 s were collected from WorldClim (http://www.worldclim.org/). At the same time, we recorded the GPS information of the sampling locations, and downloaded the GPS information of *A. viridiflora*from CVH (http://www.cvh.ac.cn) and GBIF (http://www.gbif.org). ArcMap v.10.4 was used to limit the spatial extent according to the buffer radius (5 km) around each occurrence record. We used $|r| < 0.8$ (Pearson correlation coefficient) as a cutoff to remove highly correlated variables. The seven retained current bioclimatic factors (Bio1: annual mean temperature, Bio2: mean diurnal range, Bio3: isothermality, Bio4: temperature seasonality, Bio8: mean temperature of wettest quarter, Bio15: precipitation seasonality, Bio17: precipitation of driest quarter) were used for subsequent analysis. To identify the loci related to the seven retained current bioclimatic factors, BAYENV2 (Coop, Witonsky, Di Rienzo, & Pritchard, 2010) was used with 1,000,000 iterations and run three times separately. For each bioclimatic factor, the SNPs that were among the top 1% according to BF and among the top 5% according to the absolute Spearman's ϱ were considered as candidates.

RNA Sequencing and Differential Expression Analyses

The Qst-Fst test indicated that more floral characteristics showed local adaptation than other plant characteristics, so a bud-stage was selected for transcriptome sequencing. Considering the abovementioned population genetic analysis results and the clustering results for floral characteristics, ten populations were selected for transcriptome sequencing. Three samples were collected from each population as biological replicates to ensure the stability and reliability of the sequencing results (Table S1). Total RNA was extracted from samples using TRIzol (Invitrogen, USA) according to the manufacturer's protocol, A NanoDrop 2000 spectrophotometer (Thermo Scientific, USA) and gel electrophoresis were used to estimate the quality of the RNA. An Illumina Hiseq system at Novogene Technologies, Inc. (Beijing, China) was used for genomic library generation and sequencing with 2×150 bp paired reads. Raw reads were assessed by using FastQC (Andrew, 2010) and filtered by using Trimmomatic v.0.39 (Bolger, Lohse, & Usadel, 2014). The trimmed reads of each sample were mapped to the same reference genome employed above using STAR v.2.7.5c (Dobin et al., 2013), and the expression levels were estimated using RSEM v.1.3.0 (B. Li & Dewey, 2011). Based on the obtained read counts, the R package DESeq2 was used to calculate differential gene expression via pairwise comparisons (taking any two groups among NE, EL, CN and NW) (Love, Anders, & Huber, 2014). Transcripts with an absolute log2FoldChange (LFC) value greater than 1 were considered significantly DEGs.

Weighted Gene Correlation Network Analysis (WGCNA)

According to the analysis of the DEGs and genetic distance, we found the greatest genetic distance and more DEGs between EL and NW, so all genes expressed by EL and NW were selected for WGCNA. The analysis was conducted using the WGCNA R package following the provided tutorials (Langfelder & Horvath, 2008). The soft thresholding power was determined according to the principle of a nonscale network, and the lowest power when the correlation coefficient reached the plateau period was used as a parameter in subsequent analysis. A gene clustering tree was constructed according to the correlations between the expression levels of genes, and the co-expression modules were identified by dynamic tree cut with a minimum module size of 30. When the correlation between modules and traits was greater than 0.75 ($p < 0.05$), it indicated that the modules were significantly related to the traits.

Identification of the Key Genes Related to Adaptation

To explore the adaptive evolution mechanism of *A. viridiflora* , SNPs that were related to the environment and selection were identified. Candidate SNPs were mapped to the corresponding genes to check whether these genes were included in modules that were significantly related to the phenotype and differentially expressed. The genes obtained through the above filtration procedure were considered to be the key genes in

the adaptive evolution of *A. viridiflora* . The coding sequence (CDS) regions of candidate genes were analyzed to identify whether the mutations in the CDS regions were synonymous or nonsynonymous. In addition, Multiple Em for Motif Elicitation (MEME) v.5.3.3 (https://meme-suite.org/meme/tools/meme) (Bailey & Elkan, 1994) was used to find transcription factor binding sites in the 2 kb region upstream and downstream of the gene, and Motif Comparison Tool (Tomtom) v.5.3.3 (https://meme-suite.org/meme/tools/tomtom) (Gupta, Stamatoyannopoulos, Bailey, & Noble, 2007) was used to annotate the motif functions.

RESULTS

Phylogeny and Population Structure

After filtering, we obtained 1,064,089 high-quality SNPs from 80 individuals, 55.604% of which were located upstream and downstream, while 17.454% were located in intronic regions, and 7.07% were located in exonic regions of genes. According to the above SNPs, an ML tree and NJ tree were constructed to infer the phylogenetic relationships among*Aquilegia* species. Both topologies indicated that different populations of *A. viridiflora* shared a recently common ancestor with high support (Figure S1). Therefore, population genetics analysis could subsequently be performed on the 20 populations of *A. viridiflora* . The individuals of *A. viridiflora* were divided into eastern and western lineages, each of which contained two groups. For convenience, the four groups were designated as NE (including the WD, AE, SZ and LF populations), EL (including the LT, YS, MS and TS populations), CN (including the YT, QB, TL and YM populations), NW (including the XW, HL, HH, HD, YS, SF, ZG and JQ populations) (Figure 2).

When performing the subsequent population genetic analysis, 672,439 high-quality SNPs were used. We performed PCA to determine the grouping of populations, and principal component 1 (PC1) and principal 2 (PC2) explained 6.61% and 4.04% of the observed variation, respectively (Figure S2A). The groups obtained according to the PCA plot were the same as those indicated in the phylogenetic tree. ADMIXTURE analysis showed that the optimal K value was 4, as $K = 4$ resulted in the lowest cross-validation error; at this time, populations SZ, LT, YM, XW, HL, HH and HD showed gene introgression from other populations (Figure S2B).

Group Dynamic History Inference

We used the ABBA-BABA test to detect gene flow between the four groups. Tests involving NE and EL, NE and CN, EL and CN, and CN and NW showed a significant deviation of D-stat from zero (absolute value of Z-score greater than 2), indicating that there was gene flow between these groups (Figure 3A and Table S2). G-PhoCS analysis showed that the gene flow between NE and EL and between NE and CN was greatest and that there was less gene flow from NE to the other two groups than in the opposite direction. In contrast, there was less gene flow between NW and CN and between NW and EL, and there was less gene flow from NW to the other groups than in the opposite direction. For the ease of reference, the lineage composed of NE and EL was called NL, and the lineage composed of CN and NW was called CW. The effective population size (Ne) was 24,016,900 for the ancestors. The ancestry population was indicated to have differentiated into NL and CW in the middle Miocene, approximately 6.52 Mya, and their effective population size were 9,097,225 and 9,932,925, respectively. Then, CN and NW separated in the middle Miocene, approximately 6.04 Mya, and NE and EL separated in the early Pliocene approximately 5.82 Mya. The current effective population sizes of NE, NW, EL and CN were indicated to be 6,462,550, 1,617,175, 1,510,075 and 2,698,350, respectively (Figure 3B and Table S3). All the Ne values of the other groups were lower than that of their common ancestor. Linkage disequilibrium analysis showed that CN and EL presented a greater degree of LD, while NE and NW showed less linkage disequilibrium (indicated by r^2). When $r^2 = 0.1$, the decay distances of NE, EL, CN and NW were 10 kb, 37 kb, 44 kb and 6.9 kb, respectively (Figure S2C). The rapid decay of NE and NW may have been due to the higher genetic diversity of their genomes relative those of EL and CN.

