The Complete Chloroplast Genome Sequence of *Viburnum odoratissimum* and Phylogenetic Relationship with Other Close Species in the Adoxaceae Family

Le Thi Yen, Joonho Park

Department of Fine Chemistry, Seoul National University of Science and Technology, Seoul, South Korea

Email address: Jhpark21@seoultech.ac.kr (J. Park)

*Corresponding author

To cite this article:

Received: April 5, 2021; Accepted: April 27, 2021; Published: May 8, 2021

**Abstract:** The chloroplast genome structure and gene content are highly conserved among land plants, providing valuable information for the studies of taxonomy and plant evolution. *Viburnum odoratissimum* is a well-known evergreen shrub widely distributed in Asia. It possesses excellent medicinal properties used as traditional medicine for menstrual, stomach, and kidney cramps. In this study, the complete chloroplast genome (cpDNA) of *V. odoratissimum* is reported and compared with five close *Viburnum* species and an outgroup. The cpDNA of *V. odoratissimum* is 158,744 bp in length and contains 130 genes with 17 genes duplicated in the inverted repeat region. The gene content, gene organization and GC content in *V. odoratissimum* are highly similar to other *Viburnum* species. A total of 270 tandem repeats is found in these plastomes, most of which are distributed in intergenic space. Differences in the location of the IR/SC boundaries reflect expansions and contractions of IR regions in all species studied. Phylogenetic analysis based on complete chloroplast genomes and the combination of barcodes indicates a sister relationship between *V. odoratissimum* and *V. brachybotryum*. Furthermore, a comparative cpDNA analysis identifies three DNA regions (trnC-petN-psbM, trnH-psbA, ndhC-trnV) containing high divergence among seven studied species that could be used as potential phylogenetic markers in taxonomic studies.

**Keywords:** Adoxaceae, Barcodes, Phylogenetic Relationship, *Viburnum odoratissimum*

1. Introduction

The genus *Viburnum* comprise about 200 species of deciduous shrubs, evergreen and small trees, which are broadly distributed in subtropical and temperate Northern Hemisphere and spread across the mountain regions of South Asia and South America, Mexico, and Columbia [1, 2]. *Viburnum* together with *Sambucus* and *Adoxa* were members of Caprifoliaceae but was recently moved to the new group, Adoxaceae, according to phylogenetic analyses [3]. Most *Viburnum* species have become popular ornamental plants because of their eye-catching flowers with a light fragrance and berries. Moreover, many species in the genus *Viburnum* have utilized as the traditional folk in China, Russia and Ukraine for a number of diseases, such as menstruation, hypertension, flu, tuberculosis, renal infection, stomach ache, duodenal ulcers [4-6]. These species possessed a considerable number of secondary metabolites: monoterprenes, sesquiterprenes, diterpenes [6], diterpenoids, triterpenoids, iridoids [7, 8], chlorogenic [5], amyrin, lupeol [4], resulting in many biological properties, including anti-inflammatory, antibacterial, antioxidant, antitussive activities [5, 9, 10].

The phylogenetic relationship within the genus *Viburnum* has been extensively elucidated using not only morphological characters [11] but also nuclear DNA regions, such as the nuclear ribosomal internal transcribed spacer (ITS), the granule-bound starch synthase gene (GBSSI) [1, 12]. Recently, the chloroplast nucleotide sequences are highly supportive in deciphering the phylogenetic relationship among *Viburnum* genus using *trnK*, *matK*, *rbcL*, *psbA-trnH*, *rpl32-trnL* [1, 13, 14]. However, molecular phylogenetic studies based on several chloroplast markers remain a number of issues that can cause misleading evaluation of the
relationship [15]. Complete chloroplast genomes have been widely used in phylogeny reconstruction to overcome this problem because it provides valuable information on plant evolution, and a rich source of data to estimate of phylogenetic relationships [16].

In this study, we report the complete nucleotide sequence of *Viburnum odoratissimum*, along with a comparative analysis with other species in the genus *Viburnum*. The comparison with other published chloroplast genomes in related families is performed to expand understanding of the plastid genome diversity of *Viburnum* species. Furthermore, some new DNA barcodes containing high nucleotide divergence are identified. These hotspots could be considered as potential molecular markers for phylogenetic tree reconstruction within the *Viburnum* genus.

2. Materials and Methods

2.1. Sampling and Sequencing

The sample of *V. odoratissimum* was collected from National Institute of Biological Resources, Incheon, Korea (NIBRGR0000081148).

Approximately 5g of the leaves was used for isolation total genomic DNA following a modified CTAB method [17] with a minor modification. The quality of the extracted DNA was assessed by using spectrophotometry and electrophoresis on 1 % (w/v) agarose gel. A total of 10 µg purified genomic DNA was utilized to sequence the chloroplast genome using PacBio RS II system. The quality of the raw data was assessed to remove low-quality reads. The published complete chloroplast genome of *V. erosa* (MN641480.1) was downloaded from NCBI for a comparison.

2.2. Chloroplast Genome Assembly and Annotation

The filtered subreads were mapped to the referenced genome using BWA Aligner [18]. The matched subreads were selected for the de novo assembly with CANU version 1.8 [19]. All contigs were checked overlapped region using nucmer and mummerplot. Annotation and visualization for the assembled chloroplast genome were performed with an Annotation tool – GeSeq [20]. Finally, the circular gene map was drawn with OGDraw version 1.3.1 [21]. The complete chloroplast genome of *V. odoratissimum* was deposited in GenBank with accession number MN836381.

