Research Article

Comparative Analysis of the Complete Chloroplast Genomes in Allium Subgenus Cyathophora (Amaryllidaceae): Phylogenetic Relationship and Adaptive Evolution

Xin Yang, Deng-Feng Xie, Jun-Pei Chen, Song-Dong Zhou, Yan Yu, and Xing-Jin He*  

Key Laboratory of Bio-Resources and Eco-Environment of Ministry of Education, College of Life Sciences, Sichuan University, Chengdu, China

Correspondence should be addressed to Xing-Jin He; xjhe@scu.edu.cn

Received 13 August 2019; Accepted 7 December 2019; Published 17 January 2020

Academic Editor: Marcelo A. Soares

Copyright © 2020 Xin Yang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Recent advances in molecular phylogenetics provide us with information of Allium L. taxonomy and evolution, such as the subgenus Cyathophora, which is monophyletic and contains five species. However, previous studies detected distinct incongruence between the nrDNA and cpDNA phylogenies, and the interspecies relationships of this subgenus need to be furtherly resolved. In our study, we newly assembled the whole chloroplast genome of four species in subgenus Cyathophora and two allied Allium species. The complete cp genomes were found to possess a quadripartite structure, and the genome size ranged from 152,913 to 154,174bp. Among these cp genomes, there were subtle differences in the gene order, gene content, and GC content. Seven hotspot regions (infA, rps16, rps15, ndhF, trnG-UCC, trnC-GCA, and trnK-UUU) with nucleotide diversity greater than 0.02 were discovered. The selection analysis showed that some genes have elevated Ka/Ks ratios. Phylogenetic analysis depended on the complete chloroplast genome (CCG), and the intergenic spacer regions (IGS) and coding DNA sequences (CDS) showed same topologies with high support, which revealed that subgenus Cyathophora was a monophyletic group, containing four species, and A. cyathophorum var. farreri was sister to A. spicatum with 100% bootstrap value. Our study revealed selective pressure may exert effect on several genes of the six Allium species, which may be useful for them to adapt to their specific living environment. We have well resolved the phylogenetic relationship of species in the subgenus Cyathophora, which will contribute to future evolutionary studies or phylogeographic analysis of Allium.

1. Introduction

Subgenus Cyathophora (R. M. Fritsch) R. M. Fritsch is a small group of Allium that has been put forward lately [1]. The special subgenus Cyathophora contains about six species and one variety according to Li et al. [2]; besides, A. spicatum (Prajin) N. Friesen has a wide distribution range, extending from China to Nepal, while the rest of them are endemic species in China and mainly distributed in the southeastern margin of the Qinghai-Tibet Plateau (QTP): A. mairei Lév, A. kingdonii Stearn, A. rhynchogynum Diels, A. trifurcatum (F. T. Wang and Tang) J. M. Xu, and A. cyathophorum Bur. and Franch and its variety A. cyathophorum var. farreri (Stearn) Stearn. Although it contains a small number of species, the boundary of subgenus Cyathophora and the involved species have experienced some alterations with the development of molecular biology. In previous study, A. spicatum was classified at different taxonomic levels because of its idiographic spicate inflorescence based on morphological and molecular evidences [3]. Five species have been proposed by Huang et al. about subgenus Cyathophora [4]: A. mairei, A. rhynchogynum, A. cyathophorum, A. cyathophorum var. farreri, and A. spicatum, while A. kingdonii and A. trifurcatum did not belong to this group. Micromorphological and cytological features supported that the subgenus Cyathophora is a monophly and contains five species [5, 6]. Among species of the subgenus Cyathophora, A. spicatum grows in the droughty western QTP with the extremely abnormal spicate inflorescence [3], while A. cyathophorum and A. cyathophorum var. farreri with the
umbel inflorescence stretch to the moist HMR [7] (Figure 1). Furthermore, Li et al. [6] suggested that A. cyathophorum and A. farreri were independent species based on molecular phylogeny and the striking distinctiveness in micromorphology. A. rhynochogynum has never been sampled since it was published in 1912 [8]. We also performed a lot of field work to collect it but failed. The Flora of China recorded that A. rhynochogynum only distributed in northwest of Yunnan province in China. Therefore, we speculate that A. rhynochogynum might become extinct or there is an identification error in previous research studies. Li et al. [6] performed phylogenetic and biogeographic analyses for A. cyathophorum and A. spicatum based on chloroplast and nuclear ribosomal DNA and detected distinct different topologies between these two molecular methods, in which A. cyathophorum showed close relationship with A. spicatum in nuclear DNA tree but was sister to A. cyathophorum var. farreri in cpDNA tree [4, 6]. Other than this, the relationship between these species is not exactly determined, and phylogenetic analysis using single or several combined chloroplast fragments does not solve the problem effectively, and the complete cp genome can well resolve the relationship of subgenus Cyathophora. Hence, it is imperative to reconstruct the relationship of subgenus Cyathophora and clarify the contained species depending on the complete chloroplast genomes. To evaluate the subgenus Cyathophora resources comprehensively, we also need more efficient molecular markers.

Chloroplast is one of the basic organelles in plant cells, which is in charge of photosynthesis of green plants [9]. The chloroplast genomes have a highly conserved structure and gene content, which have a quadripartite structure composed by large single-copy (LSC) and small single-copy (SSC) regions separated by two parts of inverted repeat (IR) [10, 11]. Previous studies suggested that genome size of angiosperms ranged from 120 kb to 170 kb with gene number changed from 120 to 130 [12]. Complete chloroplast genome has long been a core issue in plant molecular evolution and systematic studies because of its over-simplified structure, highly conservative sequence, and maternal hereditary traits [13]. Since the complete cp genome analysis can provide more genetic information contrasted with just single or few cpDNA fragments [14], by using cp genome sequences, many long existing phylogenetic problems of different angiosperms at various taxonomic levels have been successfully resolved [15–20].