Local Phenotypic Adaptation

ANOVA indicated that each phenotype differed significantly among different groups (Table 1). The K-means

cluster analysis of phenotypes showed that Dim1 and Dim2 could explain 61.3% and 25.6% of the observed variation, respectively, and the cumulative contribution to the observed phenotypic variation was close to 90%. All individuals of CN clustered into one clade; the individuals of the MS and TS populations in the EL group clustered into one clade, which was clearly separated from the other two populations in EL; the phenotypes of the individuals in NW and NE showed only small differences and could not be clearly distinguished; and the YS and LT populations in EL were grouped with the individuals of NW and NE (Figure S3). Among the four groups, CN showed the smallest corolla diameter, the shortest petal length, spur length, pistil length, stamen length and calyx length and the lowest chlorophyll content, followed by EL, while NW and NE showed the largest values of these phenotypes (the difference between the last two groups was small); plant height and the number of inflorescences showed the reverse order of the above traits. EL showed the largest leaf area and perimeter, followed by CN, while the leaf area of NW and NE presented little difference, but the leaf perimeter of NW was larger than that of NE. The angle of CN was the largest, followed by those of NE and NW (the difference between of them was not much), and the angle of EL was the smallest (Figure 4).

Most of the floral characteristics of *A. viridiflora* showed a significant negative correlation with nutritional traits; the other floral characteristics except for stamen length and angle were negatively correlated with the number of inflorescences, while the number of inflorescences was significantly positively correlated with height (Figure S4). The mean Fst for the four groups was 0.1519 (0.14949 – 0.15074), and the overall Qst was higher than the mean Fst for seven traits, including corolla diameter, petal length, angle, spur length, pistil length, inflorescence number and leaf area, indicating that these traits showed local adaptations (Table 1).

Identify of Selected Sites Driven by Environment

The population genetic parameters, including nucleotide diversity (π) and Tajima's D, of the NE, EL, CN and NW populations were calculated throughout the genome. Among the four groups, NW showed the highest nucleotide diversity, and EL showed the lowest nucleotide polymorphism; the Tajima's D values of the four groups were all > 0 (Figure S5A and B). F_{ST} was calculated for the four groups to infer population genetic differentiation. At the overall level of the genome, the F_{ST} values between NE and EL, NE and CN, NE and NW, EL and CN, EL and NW, and CN and NW were 0.160, 0.151, 0.152, 0.159, 0.193 and 0.161, respectively. These results indicated that EL and NW were highly differentiated, while NE and CN were poorly differentiated. The F_{ST} values calculated for the four groups using 10 kb windows across genomes were consistent with the F_{ST} at the overall level of the genome (Figure S5C). Based on the Bayesian method applied in BAYESCAN, we identified 20,796 outlier SNPs from the 151,577 filtered SNP according to a 0.05 threshold for the q-value among the four groups. These outlier SNPs might putatively be considered to have been selected among the four groups. In addition, 1730 outlier SNPs were identified by BAYENV2, which were associated with seven environmental variables. There were 1,027 and 862 SNPs associated with Bio2 and Bio8, respectively. Considering the intersection of the SNPs identified by BAYESCAN and BAYENV2, 347 SNPs located in the intersection might be presumed to be under the environmental selection, which were located in 157 genes (Table S4).

Identification of DEGs and Gene Modules Analysis

To identify the adaptative genes, a population transcriptome analysis was performed. A total of 1,323,834,776 raw reads (456 Gb) were obtained from 30 samples from ten populations. When the genomic variation and phenotypic variation results were combined, the ten populations were again divided into four groups (NE, EL, CN and NW) for differential expression analysis. The greatest total numbers of DEGs were found between NE vs. EL, EL vs. NW and CN vs. NW (928, 1716 and 855 DEGs, respectively). There were fewer DEGs between EL vs. CN, NE vs. CN and NE vs. NW (428, 666 and 513 DEGs, respectively). Gene Ontology (GO) enrichment analyses were conducted for the DEGs identified between different groups. The DEGs identified from different groups were related to biological process (BP) categories including oxidation-reduction process, carbohydrate metabolic process, response to biotic stimulus, defense response, response to stress and fatty acid biosynthetic process; molecular function (MF) categories including hydrolase activ-

ity, hydrolyzing O-glycosyl compounds, oxidoreductase activity, fatty-acyl-CoA reductase (alcohol-forming) activity, endopeptidase inhibitor activity, iron ion binding, oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen, heme binding, chitin binding, serine-type carboxypeptidase activity, serine-type endopeptidase inhibitor activity, cysteine-type peptidase activity, cysteine-type peptidase activity and hydrolase activity, acting on ester bonds; and cellular component (CC) categories including cell wall and membrane (Table S5 – S10).

Because the greatest number of DEGs was found between EL and NW, we used 23,418 genes expressed in both EL and NW for WGCNA to identify gene modules related to adaptive floral characteristics (including corolla diameter, petal length, angle, spur length, pistil length and inflorescence number). The analysis identified 33 gene modules, the number of genes in each module ranged from 58 (darkolivegreen) to 3229 (turquoise), and 24 genes were not classified into any modules. Among the six adaptive floral characteristics, only spur length presented a significant correlation with four modules, namely Magenta (640 genes), Darkgray (257 genes), Saddlebrown (134 genes) and Orange (201 genes). No gene modules were associated with other traits (Figure 5). GO enrichment analysis showed that the genes in the Magenta module were related to the kinesin complex (24 genes), microtubule motor activity (24 genes), microtubule-based movement (24 genes), microtubule binding (24 genes), DNA binding (53 genes), DNA replication (16 genes), nucleus (35 genes), nucleosome (8 genes), ATP binding (68 genes), MCM complex (3 genes), mitotic chromosome condensation (4 genes), DNA replication initiation (3 genes) and microtubule (5 genes). The genes in the Darkgrey module were related to the following categories: structural constituent of ribosome (83 genes), ribosome (82 genes), translation (83 genes), intracellular (62 genes), 5S rRNA binding (4 genes) and large ribosomal subunit (4 genes). The genes in the Saddlebrown module were related to transporter activity (8 genes) and membrane (15 genes). The genes in the Orange module were not significantly enriched in any GO terms.

Key Genes Regulating the Adaptive Evolution of Floral Size

Among the 157 genes affected by environmental selection, 21 genes were DEGs. Most of these genes were located on chromosomes 3, 5 and 7, among which the greatest number of candidate genes was distributed on chromosome 7 (Figure 6A and Figure S6). The results of QTL mapping also revealed the existence of genes related to controlling organ morphology on chromosome 7 of *Aquilegia* (Edwards et al., 2021). Six genes whose variations were affected by environmental selection showed nonsynonymous mutation: *Aqcoe1G340100* (Ankyrin repeat family protein), *Aqcoe1G488100* (ascorbate peroxidase 1),*Aqcoe3G242600* (glutamate dehydrogenase 2), *Aqcoe5G182600*(cytokinin oxidase 5), *Aqcoe5G459400* (OBP3-responsive gene 1), and *Aqcoe7G135600* (pleiotropic drug resistance 11) (Figure 6A). Except for *Aqcoe1G488100* , which was related to Bio8, the rest of the genes were related to Bio2. These six genes were essential for plant growth and development. In addition, *Aqcoe5G459400* and*Aqcoe7G135600* were located in the Orange module and Saddlebrown module related to the spur length, respectively. The expression of*Aqcoe5G459400* was significantly negatively correlated with the spur length ($p = 0.007$) (Figure 6B), while that of*Aqcoe7G35600* showed no significant correlation with spur length ($p = 0.3481$). Therefore, *Aqcoe5G459400* may be a candidate gene for *A. viridiflora* spur length to adapt to environmental evolution. WGCNA showed that the genes that interacted with*Aqcoe5G459400* were *Aqcoe2G056600* , *Aqcoe2G172400* ,*Aqcoe3G439700* , *Aqcoe7G244300* and *Aqcoe7G395200*(Figure 6C). In addition, there was a binding site for the ABI3/VP1 transcription factor located 1.8 kb upstream of the *Aqcoe5G459400*gene.

