Article

Complete Chloroplast Genome Sequences of Clematis: IR Expansion and Relative Rates of Synonymous Substitutions

Kyoung Su Choi¹, Keum Seon Jeong², Young-Ho Ha¹, Kyung Choi¹*

¹Division of Forest Biodiversity, Korea National Arboretum of the Korea forest Service, Pochen 11186, Republic of Korea
²Division of Education Service Team, Baekdudaegan National Arboretum, Bonghwa-gun 36209, Republic of Korea

* Correspondence: kchoi69@korea.kr; Tel.: +82-31-540-1090

Abstract: Genus Clematis is one of the largest within Ranunculaceae. Here we report the chloroplast genome of two Clematis species, C. brachyura and C. trichotoma endemic to Korea. The chloroplast genome lengths of C. brachyura and C. trichotoma are 159,532 bp and 159,170 bp, respectively. Gene contents in the complete chloroplast genomes of these two Clematis species are identical to that of most Ranunculaceae and other angiosperms. However, our data results demonstrated that genus Clematis has inversion and rearrangement events concerning gene rps4 gene, rps16 to trnH region, and trnL to ndhC region, and IR regions expansion. Comparison of IR regions among Ranunculaceae species revealed that Clematis species contained six protein coding genes (infA, rps8, rpl14, rpl16, rps3, and rpl22) usually found in the long single copy (LSC) region of other species. Phylogenetic analysis demonstrated that genus Clematis is closely related to genus Ranunculus. Differences in repeat structure, substitution rates, and IR expansion in genera Clematis and Ranunculus, explained their relationship. Clematis species showed slightly higher tandem repeats content than Ranunculus species. The six protein-coding genes showed lower synonymous substitution rates in the IR of Clematis species than in the LSC of Ranunculus species. Overall, the chloroplast genomes and results presented here provide important information on the evolution of Ranunculaceae.

Keywords: Clematis; Chloroplast genome; Rearrangement; Inversion; IR expansion; Synonymous substitution rate

1. Introduction

Family Ranunculaceae comprises 59-62 genera and approximately 2,500 species that have two chromosome types: R (Ranunculus) type with large chromosomes and T (Thalictrum) type with small chromosomes [1, 2]. Ranunculaceae were classified based on morphological characters until chromosomal characters became the most important criterion for classification [1, 3]. Genus Clematis, a member of Ranunculaceae, consists of 250-350 species, including some medicinal plants [4], which are distributed in all continents except Antarctica. Clematis brachyura and C. trichotoma are particularly valuable plants endemic to Korea [5].

The chloroplast (cp) genome consists of a large single copy (LSC), a small single copy (SSC), and two inverted repeat (IR) regions. In angiosperms, cp genomes of generally range from 120 to 160 kb in length and contain 110 to 130 genes. Gene contents and gene order are highly conserved in most angiosperms [6, 7] although many studies have reported changes in cp genome including gene loss...
inversion or deletions [10, 11], and expansion or contraction of the IR region [12-16]. The IR regions range from 15 to 30 kb in length and contain four ribosomal RNA (rRNA) genes, five transfer RNA (tRNA) genes and four coding genes. In recent studies, the IR regions have been used as important markers given their variability [17, 18], especially in plants, presenting large-scale expansions [14, 19], or loss of IR region [20, 21]. Previous studies suggested that mononucleotide, dispersed (>16 bp), tandem repeats and substitution rates are important contributors to cp genome size and structural evolution [13, 15, 22, 23].

Most recently, sequencing of complete cp genomes in Ranunculaceae revealed the transfer of rpl32 gene to the nucleus [24], phylogenetic and genomic structure analyses in Aconitum spp. [25, 26] and the evolutionary history of Ranunculaceae [27]. Many studies focused on Aconitum species because this is a traditional herbal medicine in Asia [26, 28, 29].

The present study reports the complete cp genome of two species (C. brachyura and C. trichotoma) within genus Clematis and its major objectives were to 1) characterize and compare the structure, gene organization, and phylogenetic relationships within Ranunculaceae, 2) analyze the inversions, rearrangements, and IR expansions within Clematis.