In addition to exploring phylogenetic studies, the whole cp genome has important significance to reveal the photosynthesis mechanism, metabolic regulation, and adaptive evolution of plants. Research has shown that adaptive evolution is mainly promoted by evolutionary processes like natural selection, which affects genetic changes caused by genetic recombination and mutations [21]. Many recent studies have analyzed the selection pressures that undergo by species in the evolutionary processes based on complete chloroplast genome, for example, a positive selection for the \textit{atpF} gene may suggest that it has made an important impact on the divergence in deciduous and evergreen oak tree [9], and there also existed positive selection on \textit{ycf2} in watercress chloroplasts [22]. With the development of sequencing technology, the number of cp genomic sequences has increased dramatically in recent years. However, a few plastid genomes of \textit{Allium} were reported until now, and it is necessary to develop more complete chloroplast genome in \textit{Allium} for future phylogenetic and evolutionary research studies.

In our report, we assembled and characterized the complete cp genome sequence of the six \textit{Allium} species using next-generation sequencing technologies to (1) reveal common structural patterns and hotspot regions, (2) gain a better understanding of the relationship about subgenus Cyathophora based on complete chloroplast genome, and (3) investigate adaptive evolution in the cp genomes of the six \textit{Allium} species. We hope our study will provide valuable genetic resources for further evolutionary studies about subgenus Cyathophora.

2. Materials and Methods

2.1. Plant Materials, DNA Extraction, and Sequencing. Fresh leaves of A. cyathophorum, A. cyathophorum var. farreri, A. spicatum, A. mairei, A. trifurcatum, and A. kingdonii were collected from different places (Table 1).

2.2. Chloroplast Genome Sequence Assembly and Annotation. We firstly used FastQC v0.11.7 to assess the quality of all reads [24]. To select the best reference, we filtered the chloroplast genome related reads by mapping all reads to the published chloroplast genome sequences in \textit{Allium}. SOAPdenovo2 was used to assemble all relevant reads into contigs [25]. The clean reads were assembled using the program NOVOPlasty [26] with the complete chloroplast genome of its close relative \textit{A. cepa} as the reference (GenBank accession no. KM088014). Geneious 11.0.4 was used to finish the annotation of the assembled chloroplast genome, and it was corrected manually after comparison with references [27]. The circular plastid genome maps were generated utilizing the OGDRAW program [28]. The GenBank accession numbers of A. cyathophorum, A. cyathophorum var. farreri, A. spicatum, A. mairei, A. trifurcatum, and A. kingdonii are MK820611, MK931245, MK931246, MK820615, MK931247, and MK294559, respectively.

2.3. Repeat Sequences and Simple Sequence Repeat (SSR) Analysis. REPuter [29] was selected to investigate the location and size of repeat sequences, which included four types of repeats in the chloroplast genomes about the six
Allium species. The sequence identity and minimum length of repeat size was set to >90% and 30bp, with the hamming distance of 3. We used IMEx, ImperfectMicrosatelliteExtractor (http://43.227.129.132:8008/IMEX/imex_advanced.html) [30], to find chloroplast SSRs in six chloroplast genome sequences of Allium. Its specifications were set up as follows: the minimum number of repeats for mononucleotide, dinucleotides, trinucleotides, tetranucleotides, pentanucleotide and hexanucleotides was 10, 5, 5, 4, 3, and 3, respectively, the repeat type was imperfect, the imperfection % was Mono: 10%, Di: 10%, Tri: 15%, Tetra: 20%, Penta: 5%, and Hexa: 5%, mismatches allowed in pattern were Mono: 1, Di: 1, Tri: 2, Tetra: 4, Penta: 0, and Hexa: 3, and the level of standardization was level 1 standardization.

2.4. Codon Usage Analysis. Codon usage of the species in subgenus Cyathophora was analyzed by the software of CodonW [31]. Protein-coding genes (CDS) were selected with the following filter requirements: (1) each CDS was longer than 300 nucleotides [18, 32]; (2) repeat sequences

Table 1: Samples information.

<table>
<thead>
<tr>
<th>Species</th>
<th>Location</th>
<th>Geographical coordinate</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. cyathophorum</td>
<td>Mangkang, Tibet</td>
<td>29°43′24″N, 98°31′51″E</td>
</tr>
<tr>
<td>A. cyathophorum var. farreri</td>
<td>Zhouqu, Gansu</td>
<td>33°47′10″N, 104°22′7″E</td>
</tr>
<tr>
<td>A. spicatum</td>
<td>NaiDong, Tibet</td>
<td>29°13′40″N, 91°45′36″E</td>
</tr>
<tr>
<td>A. mairei</td>
<td>Shangri-La, Yunnan</td>
<td>27°25′19″N, 100°09′22″E</td>
</tr>
<tr>
<td>A. trifurcatum</td>
<td>Shangri-La, Yunnan</td>
<td>27°52′22″N, 99°43′10″E</td>
</tr>
<tr>
<td>A. kingdonii</td>
<td>Nyingchi, Tibet</td>
<td>29°37′27″N, 94°39′2″E</td>
</tr>
</tbody>
</table>

Figure 1: The morphological characters of flowers of subgenus Cyathophora species. (a) A. cyathophorum. (b) A. cyathophorum var. farreri. (c) A. mairei. (d) A. spicatum.
were deleted. Totally, 53 CDS of each species in *Allium* were selected for further study.