DISSCUSSION

The Phylogeny of *Aquilegia viridiflora*

Angiosperms originated in the late Permian (*c.* 250 million years ago), through adaptive radiation evolution leading to rapid diversification (Shi, Herrera, Herendeen, Clark, & Crane, 2021). Currently, there are approximately 400,000 species in this group (Soltis, Folk, & Soltis, 2019). When adaptive radiation evolution occurs, a species exabits a wide range of morphological diversity but relatively few genetic differences, which causes difficulties in species delimitation (Pinheiro, Dantas-Queiroz, & Palma-Silva, 2018). The lower eudicot genus *Aquilegia* , which represents a phylogenetic midpoint between model species such as *Arabidopsis* and

Oryza and has recently undergone adaptive radiation evolution, has become an ideal model for studying adaptive evolution (Kramer, 2009). Due to its wide distribution and obvious morphological differences, *A. viridiflora* has attracted extensive attention from plant taxonomists. According to the records of the Flora of China (Wu et al., 2001), the 66 individuals in the 20 populations shown in Figure 1 may be regarded as*A. viridiflora* . However, according to Andrey S. Erst's descriptions of the morphological characteristics of new species, the WD population should be referred to as *A. kamelinii* (A. Erst et al., 2013); and the TS, MS, YM, TL, QB and YT populations should be referred to as *A. hebeica* (A. S. Erst et al., 2017b). In species that undergo adaptive radiation speciation, it is particularly difficult to distinguish between new species and populations. However, the correct division of species is the first step in understanding speciation, while also ensuring accurate sampling. In our study, phylogenetic analysis based on the genome showed that *A. kamelinii* , *A.hebeic* a and *A. viriflora* constituted a fully supported monophyletic group and that *A. kemelinii* showed the closest relationship with*A. viridiflora* in Northeast China; *A. hebeica* showed two different origins, one originated on Chinese Peninsula (including the Liaodong Peninsula and Shandong Peninsula) and the other originated in Northwestern China. The two different origins corresponded to two separate lineages (NL and CW). Species formed by adaptive radiation are usually composed of recently differentiated lineages whose speciation is incomplete, so molecular data alone are not sufficient to define species (Rundell & Price, 2009). Therefore, according to the cluster analysis of morphological traits, *A. kamelinii* cannot be distinguished from *A. viridiflora* ; while *A. hebeica* was significantly different from *A. viridiflora* ; these species were included in the TS, MS populations and the YT, QB, YM, TL populations were also significantly different. Additionally, the ongoing gene flow between the four groups also showed that *A. viridiflora* was not fully differentiated. Hence, the combination of phenotypic and molecular data indicated that *A. viridiflora* is a "species on the speciation way" that has not formed a mature species but could be regarded as a single polymorphic species (Liu, 2016). As a result, it is recommended that *A. kamelinii* still be reffered to as *A. viridiflora*f. *atropurpurea* . Additionally, *A. hebeica* is not a monophyletic group, and its phylogenetic position is worthy of further study.

Demographic History and Gene Flow among *A. viridiflora*

A. viridiflora is widely distributed as an ornamental flower in northern China. Both the ML tree and PCA showed four groups within*A. viridiflora* . Additionally, the optimum K value according to the ADMIXTURE result was 4, which verified the existence of four lineages. The ADMIXTURE results showed that all groups had admixed genetic backgrounds, indicating that there may be ongoing gene flow among them. D-statistics show that in the four lineages of *A. viridiflora* , there was gene flow between NE and EL, NE and CN, EL and CN, and CN and NW. The gene flow from warm and humid regions (EL and CN) to the cold and arid regions (NE and NW) was greater than the gene flow in the opposite direction, which was beneficial for increasing the genetic diversity of *A. viridiflora* in the cold and arid environments and enhancing its local adaptability. This result is in accord with studies on different populations of *Silene ciliata* ,*Pseudotsuga menziesii* and *Eperua falcata* (Brousseau, Fine, Dreyer, Vendramin, & Scotti, 2021; George et al., 2021; Morente-López et al., 2021). The G-phoCS results showed that *A. viridiflora*differentiated into two lineages, NL and CW, 6.52 Mya. Subsequently, CW differentiated into two groups, CN and NW, 6.04 Mya, and NL differentiated into NE and EL groups 5.82 Mya. The divergence times of these groups were all in the late Miocene. At this time, a global cooling event (late Miocene coding, LMC) caused by the uplift of the northeastern part of the Qinghai-Tibet Plateau, and a general climate pattern of monsoons in eastern China and arid inland areas in northwestern China formed (Chen et al., 2019; Steinthorsdottir et al., 2021). Therefore, temperature may be an important driving force for the differentiation of *A. viridiflora* . Environmental changes in the late Miocene also drove the lineage differentiation of the*Psammobates tentorius* complex and angraecoids in Aferica (Farminhão et al., 2021; Zhao, Heideman, Bester, Jordaan, & Hofmeyr, 2020). Furthermore, 5 of the candidate genes were related to resistance to environmental stress, including *Aqcoe1G340100* (Seong et al., 2007), *Aqcoe1G488100* (Koussevitzky et al., 2008),*Aqcoe3G242600* (Magadlela et al., 2019), *Aqcoe5G182600* (S. Li et al., 2019) and *Aqcoe7G135600* (He et al., 2019), indicating that the differences in the growth environment drive the genetic differentiation of different groups of *A. viridiflora* . Although the divergence times between the four groups of *A. viridiflora*were long, the degree of genome differentiation was low (Fst

are all less than 0.25), which may be caused by the gene flow between the four groups. Continuous gene flow inhibited population differentiation and maintained the boundaries of species.

In summary, the group in northwestern China presented a high level of nucleotide polymorphism and the low level of linkage disequilibrium, indicating that this group may be the origin *A. viridiflora* . These groups differentiated in the late Miocene, when one branch migrated to the North China region, and the other branch migrated to the northeastern region. Then, the group in the northeastern region diverged and migrated southward, forming a group in the East Shandong South Liaoning area.