2. Materials and Methods

2.1. Chloroplast genome sequencing, assembly and annotation

Total genomic DNA from Clematis brachyura and C. trichotoma were isolated from fresh leaf tissues of a single plant using the DNeasy Plant Mini Kit (Qiagen Inc., Valencia, CA, USA) and sequenced using the Illumina Miseq platform (Illumina Inc., San Diego, Ca). A total 8,572,072 bp and 7,505,088 bp pair-end sequence read of 300 bp was generated from the sequencing library, respectively. All contigs were de novo assembled into contigs with Velvet v. 1.2.03 [30]. Thereafter, complete chloroplast sequences were assembled with reference genome of Clematis terniflora (NC_028000) and assembled into a complete genome by overlaps using Geneious v 10.2.2 [31].

Complete cp genome sequences were annotated using DOGMA [32] and all identified tRNA genes were annotated using tRNAscan-SE [33]. Circular genome maps were drawn in OGDRAW [34]. To compare the structure and genes present in Ranunculaceae cp genomes, sequences from the different plants were aligned using MAFFT [35] and Geneious v. 10.2.2.

2.2. Phylogenetic analysis

A total of 71 coding genes of 15 Ranunculaceae were compiled in to a single file comprising 55,759 bp and aligned with MAFFT. Fourteen Ranunculaceae (including C. brachyura and C. trichotoma) were selected as in-groups and Nuphar (DQ354691) was included as outgroup (Table 1). Maximum likelihood (ML) analyses were performed using RAxML v7.4.2 using 1000 bootstrap replicates and the GRT+I+G model [38].

2.3. Repeat structure and substitution rate analysis

Tandem repeats (forward and reverse) were detected using REPuter [36]. The minimum repeats sizes was set at 30 bp and at a sequence indent greater than 90%. The simple sequence repeats (SSRs) were detected using Phobos v. 3.3.12 (http://www.ruhr-uni-bochum.de/ecoevol/cm/com_phobos.htm). The total number of repeats was compared across species considering mononucleotide (>8bp). One copy of the large IR was removed from each genome before repeat analyses.
Protein-coding genes (rpl14, rpl16, rpl2, rpl22, rps19, rps3 and rps8) of IR expansion region were extracted from each sequenced genome and previously published genomes Clematis (NC_02800 and KM652489), Ranunculus (DQ359689, KX557270 and KX639503) and Aconitum (NC_029829) genomes, were included as outgroup. Extracted sequences were aligned using MAFFT. Synonymous substitution (Ks) rates were analyzed in DnaSP [37].

3. Results

3.1. Chloroplast genome and gene contents in genus Clematis 3.1.1. Subsection

Total cp genome size was 159,532 bp in C. brachyura (KM104710) and 159,170 bp in C. trichotoma (KM104711) and ranged within these values (Figure 1, Table 1), which is consistent with that from C. fusca var. coreana (KM652489) and C. terniflora (NC_02800). All of the Clematis genomes displayed a typical quadripartite structure, comprising two IRs (17,997-18,105 bp) separated by the LSC (79,339-79,478 bp) and SSC (17,997-18,105 bp) regions. The GC content of the cp genomes obtained here is 38%, which is consistent with all Clematis cp genomes. There are 112 unique genes, including 79 protein-coding genes, 29 tRNA genes and four rRNA genes.

Compared with Ranunculaceae, total genome sizes ranged from 146.9 to 160.8 kb. Species within Clematis have an expanded IR with approximately 5kb, comprising five or six coding genes (Table 1).
Figure 1. Chloroplast genome of *Clematis brachyura* and *C. trichotoma*. Genes inside the circle are transcribed clockwise, while genes outside the circle are transcribed counter-clockwise. Dark gray and light gray inner circles corresponds to GC and AT contents, respectively.

3.2. Phylogenetic relationships and gene loss in Ranunculaceae

The phylogenetic analysis conducted for the 71 protein-coding genes of 15 Ranunculaceae (Figure 2) revealed several monophyletic groups with strong support (100% bootstrap values). Genera *Clematis* and *Ranunculus* form two well-supported monophyletic sister groups (100% bootstrap values), which is consistent with previous finding for family Ranunculaceae [2, 39].