2.5. Genome Comparison (IR Contraction and Expansion). The mVISTA program was chosen to analyze the whole sequence similarity of all six *Allium* species with Shuffle-LAGAN model [33], using the chloroplast genomes to compare their difference in sequences at the chloroplast genome level and *A. cyathophorum* as the reference. The boundaries between single copy regions (LSC and SSC) and inverted repeats (IR) regions among the six chloroplast genome sequences were compared by using Geneious v11.0.4 software [27].

2.6. Hotspot Regions Identification in Subgenus Cyathophora. To analyze nucleotide diversity (Pi), we extracted the shared 112 genes of the six species in *Allium* after alignment. DnaSP 5.10 was employed to calculate the nucleotide variability [34].

2.7. Gene Selective Pressure Analysis of Six Allium Plastomes. To investigate selection pressures, nonsynonymous (Ka) and synonymous (Ks) substitution rates of 65 selected protein-coding genes between the cp genomes of subgenus *Cyathophora* and the other two *Allium* species were calculated by KaKs Calculator version 2.0 [35].

2.8. Subgenus Cyathophora Phylogenomic Analysis Based on Chloroplast Genome. Phylogenetic analysis of subgenus *Cyathophora* was totally depended on twenty-nine complete chloroplast genome sequences, which were twenty-one species of *Allium* (including 6 newly assembled species; 15 other species of *Allium* were collected from NCBI), six species of *Lilium*, and two species of *Asparagus* as the out groups (Table S1). Three different databases were used to build the phylogenetic tree, which include the complete genome sequences, the IGS sequences, and all CDS sequences, and three different methods, Bayesian-inference (MrBayes v3.2), maximum parsimony (PAUP-version4.0), and maximum likelihood (RAxML8.0), were used to build the tree. The sequences were aligned using MAFFT [36] in Geneious 11.0.4 with the set parameters and manually trimmed. GTR + I + G was selected as the best model using software ModelTest v3.7 [37]. Maximum likelihood (ML) analyses were performed using RAxML8.0 with 1000 bootstrap replications [38]. PAUP was used to conduct maximum parsimony (MP) analyses [39]. MP was run using a heuristic search with 1000 random addition sequence replicates with the tree-bisection-reconnection (TBR) branch-swapping tree search criterion. Bayesian inference (BI) was executed with Mrbayes v3.2 [40], and the Markov chain Monte Carlo (MCMC) analysis was run 1 × 10⁸ generations. The trees were sampled every 1000 generations: the first 25% were discarded as burn in and the remaining trees were used to establish a 50% majority rule consensus tree. When the average standard deviation of the splitting frequency was kept below 0.001, it was considered that the stationarity is achieved.

3. Results

3.1. Chloroplast Genome Organization and Gene Content in Six Species. These six acquired *Allium* cp genomes were detected to have a circular DNA structure of angiosperm cp genomes that comprises LSC, SSC, and two IR regions (Figure 2). The sizes of the six CP genomes ranged from 152,913 bp for *A. mairei* to 154,174 bp for *A. cyathophorum*, which were similar with other *Allium* CP genomes [41]. The size varied from 82,493 bp (*A. mairei*) to 83,423 bp (*A. kingdonii*) in the LSC region, from 17,811 bp (*A. kingdonii*) to 21,706 bp (*A. trifurcatum*) in the SSC region, and from 24,561 bp (*A. trifurcatum*) to 26,467 bp (*A. cyathophorum*) in the IR region (Figure 2). The entire GC content of the cp genome sequences was 36.8–36.9%, and the GC contents of the LSC, SSC, and IR regions were 34.6–34.8%, 29.5–31.2%, and 42.7–43.1%, respectively. A total of 132 genes were discovered from the complete cp genome: 8 ribosomal RNA (rRNA) genes, 86 protein-coding genes, and 38 transfer RNA (tRNA) genes (Table 3).

3.2. Repeat and Simple Sequence Repeat (SSR) Analysis. Many research studies of cp genomes revealed that repeat sequences have been widely used in phylogeny, population genetics, and other studies [42]. Four types of repeats (forward repeats, reverse repeats, complement repeats, and palindromic repeats) were detected in the six *Allium* species. There were only 3 complement repeats in *A. cyathophorum*, while the other species did not have. The number of repeats varied from 37 to 77 in the six species; the *A. cyathophorum* showed the most abundant number of repeats, including 29, 40, 5 and 3 palindromic forward reverse and complement repeats, respectively. The number of forward repeats ranged from 15 to 40, the number of palindromic forward repeats ranged from 17 to 29, and the number of reverse repeats ranged from 1 to 5 (Figure 3). The lengths of forward, palindromic, and reverse repeats ranged from 30 to 267 bp, and most of them were concentrated in 30–50 bp (81.48%), while those of 50–70 bp (9.09%), >100 bp (6.40%), and 70–90 bp (3.03%) were less common (Figure S1). Earlier reports recommend that the appearance of the repeats indicates that this locus is a staple hot-spot for reconfiguration of the genome [43–45]. Nevertheless, these repeats are valuable for developing genetic markers in population genetics studies [46, 47].