Key Genes Regulating the Adaptive Evolution of *A. viridiflora*

Plant growth is greatly affected by the environment, so adaptive adjustments to different environments are necessary (Anderson & Song, 2020). In this study, the four identified groups of *A. viridiflora* were located in different climate types. The areas where NE and NW were located showed large mean diurnal temperature ranges, while the areas where EL and CN were located showed small mean diurnal temperature range. Therefore, different groups of *A. viridiflora*showed different adaptive strategies in the face of different climates. While NE and EL shared an MRCA, NW and CN shared an MRCA, the differences in the number of DEGs and phenotypes between EL and CN and between NE and NW were smaller than the differences between NE and EL and between NE and NW, indicating that the EL and CN groups and the NE and NW groups have undergone convergent evolution. The molecular mechanism of phenotypic convergence is worthy of further study. Fst-Qst analysis showed that corolla diameter, petal length, angle, spur length, pistil length, inflorescence number and leaf area reflected adaptability to the environment. Among the adaptable traits, corolla diameter, petal length, and spur length, which represent flower size, were significantly negatively correlated with the inflorescence number and leaf area. NE and NW had larger and fewer flowers and smaller leaves than the other populations, while CN and EL had smaller flowers, a greater number of flowers and larger leaves. The wide mean diurnal temperature range limits the activities of pollinators. To ensure the production of offspring, *A. viridiflora* use more energy for reproductive growth and less energy for vegetative growth. Moreover, the populations in this area used the limited energy available for reproductive growth to form a small number of flowers, thus enhancing their ability to attract pollinators. In contrast, there are abundant pollinators in areas with a narrow mean diurnal temperature range, and *A. viridiflora* populations in such areas use more energy for vegetative growth to ensure their long-term survival. The lineages in such areas use their reproductive growth energy to form more flowers and produce more seeds. This phenomenon has also been reported in *Pedicularis siphonantha*populations (Dai, Amboka, Kadiori, Wang, & Yang, 2017). In addition, populations located in harsh environments exhibit longer pistils than those located in good environments, which may be to increase the chance of contact with pollinators.

Aqcoe5G182600 (Cytokinin oxidase 5) and *Aqcoe5G459400*(OBP3-responsive gene 1) might have played an important role in the adaptive evolution of *A. viridiflora* . Cytokinin oxidase 5 and its orthologous genes play an important role in plant grain yield.*Os CKX2*knockout show increases in the numbers of both panicle branches and grains per plant under stress conditions in *Oryza sativa*(Ashikari et al., 2005). Accordingly, *Aqcoe5G182600* may regulate the activity of flower primordia and control the number of inflorescences by controlling the levels of cytokinin. The appearance of nectar spur is a critical morphological characteristic of highly diversified *Aquilegia* . *Aqcoe5G459400* was shown to be related to the nectar spur by integrating the results of WGCNA, and its expression level was negatively correlated with the length of the spur.*Aqcoe2G056600* , *Aqcoe2G172400* , *Aqcoe3G439700* ,*Aqcoe7G244300* and *Aqcoe7G395200* interacting with*Aqcoe5G459400* in the module are related to histone methylation and ion transport across membranes and may therefore improve the ability of plants to resist insect pests and adverse stressed (Fedoreyeva, Vanyushin, & Baranova, 2020; Liao et al., 2018; Philippe, Ralph, Külheim, Jancsik, & Bohlmann, 2010; Rodríguez-Celma, Chou, Kobayashi, Long, & Balk, 2019; Surya, 2020). Previous studies have shown that OBP3-responsive gene 3 (*ORG3*) controls the size of petals by controlling the number of cells (Omidbakhshfard et al., 2018), and*A. ecalcarata* ceases cell division and begins cell differentiation earlier (Ballerini, Kramer, & Hodges, 2019). Therefore,*Aqcoe5G459400* may control the size of spur by controlling the number of cells in *A. viridiflora* . In addition, ABI3/VP1 and its orthologs are key genes involved in regulating

abscisic acid (ABA), which not only plays a role in the dormancy of seeds and buds, but also affects the flowering time of plants (Riboni, Robustelli Test, Galbiati, Tonelli, & Conti, 2016; Shu et al., 2018; Shu et al., 2016; Shu et al., 2013). In summary, there may be a cascade reaction between ABI3/VP1 and*ORG3* , which in turn controlls the size of the spur in *A. viridiflora* . Generally, *Aqcoe5G459400* could be considered as a key gene in the early stage of *A. viridiflora* speciation. Its expression level is lower in the lineages from regions with a wide mean diurnal temperature range, which causes *A. viridiflora* to produce larger flowers to attract more pollinators and increases the chance of successful reproduction. In contrast, its expression level is higher in the lineages from regions with a narrow mean diurnal temperature range, and *A. viridiflora* produces smaller flowers in such areas to increase the number of seeds produced. Thus, *A. viridiflora* has developed different adaptive evolution mechanisms in response to different environments.

AUTHOR CONTRIBUTIONS

X. HX. and W. HY designed the study and evaluated the results; W. HY and Z. W. collected the materials; Z. W., Z. TJ. and F. XX. participated in data analysis; Z. W. and W. HY. prepared the manuscript; all authors read and approved the final manuscript.

ACKNOWLEDGMENTS

The research was supported by the National Natural Science Foundation of China (32070244) and "the Fundamental Research Funds for the Central Universities". We acknowledge Mingzhou Sun for his help in materials collection.

DATA AVAILABILITY STATEMENT

Raw sequence data is available from the National Center for Biotechnology Information's (NCBI) Sequence Read Archive (SRA) under the submission XXX. The phenotypic and climate data have been archived in dryad (https://doi.org/10.5061/dryad.sqv9s4n4m).

ORCID

Hongxing Xiao https://orcid.org/0000-0002-6040-5443

REFERENCES

Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome research, 19* (9), 1655-1664.

Anderson, J. T., & Song, B. H. (2020). Plant adaptation to climate change—Where are we? *Journal of Systematics and Evolution, 58* (5), 533-545.

Andrew, S. (2010). A quality control tool for high throughput sequence data. *Fast QC, 532* .

Ashikari, M., Sakakibara, H., Lin, S., Yamamoto, T., Takashi, T., Nishimura, A., . . . Matsuoka, M. (2005). Cytokinin oxidase regulates rice grain production. *science, 309* (5735), 741-745.

Bailey, T. L., & Elkan, C. (1994). Fitting a mixture model by expectation maximization to discover motifs in bipolymers.

Ballerini, E. S., Kramer, E. M., & Hodges, S. A. (2019). Comparative transcriptomics of early petal development across four diverse species of *Aquilegia* reveal few genes consistently associated with nectar spur development. *BMC Genomics, 20* (1), 668.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. *arXiv preprint arXiv:1406.5823* .

Bennett, K. L., McMillan, W. O., & Loaiza, J. R. (2021). The genomic signal of local environmental adaptation in *Aedes aegypti*mosquitoes. *Evolutionary Applications* .

Bian, J., Cui, L., Wang, X., Yang, G., Huo, F., Ling, H., . . . Levi, B. (2020). Genomic and Phenotypic Divergence in Wild Barley Driven by Microgeographic Adaptation. *Advanced Science, 7* (24), 2000709.

Bierne, N., Welch, J., Loire, E., Bonhomme, F., & David, P. (2011). The coupling hypothesis: why genome scans may fail to map local adaptation genes. *Molecular ecology, 20* (10), 2044-2072.

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics, 30* (15), 2114-2120.

Brousseau, L., Fine, P. V., Dreyer, E., Vendramin, G. G., & Scotti, I. (2021). Genomic and phenotypic divergence unveil microgeographic adaptation in the Amazonian hyperdominant tree *Eperua falcata*Aubl.(Fabaceae). *Molecular ecology, 30* (5), 1136-1154.

Browning, S. R., & Browning, B. L. (2007). Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *The American journal of human genetics, 81* (5), 1084-1097.

Cahill, A. E., Aiello-Lammens, M. E., Caitlin Fisher-Reid, M., Hua, X., Karanewsky, C. J., Ryu, H. Y., . . . Wiens, J. J. (2014). Causes of warm-edge range limits: systematic review, proximate factors and implications for climate change. *Journal of Biogeography, 41* (3), 429-442.