Previous studies revealed that gene *infA* was lost or pseudogenation in angiosperm [18, 20, 40]. In the present study, *Ranunculus*, *Thalictrum*, and *Gymnaconitum* show an internal stop codon in *infA*. Park et al. [24] reported that *rpl32* was lost in *Thalictrum*. According to the present study, *Aconitum*, *Gymnaconitum* and *Actaea* also lost *rpl32* (Figure 2).
Table 1. Ranunculaceae chloroplast genomes compared in this study.

<table>
<thead>
<tr>
<th>Species</th>
<th>Total Size (bp)</th>
<th>LSC size (bp)</th>
<th>SSC size (bp)</th>
<th>IR size (bp)</th>
<th>Genes in IR region (coding genes/tRNA/rRNA)</th>
<th>GC content in IR region</th>
<th>Accession number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clematis brachyura (CB)</td>
<td>159,532</td>
<td>79,341</td>
<td>18,105</td>
<td>31,043</td>
<td>13/7/4</td>
<td>42.1%</td>
<td>KM104710</td>
</tr>
<tr>
<td>Clematis trichotoma (CT)</td>
<td>159,170</td>
<td>79,339</td>
<td>17,997</td>
<td>30,917</td>
<td>12/7/4</td>
<td>42.0%</td>
<td>KM104711</td>
</tr>
<tr>
<td>Clematis fusca var. coreana (CF)</td>
<td>159,609</td>
<td>79,478</td>
<td>18,044</td>
<td>31,048</td>
<td>13/7/4</td>
<td>42.0%</td>
<td>KM652489</td>
</tr>
<tr>
<td>Clematis terniflora (CTE)</td>
<td>159,528</td>
<td>79,328</td>
<td>18,092</td>
<td>31,054</td>
<td>13/7/4</td>
<td>42.0%</td>
<td>NC_028000</td>
</tr>
<tr>
<td>Ranunculus macranthus (RM)</td>
<td>155,129</td>
<td>84,637</td>
<td>18,910</td>
<td>25,791</td>
<td>6/7/4</td>
<td>43.5%</td>
<td>DQ359689</td>
</tr>
<tr>
<td>Ranunculus occidentalis (RO)</td>
<td>154,474</td>
<td>83,532</td>
<td>21,282</td>
<td>24,830</td>
<td>6/7/4</td>
<td>43.6%</td>
<td>KX557270</td>
</tr>
<tr>
<td>Ranunculus australis-oreganus (RA)</td>
<td>154,493</td>
<td>83,582</td>
<td>21,249</td>
<td>24,831</td>
<td>6/7/4</td>
<td>43.6%</td>
<td>KX639503</td>
</tr>
<tr>
<td>Aconitum chiisanense (AC)</td>
<td>155,934</td>
<td>86,559</td>
<td>17,054</td>
<td>26,161</td>
<td>6/7/4</td>
<td>43.0%</td>
<td>NC_029829</td>
</tr>
<tr>
<td>Gymnacodium gymnantrum</td>
<td>157,327</td>
<td>88,107</td>
<td>16,940</td>
<td>26,140</td>
<td>6/7/4</td>
<td>41.1%</td>
<td>NC_033341</td>
</tr>
<tr>
<td>Thalictrum coreanum</td>
<td>155,088</td>
<td>84,733</td>
<td>17,549</td>
<td>26,403</td>
<td>7/7/4</td>
<td>43.3%</td>
<td>NC_026103</td>
</tr>
<tr>
<td>Megaleranthis saniculifolia</td>
<td>159,924</td>
<td>88,326</td>
<td>18,382</td>
<td>26,608</td>
<td>7/7/4</td>
<td>43.1%</td>
<td>FJ597983</td>
</tr>
<tr>
<td>Tollius chinensis</td>
<td>160,191</td>
<td>88,522</td>
<td>18,405</td>
<td>26,632</td>
<td>7/7/4</td>
<td>43.1%</td>
<td>NC_031849</td>
</tr>
<tr>
<td>Coptis chinensis</td>
<td>155,484</td>
<td>84,567</td>
<td>17,393</td>
<td>26,762</td>
<td>5/7/4</td>
<td>43.0%</td>
<td>NC_036485</td>
</tr>
<tr>
<td>Actaea racemosae</td>
<td>146,906</td>
<td>92,594</td>
<td>18,674</td>
<td>17,819</td>
<td>5/7/4</td>
<td>45.0%</td>
<td>NC_034704</td>
</tr>
<tr>
<td>Nuphar advena</td>
<td>160,866</td>
<td>90,379</td>
<td>18,817</td>
<td>25,835</td>
<td>7/7/4</td>
<td>43.3%</td>
<td>DQ354691</td>
</tr>
</tbody>
</table>
Figure 2. Phylogenetic tree reconstruction of 15 taxa using maximum likelihood, based on concatenated sequences of 71 protein-coding genes and gene loss.
3.3. Inversion, rearrangement, and IR expansion in genus Clematis