SSRs, also called as microsatellites, are 1- to 6-bp repeating sequences that are extensively distributed in the chloroplast genome. SSRs are highly polymorphic and dominant, which are valuable markers for study involving gene flow, population genetics, and gene mapping [48]. In this study, six classes of SSRs (mono-, di-, tri-, tetra-, penta-, and hexanucleotide repeats) were found in the cp genome of the six species, whereas the number of hexanucleotide repeats ranged from 1 to 3 in the six species, and the
pentanucleotide repeats just existed in *A. kingdonii* and *A. spicatum*. The total number of SSRs in the genome of the six *Allium* species was 185 in *A. cyathophorum*, 158 in *A. cyathophorum* var. *farreri*, 159 in *A. spicatum*, 171 in *A. mairei*, 201 in *A. trifurcatum*, and 165 in *A. kingdonii* (Figure 4(a)). The highest number was mononucleotide repeat, which accounted for about 30.41% of the total SSRs (Figure 4(c)); the number ranged from 36 in *A. kingdonii* to 71 in *A. trifurcatum*, and all mononucleotide repeats are composed of A or T bases; these conclusions were unanimous in previous studies that SSRs in cp genomes usually contained short polyA or polyT repeats [49], while those of dinucleotide repeats (28.20%), trinucleotide repeats (20.79%), tetranucleotide repeats (19.15%), pentanucleotide repeats (0.38%), and hexanucleotide repeats were the least abundant (1.06%). In the whole SSR locus, the SSRs located in the LSC area are much more than those in the SSC and IR areas (Figure 4(b)), which is identical with previous research studies that SSRs are unevenly distributed in cp genomes [50].

![Gene map of *A. cyathophorum*, *A. cyathophorum* var. *farreri*, *A. spicatum*, *A. mairei*, *A. trifurcatum*, and *A. kingdonii* complete chloroplast genomes. The circle inside genes is transcribed clockwise, and the outside genes are transcribed counterclockwise. Genes belonging to different functional groups are represented by distinct colors.](image-url)
universal amino acids, respectively. As recently discovered, the result indicated that a total of 13218 codons encoding usage frequency was calculated for the six sequences of 53 protein-coding genes (CDS), the codon 1 stands for more frequent use of a codon. In view of the demonstrates lack of bias, whereas RSCU value greater than codon usage in coding sequences, in which RSCU less than 1 (RSCU) is a method that estimates nonuniform synonymous selections [51, 52]. he relative synonymous codon usage by evolutionary factors that affect gene mutations and different frequencies of use in plant genomes, which was caused nomenon that the synonymous codons usually have dif-

Codon usage bias is a phenomenon that the synonymous codons usually have different frequencies of use in plant genomes, which was caused by evolutionary factors that affect gene mutations and selections [51, 52]. The relative synonymous codon usage (RSCU) is a method that estimates nonuniform synonymous codon usage in coding sequences, in which RSCU less than 1 demonstrates lack of bias, whereas RSCU value greater than 1 stands for more frequent use of a codon. In view of the sequences of 53 protein-coding genes (CDS), the codon usage frequency was calculated for the six Allium cp genomes (Table 4). Altogether, the number of codons ranged from 33058 in A. kingdonii to 23791 in A. mairei. In addition, the result indicated that a total of 13218 codons encoding leucine in the cp genomes of the six species and 1453 codons encoding cysteine as the most common and least common universal amino acids, respectively. As recently discovered in other cp genomes of plants, our study revealed that except tryptophan and methionine, there was preference in the use of synonymous codons, and the RSCU value of 30 codons exceeded 1 for each species, and they were A or T-ending codons. The result is in accordance with other researches, which showed that the codon usage preference for A/T ending in plants [53–55].

3.4. Comparative Analysis of the Chloroplast Genomes among Six Species of Allium. mVISTA online software in the Shuffle-LAGAN mode was employed to analyze the comprehensive sequence discrepancy of the six chloroplast genomes of Allium with the annotation of A. cyathophorum as a reference. In this study, the whole chloroplast genome alignment showed great sequence consistency of the six cp genomes, indicating that Allium cp genomes are very similar in sequence.
conservative (Figure 5). We found that among the six cp genomes, their IR region is more conserved compared to the LSC and SSC regions, which is similar with other plants [56, 57]. Furthermore, as we have found in other angiosperms, the coding areas were more conserved than the noncoding areas, and there were more variations in the intergenic spacers of the LSC and SSC areas, whereas the IR areas presented a lower sequence divergence [58, 59]. *A. cyathophorum* var. *farreri* had the highest sequence similarity to *A. cyathophorum* in sequence identity analysis. Noncoding regions displayed varying degrees of sequence differences in these six *Allium* cp genomes, including *trnK-rps16, trnS-trnG, atpH-atpI, petN-psbM, trnT-psbD, trnF-ndhI, accD-psal*, and *petA-psbL*. The coding areas with significant diversity contain *matK, rps16, rpoC2, infA, ycf1, ndhF*, and *rps15* genes. The highly diverse regions found in this study may be used to develop molecular markers that can improve efficiency to study phylogenetic relationships within the *Allium* species.

Though the cp genome is usually well conserved, having typical quadripartite structure, gene number, and order, a phenomenon recognized as ebb and flow exists, and this is where the IR area often expands or contracts [60]. Expansion and contraction of IR region is related to the size variations in the cp genome and has great differences in its evolution [61, 62]. We compared the IR/SC boundary areas of the six *Allium* cp genomes, and we found that there are obvious differences in the IR/LSC and IR/SSC connections (Figure 6). At the boundary of LSC/IRa junction, *rps19* gene of different species distance the boundary were from 1 to 81 bp, while the *rpl22* genes distance the border were from 29 to 273 bp. At the boundary of LSC/IRb connections, the *psbA* genes distance the border were from 108 to 605 bp. The inverted repeat b (IRb)/SSC border located in the coding region, and the *ycf1* genes of the six species with a region ranged from 4193 to 5223 bp located in the SSC regions, while the *ycf1* gene of *A. trifurcatum* all located in the SSC region. The shorter *ycf1* gene crossed the inverted repeat (IRa)/SSC boundary, with 56–919 bp locating in the SSC regions. And the *ndhF* genes were situated in the SSC region, which distance from the IRa/SSC boundary ranged from 1 to 1962 bp. Undoubtedly, the full-length differences in the sequence of the six cp genomes are caused by changes in the IR/SC boundaries.