Chen, C., Bai, Y., Fang, X., Guo, H., Meng, Q., Zhang, W., . . . Murodov, A. (2019). A Late Miocene Terrestrial Temperature History for the Northeastern Tibetan Plateau's Period of Tectonic Expansion. *Geophysical Research Letters, 46* (14), 8375-8386.

Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., . . . Ruden, D. M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly, 6* (2), 80-92.

Coop, G., Witonsky, D., Di Rienzo, A., & Pritchard, J. K. (2010). Using environmental correlations to identify loci underlying local adaptation. *Genetics, 185* (4), 1411-1423.

Dai, W.-K., Amboka, G. M., Kadiori, E. L., Wang, Q.-F., & Yang, C.-F. (2017). Phenotypic plasticity of floral traits and pollination adaption in an alpine plant *Pedicularis siphonantha* D. Don when transplanted from higher to lower elevation in Eastern Himalaya. *Journal of Mountain Science, 14* (10), 1995-2002.

Dalgleish, H. J., Koons, D. N., Hooten, M. B., Moffet, C. A., & Adler, P. B. (2011). Climate influences the demography of three dominant sagebrush steppe plants. *Ecology, 92* (1), 75-85.

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., . . . Sherry, S. T. (2011). The variant call format and VCFtools. *Bioinformatics, 27* (15), 2156-2158.

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., . . . Gingeras, T. R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics, 29* (1), 15-21.

Doyle, J. J., & Doyle, J. L. (1987). *A rapid DNA isolation procedure for small quantities of fresh leaf tissue* . Retrieved from

Edwards, M. B., Choi, G. P., Derieg, N. J., Min, Y., Diana, A. C., Hodges, S. A., . . . Ballerini, E. S. (2021). Genetic architecture of floral traits in bee-and hummingbird-pollinated sister species of*Aquilegia* (columbine). *bioRxiv* .

Erst, A., Shaulo, D., & Schmakov, A. (2013). *Aquilegia kamelinii*(Ranunculaceae)–a new species from North Asia. *Turczaninowia, 16* (3), 8-10.

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

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

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

Farminhao, J. N., Verlynde, S., Kaymak, E., Droissart, V., Simo-Droissart, M., Collobert, G., . . . Stevart, T. (2021). Rapid radiation of angraecoids (Orchidaceae, Angraecinae) in tropical Africa characterised by multiple karyotypic shifts under major environmental instability. *Molecular Phylogenetics and Evolution, 159* , 107105.

Fedoreyeva, L. I., Vanyushin, B. F., & Baranova, E. N. (2020). Peptide AEDL alters chromatin conformation via histone binding. *AIMS Biophysics, 7* (1), 1-16.

Flood, P. J., & Hancock, A. M. (2017). The genomic basis of adaptation in plants. *Current opinion in plant biology, 36* , 88-94.

Foll, M., & Gaggiotti, O. (2008). A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics, 180* (2), 977-993.

Futuyma, D. J. (2005). *Evolution* : Evolution.

George, J. P., Schueler, S., Grabner, M., Karanitsch-Ackerl, S., Mayer, K., Stierschneider, M., . . . van Loo, M. (2021). Looking for the needle in a downsized haystack: Whole-exome sequencing unravels genomic signals of climatic adaptation in Douglas-fir (*Pseudotsuga menziesii*).*Ecology and Evolution* .

Gilbert, K. J., & Whitlock, M. C. (2015). QST–FST comparisons with unbalanced half-sib designs. *Molecular ecology resources, 15* (2), 262-267.

Gronau, I., Hubisz, M. J., Gulko, B., Danko, C. G., & Siepel, A. (2011). Bayesian inference of ancient human demography from individual genome sequences. *Nature genetics, 43* (10), 1031.

Guo, B., Lu, D., Liao, W. B., & Merila, J. (2016). Genomewide scan for adaptive differentiation along altitudinal gradient in the Andrew's toad*Bufo andrewsi* . *Molecular ecology, 25* (16), 3884-3900.

Gupta, S., Stamatoyannopoulos, J. A., Bailey, T. L., & Noble, W. S. (2007). Quantifying similarity between motifs. *Genome Biology, 8* (2), 1-9.

He, Y., Xu, J., Wang, X., He, X., Wang, Y., Zhou, J., . . . Meng, X. (2019). The *Arabidopsis* pleiotropic drug resistance transporters PEN3 and PDR12 mediate camalexin secretion for resistance to Botrytis cinerea. *The Plant Cell, 31* (9), 2206-2222.

Hoffmann, A. A., & Sgro, C. M. (2011). Climate change and evolutionary adaptation. *Nature, 470* (7335), 479-485.

Kim, E., & Donohue, K. (2013). Local adaptation and plasticity of*Erysimum capitatum* to altitude: its implications for responses to climate change. *Journal of Ecology, 101* (3), 796-805.

Koussevitzky, S., Suzuki, N., Huntington, S., Armijo, L., Sha, W., Cortes, D., . . . Mittler, R. (2008). Ascorbate peroxidase 1 plays a key role in the response of *Arabidopsis thaliana* to stress combination. *Journal of Biological Chemistry, 283* (49), 34197-34203.

Krak, K., Vit, P., Belyayev, A., Douda, J., Hreusova, L., & Mandak, B. (2016). Allopolyploid origin of *Chenopodium album* s. str.(Chenopodiaceae): a molecular and cytogenetic insight. *PLoS One, 11* (8), e0161063.

Kramer, E. M. (2009). *Aquilegia* : a new model for plant development, ecology, and evolution. *Annual review of plant biology, 60* , 261-277.

Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA X: molecular evolutionary genetics analysis across computing platforms.*Molecular biology and evolution, 35* (6), 1547-1549.

Langfelder, P., & Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. *BMC bioinformatics, 9* (1), 1-13.

Leimu, R., & Fischer, M. (2008). A meta-analysis of local adaptation in plants. *PLoS One, 3* (12), e4010.

Li, B., & Dewey, C. N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC bioinformatics, 12* (1), 323.

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics, 25* (14), 1754-1760.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., . . . Durbin, R. (2009). The sequence alignment/map format and SAMtools.*Bioinformatics, 25* (16), 2078-2079.

Li, M.-R., Wang, H.-Y., Ding, N., Lu, T., Huang, Y.-C., Xiao, H.-X., . . . Li, L.-F. (2019). Rapid divergence followed by adaptation to contrasting ecological niches of two closely related columbine species*Aquilegia japonica* and *A. oxysepala* . *Genome biology and evolution, 11* (3), 919-930.

Li, S., An, Y., Hailati, S., Zhang, J., Cao, Y., Liu, Y., . . . Yang, P. (2019). Overexpression of the cytokinin oxidase/dehydrogenase (CKX) from*Medicago sativa* enhanced salt stress tolerance of*Arabidopsis* . *Journal of Plant Biology, 62* (5), 374-386.

Liao, Q., Zhou, T., Yao, J. Y., Han, Q. F., Song, H. X., Guan, C. Y., . . . Singh, A. K. (2018). Genome-scale characterization of the vacuole nitrate transporter Chloride Channel (CLC) genes and their transcriptional responses to diverse nutrient stresses in allotetraploid rapeseed. *PLoS One, 13* (12).

Lischer, H. E., & Excoffier, L. (2012). PGDSpider: an automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics, 28* (2), 298-299.