Chloroplast genomes of Ranunculaceae are highly conserved. However, rearrangements, inversions and IR expansions have been detected in cp genomes of Clematis species. A small inversion (Figure 3A) and a large inversion (Figure 3B) were detected in the LSC region between genes rps4 gene and rps16 and to the trnH region, respectively (Figure 3). Rearrangements such as shifting trnL to ndhC region (Figure 3C) and shifting rpl22 to the infA region (Figure 4) were also detected. Shifting rpl22 to the infA region is considered an IR expansion in genus Clematis.

Figure 3. Inversions and rearrangements in genus Clematis. A: rearrangement of gene rps4. B: inversion of rps16 to the trnH region C: rearrangement of trnL to the ndhC region.

Figure 4. The inverted repeat (IR) region expansion in genus Clematis. A: IR region of Clematis brachyuran B: IR region of C. trichotoma, C: IR region of genus Ranunculus species. SSC, small single copy, LSC, long single copy.
3.4. Repeat content

Tandem repeats and repetitive sequences were compared between Clematis and Ranunculus species (Table 2). Repetitive sequences were longer in Clematis species (1,369 to 1,402 bp) than in Ranunculus species (1,073 to 1,487 bp) and tandem repeats were also longer in Clematis species (385 to 461 bp) than Ranunculus species (212 to 265 bp). Among the eight cp genomes compared (including Aconitum), four Clematis species contained the longest tandem repeats (more than 73 bp).

Table 2. Summary of repetitive sequences and tandem repeats contents Clematis and Ranunculus species (*outgroup).

<table>
<thead>
<tr>
<th>Species</th>
<th>Size (bp)</th>
<th>Repetitive sequence (bp)</th>
<th>Tandem repeats (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clematis brachyura (CB)</td>
<td>128,500</td>
<td>1,381</td>
<td>385</td>
</tr>
<tr>
<td>Clematis trichotoma (CT)</td>
<td>128,248</td>
<td>1,369</td>
<td>461</td>
</tr>
<tr>
<td>Clematis fusca var. coreana (CF)</td>
<td>128,561</td>
<td>1,402</td>
<td>427</td>
</tr>
<tr>
<td>Clematis terniflora (CTE)</td>
<td>128,474</td>
<td>1,390</td>
<td>385</td>
</tr>
<tr>
<td>Ranunculus macranthus (RM)</td>
<td>129,338</td>
<td>1,487</td>
<td>265</td>
</tr>
<tr>
<td>Ranunculus occidentalis (RO)</td>
<td>129,644</td>
<td>1,073</td>
<td>212</td>
</tr>
<tr>
<td>Ranunculus austro-oreganius (RA)</td>
<td>129,662</td>
<td>1,136</td>
<td>212</td>
</tr>
<tr>
<td>Aconitum chiisanense (AC)*</td>
<td>129,773</td>
<td>1,155</td>
<td>313</td>
</tr>
</tbody>
</table>