### 3.5. Hotspot Regions Identification in Subgenus Cyathophora

We totally extracted the shared 112 genes of the six species in chloroplast genomes; the nucleotide variability (Pi) ranged from 0.00041 (*rrn16*) to 0.08125 (*infA*) among these shared...
genes (Figure 7; Table S2). Seven genes (infA, rps16, rps15, ndhF, trnG-UCC, trnC-GCA, and trnK-UUU) were considered to be hotspot regions with a nucleotide diversity greater than 0.02. These regions can be used to develop useful markers for phylogenetic analysis and distinguish the species in *Allium*.

3.6. Synonymous (Ks) and Nonsynonymous (Ka) Substitution Rate Analysis. The Ka/Ks ratio is a significant index for understanding the evolution of protein-coding genes to assess gene differentiation rates and to determine whether positive, purified, or neutral selections have been performed; a Ka/Ks ratio >1 illustrates positive selection and Ka/Ks < 1 illustrates purifying selection, while the ratio of Ka/Ks close to 1 illustrates neutral selection [63]. In our study, the Ka/Ks ratio was calculated for 65 shared protein-coding genes in all six chloroplast genomes (Table S3), and the results are shown in Figure 8. The conservative genes with Ka/Ks ratio of 0.01, indicating powerful purifying selection pressure, were rpl2, rpl32, psaC, psbA, rpoC2, petN, psbZ, psaB, psaL, and psbT, when the averaging Ka/Ks method showed ycf1 and ycf2 genes with Ka/Ks > 1, which shows that they may undergo some selective pressure among the six *Allium* species. The Ka/Ks ratios ranging from 0.5 to 1 were found for matK, rps16, psaI, cemA, petA, and rpl20, representing relaxed selection. The majority (56 of 65 genes) had an average Ka/Ks ratio ranging from 0 to 0.49 for the six compared groups, indicating that most genes were under purifying selection. Other than this, four genes (matK, rpoB, petA, and rpoA) with Ka/Ks > 1 in one or more pairwise comparisons (Figure 8) suggest that these genes may undergo selective pressure which is unknown, which is very important for researching the evolution of species.