Liu, J. (2016). "The integrative species concept" and "species on the speciation way". . *Biodiversity Science, 24* (9), 1004-1008. doi:10.17520/biods.2016222

Love, M., Anders, S., & Huber, W. (2014). Differential analysis of count data–the DESeq2 package. *Genome Biol, 15* (550), 10.1186.

Lu, T., Li, M.-R., Ding, N., Wang, Z.-H., Lan, L.-Z., Gao, X., & Li, L.-F. (2019). Genetic and epigenetic mechanisms underpinning the adaptive radiation of *Aquilegia* species. *bioRxiv* , 782821.

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

Magadlela, A., Morcillo, R. J. L., Kleinert, A., Venter, M., Steenkamp, E., & Valentine, A. (2019). Glutamate dehydrogenase is essential in the acclimation of V*irgilia divaricata* , a legume indigenous to the nutrient-poor Mediterranean-type ecosystems of the Cape Fynbos.*Journal of plant physiology, 243* , 153053.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., . . . Daly, M. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data.*Genome research, 20* (9), 1297-1303.

Morente-Lopez, J., Lara-Romero, C., Garcia-Fernandez, A., Rubio Teso, M. L., Prieto-Benitez, S., & Iriondo, J. M. (2021). Gene flow effects on populations inhabiting marginal areas: origin matters. *Journal of Ecology, 109* (1), 139-153.

Mulcare, D. M. (2004). NGS Toolkit, Part 8: The National Geodetic Survey NADCON Tool. *Professional Surveyor Magazine* .

Munz, P. A. (1946). *Aquilegia: the cultivated and the wild Columbines* : Bailey Hortorium.

Nagano, A. J., Kawagoe, T., Sugisaka, J., Honjo, M. N., Iwayama, K., & Kudoh, H. (2019). Annual transcriptome dynamics in natural environments reveals plant seasonal adaptation. *Nature plants, 5* (1),

74-83.

Nguyen, L.-T., Schmidt, H. A., Von Haeseler, A., & Minh, B. Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular biology and evolution, 32* (1), 268-274.

Omidbakhshfard, M. A., Fujikura, U., Olas, J. J., Xue, G.-P., Balazadeh, S., & Mueller-Roeber, B. (2018). GROWTH-REGULATING FACTOR 9 negatively regulates arabidopsis leaf growth by controlling ORG3 and restricting cell proliferation in leaf primordia. *PLoS genetics, 14* (7), e1007484.

Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., . . . Reich, D. (2012). Ancient admixture in human history. *Genetics, 192* (3), 1065-1093.

Philippe, R. N., Ralph, S. G., Kulheim, C., Jancsik, S. I., & Bohlmann, J. (2010). Poplar defense against insects: genome analysis, full-length cDNA cloning, and transcriptome and protein analysis of the poplar Kunitz-type protease inhibitor family. *New Phytologist, 184* (4), 865-884.

Pinheiro, F., Dantas-Queiroz, M. V., & Palma-Silva, C. (2018). Plant species complexes as models to understand speciation and evolution: a review of South American studies. *Critical reviews in plant sciences, 37* (1), 54-80.

Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A., & Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nature genetics, 38* (8), 904-909.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., . . . Daly, M. J. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American journal of human genetics, 81* (3), 559-575.

Riboni, M., Robustelli Test, A., Galbiati, M., Tonelli, C., & Conti, L. (2016). ABA-dependent control of GIGANTEA signalling enables drought escape via up-regulation of FLOWERING LOCUS T in *Arabidopsis thaliana* . *Journal of experimental botany, 67* (22), 6309-6322.

Ritchie, M. D., Holzinger, E. R., Li, R., Pendergrass, S. A., & Kim, D. (2015). Methods of integrating data to uncover genotype–phenotype interactions. *Nature reviews genetics, 16* (2), 85-97.

Rodriguez-Celma, J., Chou, H., Kobayashi, T., Long, T. A., & Balk, J. (2019). Hemerythrin E3 Ubiquitin Ligases as Negative Regulators of Iron Homeostasis in Plants. *Other, 10* .

Ronco, F., Matschiner, M., Bohne, A., Boila, A., Buscher, H. H., El Taher, A., . . . Kahmen, A. (2021). Drivers and dynamics of a massive adaptive radiation in cichlid fishes. *Nature, 589* (7840), 76-81.

Rundell, R. J., & Price, T. D. (2009). Adaptive radiation, nonadaptive radiation, ecological speciation and nonecological speciation. *Trends in Ecology & Evolution, 24* (7), 394-399.

Schluter, D. (2000). *The ecology of adaptive radiation* : OUP Oxford.

Seong, E.-S., Choi, D.-I., Cho, H.-S., Lim, C.-K., Cho, H.-J., & Wang, M.-H. (2007). Characterization of a stress-responsive ankyrin repeat-containing zinc finger protein of Capsicum annuum (CaKR1). *BMB Reports, 40* (6), 952-958.

Shi, G., Herrera, F., Herendeen, P. S., Clark, E. G., & Crane, P. R. (2021). Mesozoic cupules and the origin of the angiosperm second integument. *Nature* . doi:10.1038/s41586-021-03598-w

Shu, K., Chen, F., Zhou, W., Luo, X., Dai, Y., Shuai, H., & Yang, W. (2018). ABI4 regulates the floral transition independently of ABI5 and ABI3. *Molecular Biology Reports, 45* (6), 2727-2731.

Shu, K., Chen, Q., Wu, Y., Liu, R., Zhang, H., Wang, P., . . . Liu, C. (2016). ABI 4 mediates antagonistic effects of abscisic acid and gibberellins at transcript and protein levels. *The Plant Journal, 85* (3), 348-361.

Shu, K., Zhang, H., Wang, S., Chen, M., Wu, Y., Tang, S., . . . Xie, Q. (2013). ABI4 regulates primary seed dormancy by regulating the biogenesis of abscisic acid and gibberellins in Arabidopsis. *PLoS Genet, 9* (6), e1003577.

Smith, B. J. (2007). boa: an R package for MCMC output convergence assessment and posterior inference. *Journal of statistical software, 21* (11), 1-37.

Smith, S., Bernatchez, L., & Beheregaray, L. B. (2013). RNA-seq analysis reveals extensive transcriptional plasticity to temperature stress in a freshwater fish species. *BMC Genomics, 14* (1), 1-12.

Soltis, P. S., Folk, R. A., & Soltis, D. E. (2019). Darwin review: angiosperm phylogeny and evolutionary radiations. *Proceedings of the Royal Society B, 286* (1899), 20190099.

Steinthorsdottir, M., Coxall, H., De Boer, A., Huber, M., Barbolini, N., Bradshaw, C., . . . Henderiks, J. (2021). The Miocene: The future of the past. *Paleoceanography and Paleoclimatology, 36* (4), e2020PA004037.

Suarez-Gonzalez, A., Hefer, C. A., Christe, C., Corea, O., Lexer, C., Cronk, Q. C., & Douglas, C. J. (2016). Genomic and functional approaches reveal a case of adaptive introgression from *Populus balsamifera* (balsam poplar) in *P. atrichocarpa* (black cottonwood). *Molecular ecology, 25* (11), 2427-2442.

Surya, R. (2020). In-silico characterization of *At5g18130* gene in*Arabidopsis thaliana* with emphasis on its expression patterns and functional aspects. *Journal of Pharmacognosy and Phytochemistry, 9* (5), 987-995.

Wong, E. L., Nevado, B., Osborne, O. G., Papadopulos, A. S., Bridle, J. R., Hiscock, S. J., & Filatov, D. A. (2020). Strong divergent selection at multiple loci in two closely related species of ragworts adapted to high and low elevations on Mount Etna. *Molecular ecology, 29* (2), 394-412.