3.5. Synonymous substitution rates

Synonymous substitution rates were compared between the IR and LSC regions of four Clematis and three Ranunculus species. The values of Ks were markedly higher for LSC than for IR genes in both genera (Figure 5A, B). Ranunculaceae have six protein-coding genes (infA, rps8, rpl14, rpl16, rps3, and rpl22) in the LSC region. However, in Clematis genomes, these six protein-coding genes shifted from the LSC region to IR regions (Figure 4, 5). For five of these genes (except infA), Ks values were substantially lower in IR than LSC region (Figure 5C). Gene infA was lost in genus Ranunculus. Genes (rpl14, rpl16, and rps19) exhibited extremely lower values in the IR of Clematis species than in the LSC of Ranunculus species.
Figure 5. Synonymous substitutions (Ks) in genera *Clematis* and *Ranunculus*. A: Ks in *Clematis* species. B: Ks in *Ranunculus* species. C: Ks for genes in the junction of long single copy (LSC) and inverted repeat (IR) regions, including genes relocated into or out of these regions.

4. Discussion

4.1 Structure of chloroplast genomes and comparative analyses

Previous studies showed that typical cp genomes of angiosperms have similar size, gene order and gene contents [41, 42]. Chloroplast genome sizes of Ranunculaceae generally range from generally 146-155 kb (except for *Trollius*), but *Clematis* species cp genomes are approximately 159 kb in size. Gene order within LSC region and IR regions also differs between *Clematis* and other Ranunculaceae genera. In particular, the IR regions expanded into the LSC region by shifting six protein-coding genes (*infA*, *rps8*, *rpl14*, *rpl16*, *rps3*, and *rpl22*) usually found in the LSC region. Similar patterns have been observed in *Tetracentron* and *Trochodendron* within family Trochodendraceae [43], where five protein-coding genes (*rps8*, *rpl14*, *rpl16*, *rps3*, and *rpl22*) are shifted to the IR regions. A previous study suggested poly A regions might play an important role in IR expansion [12]. In *Clematis* species, IR extension ends in the LSC region and the IR region is over 70% AT-rich with numerous possible poly A tracts that might have played a role in its expansion.

The cp genome of *Clematis* species has similar gene contents to that of other Ranunculaceae genera. However, *Clematis* species present rearrangements and inversions in the LSC region. Hoot and Palmer [44] reported that *Anemone* and related genera (*Clematis*, *Pulsatilla*, *Hepatica* and *Knowltonia*) underwent inversion and rearrangement events throughout cp genome evolution. Several studies suggested that tandem repeats or AT-rich regions occurred rearrangement and plastome size evolution in angiosperm such as *Silene* [23], *Trifolium* [45], and Geraniaceae [13]. For example, repetitive sequence and tandem repeats within the contraction/expansion IR region of some *Sileneae* species is dramatically higher than in other *Sileneae* species [23]. However, the genus *Clematis* species show tandem repeats content slightly higher than closely related taxa or outgroup (Table 2).
4.2 Gene shifts from LSC to IR reduced synonymous substitution rates

The synonymous substitution rate is lower in the IR than in the single copy (SC) regions of angiosperm cp genomes [46]. Zhu et al. [16] demonstrated that genes moved from the SC into IR regions in land plants demonstrated lower synonymous substitution rates consistent with that observed for original genes in IR regions.

In this study, we examined synonymous substitution rates of five LSC genes (rps8, rpl14, rpl16, rps3, and rpl22) shifted to IR in Clematis species and LSC located genes in Ranunculus species (Figure 5C). We found that the five shifted genes in Clematis species (LSC to IR genes) exhibited lower synonymous substitution rates than genes located in the LSC region of Ranunculus species. These results support that the duplicative nature of the IR reduces the synonymous substitution rate within this region, as stated in a previous study [16].

Acknowledgments: This work was supported by grants from Scientific Research (KNA1-1-13, 14-1) of the Korea National Arboretum.

Author Contributions: K Choi, K.S.Choi and K.S. Jeong conceived and designed the experiments; K.S.Choi and Y.H. Ha performed the experiments; K.S. Choi and Y.H. Ha analyzed the data; K.S.Choi wrote the paper.

Conflicts of Interest: The authors declare no conflict of interest.

References


24. Park, S.; Jansen, R.K.; Park, S. Complete plastome sequence of *Thalictrum cernuum* (Ranunculaceae) and transfer of the *rpl32* gene to the nucleus in the ancestor of the subfamily Thalictroideae. *BMC Plant Biol.* 2015, 15, 40.