3.7. Phylogenetic Analysis of Subgenus Cyathophora Depends on Chloroplast Genome. The cp genome of sequence is significant and helpful to construct phylogenetic relationships and explore the evolutionary history in many previous reports [64, 65]. To explore the phylogenetic relationship of the six *Allium* species, we constructed the phylogenetic tree using three different methods and databases containing twenty-one *Allium* species, six *Lilium* species, and two *Asparagus* species as the out groups (Figure 9). Three databases of the complete genome sequences, the IGS sequences, and all CDS sequences using MP, BI, and ML.
<table>
<thead>
<tr>
<th>AA</th>
<th>Codon</th>
<th>Acy</th>
<th>Afa</th>
<th>Asp</th>
<th>Ama</th>
<th>Aki</th>
<th>Atr</th>
<th>Acy</th>
<th>Afa</th>
<th>Asp</th>
<th>Ama</th>
<th>Aki</th>
<th>Atr</th>
<th>RSCU</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phe</td>
<td>UUU</td>
<td>826</td>
<td>825</td>
<td>827</td>
<td>822</td>
<td>809</td>
<td>827</td>
<td>1.33</td>
<td>1.33</td>
<td>1.33</td>
<td>1.33</td>
<td>1.33</td>
<td>1.33</td>
<td>1.33</td>
</tr>
<tr>
<td></td>
<td>UUC</td>
<td>420</td>
<td>415</td>
<td>418</td>
<td>414</td>
<td>421</td>
<td>419</td>
<td>0.67</td>
<td>0.67</td>
<td>0.67</td>
<td>0.67</td>
<td>0.67</td>
<td>0.67</td>
<td>0.67</td>
</tr>
<tr>
<td></td>
<td>UUA</td>
<td>758</td>
<td>758</td>
<td>756</td>
<td>751</td>
<td>744</td>
<td>746</td>
<td>2.05</td>
<td>2.05</td>
<td>2.05</td>
<td>2.05</td>
<td>2.05</td>
<td>2.05</td>
<td>2.04</td>
</tr>
<tr>
<td></td>
<td>UUG</td>
<td>447</td>
<td>450</td>
<td>447</td>
<td>441</td>
<td>446</td>
<td>444</td>
<td>1.21</td>
<td>1.22</td>
<td>1.21</td>
<td>1.21</td>
<td>1.21</td>
<td>1.21</td>
<td>1.21</td>
</tr>
<tr>
<td></td>
<td>CUU</td>
<td>455</td>
<td>453</td>
<td>453</td>
<td>448</td>
<td>448</td>
<td>452</td>
<td>1.23</td>
<td>1.23</td>
<td>1.23</td>
<td>1.23</td>
<td>1.23</td>
<td>1.23</td>
<td>1.24</td>
</tr>
<tr>
<td></td>
<td>CUC</td>
<td>138</td>
<td>137</td>
<td>138</td>
<td>139</td>
<td>136</td>
<td>138</td>
<td>0.37</td>
<td>0.37</td>
<td>0.37</td>
<td>0.37</td>
<td>0.37</td>
<td>0.37</td>
<td>0.38</td>
</tr>
<tr>
<td></td>
<td>CUA</td>
<td>295</td>
<td>295</td>
<td>298</td>
<td>294</td>
<td>292</td>
<td>288</td>
<td>0.8</td>
<td>0.8</td>
<td>0.8</td>
<td>0.8</td>
<td>0.8</td>
<td>0.8</td>
<td>0.79</td>
</tr>
<tr>
<td></td>
<td>CUG</td>
<td>122</td>
<td>121</td>
<td>124</td>
<td>122</td>
<td>119</td>
<td>125</td>
<td>0.33</td>
<td>0.33</td>
<td>0.33</td>
<td>0.33</td>
<td>0.33</td>
<td>0.33</td>
<td>0.34</td>
</tr>
<tr>
<td></td>
<td>AUU</td>
<td>943</td>
<td>946</td>
<td>945</td>
<td>955</td>
<td>938</td>
<td>931</td>
<td>1.49</td>
<td>1.49</td>
<td>1.49</td>
<td>1.5</td>
<td>1.48</td>
<td>1.48</td>
<td>1.48</td>
</tr>
<tr>
<td></td>
<td>AUC</td>
<td>760</td>
<td>758</td>
<td>760</td>
<td>756</td>
<td>751</td>
<td>748</td>
<td>1.32</td>
<td>1.32</td>
<td>1.32</td>
<td>1.32</td>
<td>1.32</td>
<td>1.32</td>
<td>1.32</td>
</tr>
<tr>
<td></td>
<td>AUA</td>
<td>600</td>
<td>600</td>
<td>600</td>
<td>600</td>
<td>600</td>
<td>600</td>
<td>1.32</td>
<td>1.32</td>
<td>1.32</td>
<td>1.32</td>
<td>1.32</td>
<td>1.32</td>
<td>1.32</td>
</tr>
<tr>
<td></td>
<td>AAG</td>
<td>295</td>
<td>295</td>
<td>298</td>
<td>294</td>
<td>292</td>
<td>288</td>
<td>0.8</td>
<td>0.8</td>
<td>0.8</td>
<td>0.8</td>
<td>0.8</td>
<td>0.8</td>
<td>0.79</td>
</tr>
<tr>
<td></td>
<td>AGG</td>
<td>339</td>
<td>342</td>
<td>342</td>
<td>342</td>
<td>331</td>
<td>335</td>
<td>1.26</td>
<td>1.26</td>
<td>1.26</td>
<td>1.27</td>
<td>1.27</td>
<td>1.27</td>
<td>1.24</td>
</tr>
<tr>
<td></td>
<td>AGC</td>
<td>91</td>
<td>88</td>
<td>88</td>
<td>86</td>
<td>87</td>
<td>92</td>
<td>0.34</td>
<td>0.33</td>
<td>0.33</td>
<td>0.32</td>
<td>0.32</td>
<td>0.32</td>
<td>0.34</td>
</tr>
</tbody>
</table>
methods all showed the same topologies with high support (Figures S2 and S3). The results strongly supported that subgenus Cyathophora is a monophyletic group, comprising A. cyathophorum, A. cyathophorum var. farreri, A. spicatum, and A. mairei in this study with 100% bootstrap value; subgenus Cyathophora does not contain A. kingdonii and A. trifurcatum, and the phylogenetic tree indicates that A. cyathophorum var. farreri is a direct sister to A. spicatum, which is in accordance with the results of previous molecular research studies [4, 6]. The sister relationship of A. cyathophorum var. farreri and A. spicatum strongly suggests that A. spicatum is closely related to subgenus Cyathophora though it is a special species with the significant abnormal spicate inflorescence compared to other species with capitate or umbellate inflorescence. Furthermore, Allium kingdonii was the closest relative of Allium paradoxum and Allium ursinum.

### 4. Discussion

#### 4.1. Variations among the Six Allium Species

In this research, we assembled the complete cp genome of the six species in Allium. They were very conservative in genome structure and size; it showed a typical circular DNA structure and similar cp genome sequence length, ranging from 152,913 bp in A. mairei to 154,174 bp in A. cyathophorum. The six species had the identical numbers of protein-coding, tRNA, and rRNA genes. There were some expansion or contraction of IRs among these species (Figure 6); the expansion and contraction of IR regions are related to the divergences in chloroplast genome size [66]. To some extent, it is contributed to the cp genome variation and evolution. Other than this, variations in the IR/SC boundaries in the six cp genomes lead to the distinction in the whole length of sequence [61]. Previous research studies showed that SSRs have been widely known as important resources of molecular markers and have been broadly applied in phylogenetic and biogeographic studies [67, 68]. We surveyed and analyzed the quantities and distributions of SSRs with the six species in Allium, the largest number of SSR type was mono-nucleotide repeats, and the SSRs in the LSC area are much higher than those in the SSC and IR areas (Figure 4), showing that SSRs have an unevenly distribution in cp genome [50]. Additionally, we also explored seven common genes (infA, rps16, rps15, ndhF, trnG-UCC, trnC-GCA, and trnK-UUU) with nucleotide diversity more than 0.02 in the six cp genome sequences of Allium; among them, trnK-UUU, trnG-UCC, ndhF, and rps15 have been previously known as hypervariable regions in Allium [17], and we consider that these SSRs and genes with greater nucleotide diversity can be used as helpful DNA barcodes to identify the species in Allium.