Wu, Z., Raven, P., & Hong, D. (2001). *Flora of China. Volume 6* : Science Press.

Zhang, C., Dong, S.-S., Xu, J.-Y., He, W.-M., & Yang, T.-L. (2019). PopLDdecay: a fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. *Bioinformatics, 35* (10), 1786-1788.

Zhang, W., Wang, H., Dong, J., Zhang, T., & Xiao, H. (2021). Comparative chloroplast genomes and phylogenetic analysis of Aquilegia.*Applications in plant sciences, 9* (3), e11412.

Zhao, Z., Heideman, N., Bester, P., Jordaan, A., & Hofmeyr, M. D. (2020). Climatic and topographic changes since the Miocene influenced the diversification and biogeography of the tent tortoise (Psammobates tentorius) species complex in Southern Africa. *BMC evolutionary biology, 20* (1), 1-33.

FIGURE LEGENDS

Figure . Geographical distributions of sampled *Aquilegia viridiflora* . Populations in the NE, EL, CN and NW groups have different floral traits. The scale in the figure is 1:20,000,000.

Figure . Phylogenetic relationships of the four *Aquilegia viridiflora* lineages with genetic structure. Phylogenetic relationships were inferred according to the maximum likelihood (ML) method, and the genetic structure showed ADMIXTURE proportions of genetic clusters for each individual of the four lineages at the best K value (K = 4). Pale turquoise, moccasin, light green and salmon represent the NE, EL, CN and NW lineages, respectively.

Figure . The demographic history of *Aquilegia viridiflora* . (A) Results for the D-statistics; the red dot represents the Z-scores. (B) Ancestral population sizes, population divergence times, and migration rates were assessed by G-PhoCS.

Figure . Phenotype of*Aquilegia viridiflora* . (A) Corolla diameter. (B) Petal length (not including spur). (C) Angle between petals and spurs. (D) Spur length. (E) Pistil length. (F) Stamen length. (G) Calyx length. (H) Number of inflorescences. (I) Leaf area. (J) Chlorophyll content. (K) Leaf perimeter. (L) Plant height.

Figure . WGCNA module and trait associations. Gene modules that are significantly related to a phenotype are indicated in bold.

Figure . Key genes for the adaptive evolution of *Aquilegia viridiflora* in early speciation. (A) Candidate genes for adaptive evolution that are differentially expressed in EL and NW. The yellow dots indicate that the SNP in the sliding window was selected by the environment; red and blue dots represent other regions in the genome. (B) The expression level of*Aqcoe5G459400*in flowers between different lineages *A. viridiflora* . (C) Genes related to the *Aqcoe5G459400* identified by WGCNA.

Figure S1 Phylogenetic relationship between *Aquilegia viridiflora* complex and its sympatric other columbine species. (A) Neighbor-Joining (NJ) tree; (B) Maximum likelihood (ML) tree.

Figure S2 Genetic structure of *Aquilegia viridiflora*complex. (A) Cross-validation results corresponding to different K values; (B) Principal component analysis (PCA) plot for the 66 *A. viridiflora* individuals based on the first two principal component; (C) LD decay of the four lineages of *A. viridiflora* . The x axis stands for physical distance.

Figure S3 K-means cluster analysis of phenotype based on the first two principal component.

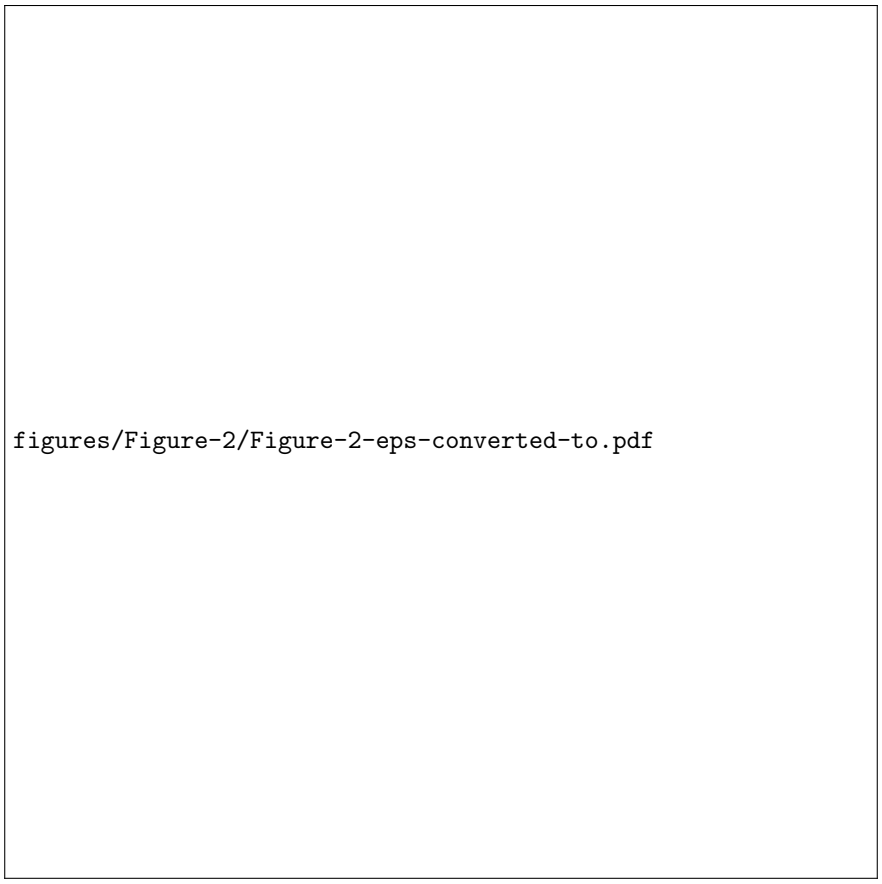
Figure S4 Results of correlation analysis between phenotypes. * represented significant relation between different phenotypes $0.01 < p < 0.05$; ** represented significant relation between different phenotypes $0.001 < p < 0.01$; *** represented significant relation between different phenotypes $p < 0.001$.

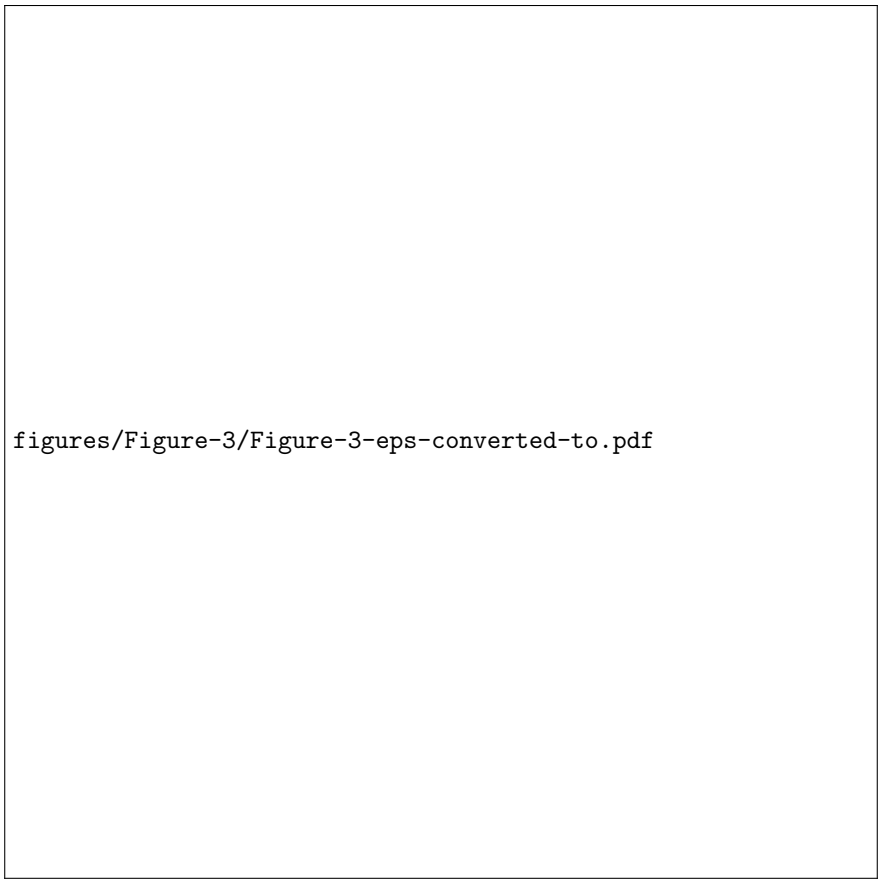
Figure S5 Population genetic analysis. (A) Nucleotide diversity (ϑ_π); (B) Taijima's D; (C) Pairwise F_{ST}between different lineages.