#### 4.2. Phylogenetic Relationships

The results of phylogenetic analysis clearly show that Allium subgenus Cyathophora is a monophyletic group, and comprise four species (A. cyathophorum, A. farreri, A. spicatum and A. mairei), A. cyathophorum var. farreri has been upgraded to the level of the species as A. farreri in a recent study [69]. Besides A. farreri is a direct sister to A. spicatum with 100% strong bootstrap value, while the previous study showed low bootstrap value by the combined plastid dataset (trnLF + rpl32-trnL) [4]. Currently, most phylogenetic relationships are obtained with chloroplast fragments, while single ITS, chloroplast fragment, or chloroplast combined fragment does not have a better effect in phylogenetic analysis compared to the whole cp genome. We convinced that the complete chloroplast genomes have more advantages to solve the phylogenetic issues about the subgenus Cyathophora. In previous studies, many phylogenetic problems in many plants have been successfully resolved by using complete cp genome sequences [18, 19, 70]; the lately published article about Allium also well resolved the phylogenetic relationship [17, 71]. Although the morphological characteristics of the A. farreri and A. spicatum are obviously different, in which A. spicatum has distinctive spicate inflorescence compared to A. farreri with umbel hemispheric inflorescence, our results undoubtedly showed A. farreri is a direct sister to A. spicatum, which is in accordance with Li et al. [6]. According to previous study, different inflorescence may imply that the umbel inflorescence was replaced by spicate inflorescence to adapt the harsh environment [6]. The phylogenetic tree revealed A. cyathophorum had a closer relationship with A. spicatum and A. farreri compared to A. mairei. Furthermore, the members of subgenus Cyathophora do not contain A. kingdonii and A. trifurcatum; A. kingdonii was the closest relative of A. paradoxum and A. ursinum, which is consistency with previous studies [4, 6]. Certainly, our study persuasively constructed reliable phylogeny relationship of subgenus Cyathophora by using the complete cp genome data.

---

Table 4: Continued.

<table>
<thead>
<tr>
<th>AA</th>
<th>Codon</th>
<th>Acy</th>
<th>Afa</th>
<th>Asp</th>
<th>Ama</th>
<th>Aki</th>
<th>Atr</th>
<th>Acy</th>
<th>Afa</th>
<th>Asp</th>
<th>Ama</th>
<th>Aki</th>
<th>Atr</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arg</td>
<td>AGA</td>
<td>390</td>
<td>392</td>
<td>395</td>
<td>388</td>
<td>398</td>
<td>400</td>
<td>1.91</td>
<td>1.91</td>
<td>1.93</td>
<td>1.93</td>
<td>1.93</td>
<td>1.94</td>
</tr>
<tr>
<td></td>
<td>AGG</td>
<td>125</td>
<td>121</td>
<td>119</td>
<td>119</td>
<td>124</td>
<td>119</td>
<td>0.61</td>
<td>0.59</td>
<td>0.58</td>
<td>0.59</td>
<td>0.6</td>
<td>0.58</td>
</tr>
<tr>
<td></td>
<td>GGU</td>
<td>489</td>
<td>487</td>
<td>487</td>
<td>493</td>
<td>483</td>
<td>489</td>
<td>1.33</td>
<td>1.33</td>
<td>1.33</td>
<td>1.34</td>
<td>1.31</td>
<td>1.33</td>
</tr>
<tr>
<td>Gly</td>
<td>GGC</td>
<td>133</td>
<td>130</td>
<td>131</td>
<td>131</td>
<td>137</td>
<td>132</td>
<td>0.36</td>
<td>0.36</td>
<td>0.36</td>
<td>0.36</td>
<td>0.37</td>
<td>0.36</td>
</tr>
<tr>
<td></td>
<td>GGA</td>
<td>620</td>
<td>621</td>
<td>625</td>
<td>622</td>
<td>623</td>
<td>616</td>
<td>1.69</td>
<td>1.7</td>
<td>1.7</td>
<td>1.69</td>
<td>1.69</td>
<td>1.67</td>
</tr>
<tr>
<td></td>
<td>GGG</td>
<td>228</td>
<td>226</td>
<td>226</td>
<td>225</td>
<td>229</td>
<td>236</td>
<td>0.62</td>
<td>0.62</td>
<td>0.62</td>
<td>0.61</td>
<td>0.62</td>
<td>0.64</td>
</tr>
</tbody>
</table>

RSCU represents relative synonymous codon usage. RSCU more than one is highlighted in bold. Acy, Afa, Asp, Ama, Aki, and Atr stand for A. cyathophorum, A. cyathophorum var. farreri, A. spicatum, A. mairei, A. kingdonii, and A. trifurcatum, respectively.
4.3. Selection Events in Protein Coding Genes. DNA base mutations can be divided into two categories based on their effects to the encoded amino acids: synonymous mutations (Ks) and nonsynonymous mutations (Ka). Synonymous mutations do not result in amino acid changes, the frequency of which is represented by Ks; nonsynonymous mutations resulted in a change of amino acid, the frequency of which is indicated by Ka [72]. The ratio (Ka/Ks) is an important indicator to reveal evolutionary rate and natural selection pressure [73]. Interestingly, the synonymous

![Sequence alignment of six Allium chloroplast genomes (A. cyathophorum as the reference). The y-axis represents the percent similarity between 50% and 100%. Different colors represent different genetic regions.](image-url)
Figure 6: Comparison of the boundaries of the LSC, SSC, and IR areas of the whole chloroplast genomes of the six species.

Figure 7: The nucleotide diversity of the shared 112 genes of the six species in chloroplast genomes.
nucleotide substitutions occurred at a higher frequency than nonsynonymous substitutions, and thus Ka/Ks ratios are constantly <1 in most genes [9, 74], and our study is similar with this. Between different regions and genes, the Ka/Ks ratios were usually specific (Figure 8). Most conserved genes (56 of 65 genes) had an average Ka/Ks value ranging from 0 to 0.49 for the fifteen comparison groups, indicating that most genes were under purifying selection. On the contrary, the average Ka/Ks values of the ycf1 and ycf2 genes were >1 in the fifteen comparison groups, revealing that some selective pressure may execute on them in six Allium species. Previous studies have shown that ycf1 and ycf2 genes were two large open reading frames; they were important to tobacco, and the gene knockout experiments showed that ycf1 and ycf2 played important role in a healthy cell [75]. Hu et al. [76] suggested that plants have a variety of adaptation
were considered as the outgroups. Purple shows the branch of the subgenus *Asparagus* branches stand for bootstrap values in the ML, MP, and posterior probabilities in the BI trees. Six species of *Allium* using three different methods: maximum likelihood (ML), maximum parsimony (MP), and Bayesian inference (BI). Numbers on the branches stand for bootstrap values in the ML, MP, and posterior probabilities in the BI trees. Six species of *Lilium* and two species of *Asparagus* were considered as the out groups. Purple shows the branch of the subgenus *Cyathophora*.

![Figure 9: Phylogenetic relationship of subgenus Cyathophora with relational species. The whole chloroplast genomes dataset was analyzed using three different methods: maximum likelihood (ML), maximum parsimony (MP), and Bayesian inference (BI). Numbers on the branches stand for bootstrap values in the ML, MP, and posterior probabilities in the BI trees. Six species of Lilium and two species of Asparagus were considered as the out groups. Purple shows the branch of the subgenus Cyathophora.](image)

strategies in response to unforeseen environmental conditions. Recent studies about *Allium* species also suggested that the selective pressure in chloroplast genomes play an important role in *Allium* species adaptation and evolution [17, 71]. In our field investigations, the species of subgenus *Cyathophora* grows in slopes or grasslands with altitude ranging from 2700 m to 4800 m. The elevated Ka/Ks ratios observed about some genes in the six *Allium* species may suggest that it is relate to their specific living environment. What is more, there were four genes (*matK, rpoB, petA*, and *rpoA*) with Ka/Ks > 1 in at one or more pairwise comparisons (Figure 8, Table S3), and among these genes, *rpoA* was also undergone positive selection in species of Annonaceae [77]. Previous study demonstrated that *rpoA* encodes the α subunit of plastid RNA polymerase (PEP), which is in charge of the expression of most photosynthesis-related genes [78]. It is generally believed that low temperature and strong ultraviolet radiation are not conducive to effective photosynthesis of plants; therefore, plants that survive and reproduce at high altitudes need a special photosynthetic protection strategy [79, 80]. In this study, the population of subgenus *Cyathophora* is mainly distributed in the Qinghai-Tibet Plateau and its adjacent high-altitude regions [2]. Therefore, we speculated that the positive selection of these genes may be related to the difference between their optimal growth environment.

5. Conclusions

Here, we sequenced, assembled, and annotated six chloroplast genomes of *Allium* with high-throughput sequencing technology. The gene contents and orders of the cp genomes were extremely conservative, and their cp genomes are also quadripartite structure. Repeated sequence and SSRs are helpful sources for developing new molecular markers. Codon usage analyses detected that some amino acids of the six species showed distinct codon usage preferences, and we should comprehend codon usage bias to learn evolution process. We also discovered seven highly variable common genes which can be used to develop useful markers for phylogenetic analysis and distinguish species in *Allium*. The Ka/Ks analysis indicated that some selective pressure may exert on several genes in the chloroplast genomes of six *Allium* L. species. The maximum likelihood (ML), BI, and MP phylogenetic results clearly showed that subgenus *Cyathophora* comprised the four assembled species: *A. cyathophorum*, *A. cyathophorum* var. *farreri*, *A. spicatum*, and *A. mairei*, and *A. cyathophorum* var. *farreri* has a closer relationship with *A. spicatum*. This study will not only provide insights into the cp genome characteristics of species in subgenus *Cyathophora* but also supply useful genetic resources for phylogenetic analysis of genus *Allium*.

Data Availability

The complete chloroplast genome sequences of *A. cyathophorum*, *A. cyathophorum* var. *farreri*, *A. spicatum*, *A. mairei*, *Allium trifulcatum*, and *Allium kingdonii* are saved in the GenBank of NCBI, and the accession numbers are MK820611, MK931245, MK931246, MK820615, MK931247, and MK294559, respectively.

Conflicts of Interest

The authors declare no conflicts of interest.
Acknowledgments

The authors appreciate the open cp genome data in NCBI and acknowledge Min Jie Li for the help in collection of materials. The authors also thank Xian Lin Guo, Fu Ming Xie, Dan Mei Su, Hao Li, and Sheng Bin Jia for their help in the use of software. This work was supported by the National Natural Science Foundation of China (grant nos. 31872647 and 31570198), National Specimen Information Infrastructure, Educational Specimen Sub-Platform (http://mnh.scu.edu.cn/), and Science and Technology Basic Work (grant no. 2013FY112100).

Supplementary Materials

Supplementary 1. Table S1: all used accession numbers of cp genome sequences from GenBank in this article.

Supplementary 2. Table S2: the nucleotide variability (Pi) of 112 shared genes.

Supplementary 3. Table S3: the Ka/Ks ratio of 65 shared protein-coding genes in all six chloroplast genomes.

Supplementary 4. Figure S1: the proportion of the different lengths of the four dispersed repeats.

Supplementary 5. Figure S2: phylogenetic tree of the subgenus Cyathophora built by the CDS sequences using maximum likelihood (ML), maximum parsimony (MP), and Bayesian inference (BI).

Supplementary 6. Figure S3: phylogenetic tree of the subgenus Cyathophora built by the IGS sequences using maximum likelihood (ML), maximum parsimony (MP), and Bayesian inference (BI).

References


[34] X. Nie, S. Lv, Y. Zhang et al., "Complete chloroplast genome sequence of a major invasive species, croton weed (Ageratina adenophora)," *PloS One*, vol. 7, no. 5, Article ID e36869, 2012.