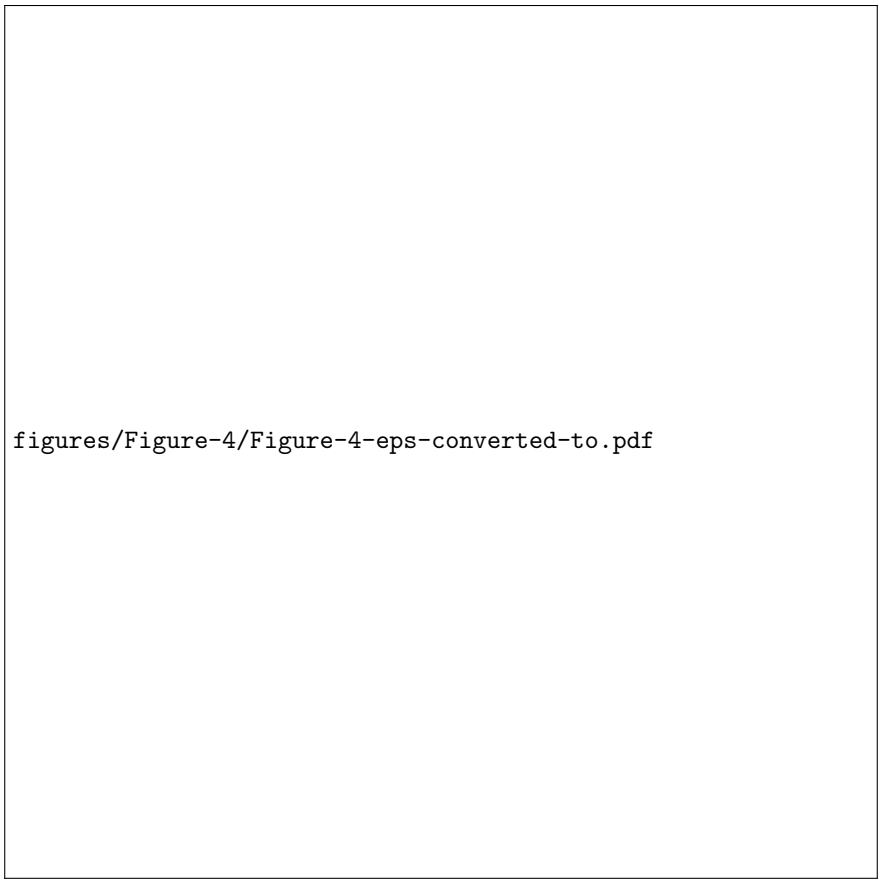
Figure S6 Candidate genes for adaptive evolution that were differentially expressed. (A) NE and EL; (B) NE and CN; (C) NE and NW; (D) EL and CN; (E) CN and NW. The yellow dots represented that the SNP in this sliding window has been selected by the environment; the red and blue dots represented other regions in the genome.

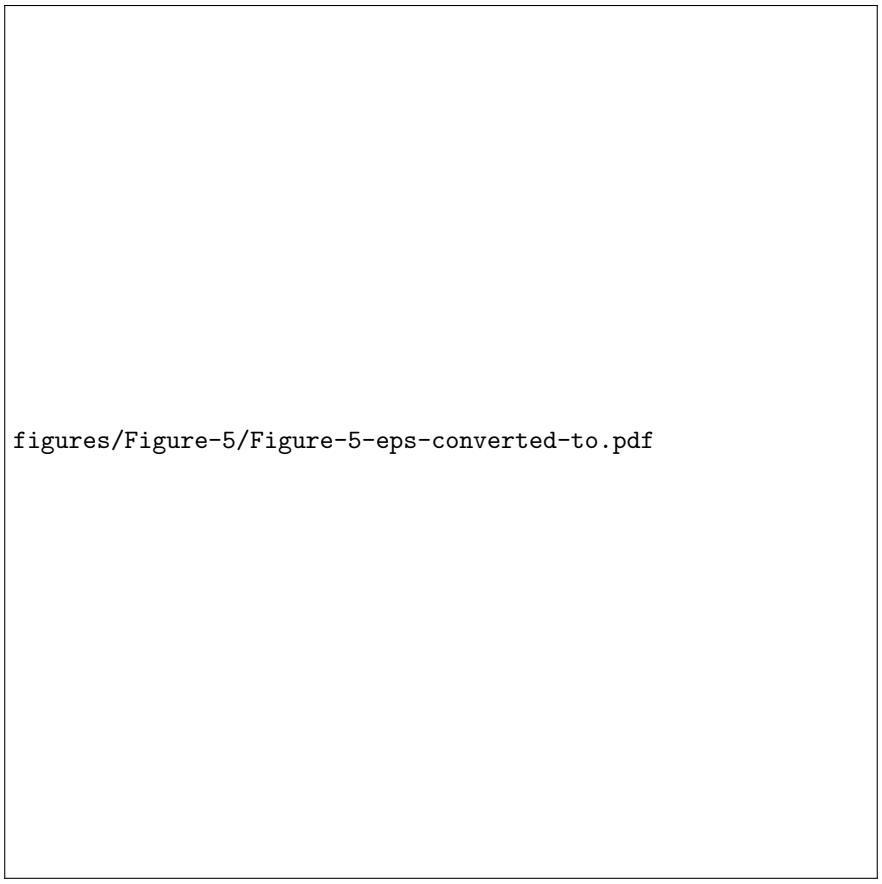
Hosted file

`Figure_1.eps` available at https://authorea.com/users/500580/articles/711447-genome-and-transcriptome-variation-in-aquilegia-viridiflora-underlies-the-adaptive-evolution-to-the-environment-in-the-early-stage-of-speciation









Hosted file

Figure_6.eps available at https://authorea.com/users/500580/articles/711447-genome-and-transcriptome-variation-in-aquilegia-viridiflora-underlies-the-adaptive-evolution-to-the-environment-in-the-early-stage-of-speciation

Hosted file

Table.docx available at https://authorea.com/users/500580/articles/711447-genome-and-transcriptome-variation-in-aquilegia-viridiflora-underlies-the-adaptive-evolution-to-the-environment-in-the-early-stage-of-speciation