2.3. Sequence Analysis

Six chloroplast genome sequences, *Viburnum utile* (KX792264), *Viburnum betulifolium* (MG738665), *Viburnum japonicum* (MH036593), *Viburnum erosa* (MN218778), *Viburnum brachybotryum* (MNS24624) and *Tetradoxa omeiensis* (NC_034793), were obtained from NCBI for a comparison.

Tandem Repeat Finder version 4.09 [22] was used to search tandem repeats. Additionally, simple sequence repeats (SSRs) were detected by MISA [23] with the following settings for numbers of repetitions: 10 for monon-, 6 for di-, 5 for all tri-, tetra-, penta-, and hexanucleotide.

The seven complete chloroplast genomes were aligned and visualized with the online comparison tool mVISTA [24] using *V. odoratissimum* as a reference. To analyze nucleotide variability, seven studied chloroplast genome were aligned using ClustalX 1.81 [25] and then conducted a sliding window analysis using DnaSP version 6.10.03 [26].

3. Results and Discussion

3.1. Characteristics of Viburnum Odoratissimum

**Chloroplast Genome**

The complete cpDNA of *V. odoratissimum* is 158,744 bp in size, with a pair of inverted repeat regions (IRs) of 26,494 bp that separate a large single-copy (LSC) region of 87,348 bp from a small single-copy (SSC) region of 18,267 bp (Figure 1). The total GC content is 38.1 %, with the highest content in IR regions (43%), followed by an LSC (36.4%), and an SSC accounting for 32.1%. All the sequences of protein-coding genes and tRNA genes in the *V. odoratissimum* cp genome are encoded by 26,278 codons. Leucine is the most frequent amino acid with 10.5% (2768) of the codon and cysteine is the least frequent with 1.1% (294).

The *V. odoratissimum* cp genome encodes 129 genes, consisting of 84 protein-coding genes, 37 tRNA genes, and eight rRNA genes. Among these genes, 17 genes are duplicated in IR regions, six functional genes, seven tRNA genes, and four RNA genes. In total, there are 22 intron-containing genes, 19 of which contain one intron, and three of which contain two introns (clpP, rps12, and ycf3). The largest intron is largest in *trnK-UUU* which itself contains the *matK* gene.

3.2. Comparative Complete Chloroplast Genomic Analysis

3.2.1. Genome Structure and Content

The complete chloroplast genome of *V. odoratissimum* (MN836381) is compared to those of five others in *Viburnum*: *Viburnum utile* (KX792264), *Viburnum betulifolium* (MG738665), *Viburnum japonicum* (MH036493), *Viburnum erosa* (MN218778), *Viburnum brachybotryum* (MNS24624), and an outgroup, *Tetradoxa omeiensis* (NC_034793). Each chloroplast genome encodes for a total of 130 genes, including 85 protein-coding genes, 37 tRNA genes, and eight rRNA genes. There are 17 genes duplicated in the IR regions. A total of 22 genes contains introns, 19 of which contains one intron while three genes have two introns (clpP, rps12, and ycf3).

The cpDNA size of *V. odoratissimum* is the largest among seven studied genomes (158,744 bp), which is larger than the smallest cp genome of *V. brachybotryum* by 1,311 bp (Table 1). There are slight differences in length between IR (26,123-26,517 bp) or SSC (18,338-18,795 bp), and the main reason for the variation in genomic size is a difference in the length of the LSC (86,526-87,348 bp).
Figure 1. Gene map of *V. odoratissimum* chloroplast genome. Genes drawn inside and outside the circle are transcribed anti-clockwise and clockwise, respectively. Genes are differently colored by the functional groups which genes are affiliated with.

Table 1. Summary of complete chloroplast genomes of six *Viburnum* species and an outgroup.

<table>
<thead>
<tr>
<th>Features</th>
<th><em>Viburnum odoratissimum</em></th>
<th><em>Viburnum utile</em></th>
<th><em>Viburnum betulifolium</em></th>
<th><em>Viburnum japonicum</em></th>
<th><em>Viburnum erosum</em></th>
<th><em>Viburnum brachybotryum</em></th>
<th><em>Tetradoxa omeiensis</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>Genome size</td>
<td>158,744</td>
<td>157,620</td>
<td>158,023</td>
<td>158,614</td>
<td>158,624</td>
<td>157,433</td>
<td>157,502</td>
</tr>
<tr>
<td>LSC Length</td>
<td>87,348</td>
<td>86,576</td>
<td>86,761</td>
<td>87,060</td>
<td>87,060</td>
<td>86,526</td>
<td>86,526</td>
</tr>
<tr>
<td>SSC Length</td>
<td>18,408</td>
<td>18,726</td>
<td>18,338</td>
<td>18,523</td>
<td>18,530</td>
<td>18,615</td>
<td>18,682</td>
</tr>
<tr>
<td>IR length</td>
<td>26,494</td>
<td>26,159</td>
<td>26,462</td>
<td>26,516</td>
<td>26,517</td>
<td>26,133</td>
<td>26,147</td>
</tr>
<tr>
<td>Coding Size</td>
<td>78,834</td>
<td>75,411</td>
<td>77,061</td>
<td>78,852</td>
<td>77,124</td>
<td>47,316</td>
<td>75,525</td>
</tr>
<tr>
<td>GC content (%)</td>
<td>38.1</td>
<td>38.1</td>
<td>38.1</td>
<td>38.1</td>
<td>38.1</td>
<td>38.1</td>
<td>37.7</td>
</tr>
<tr>
<td>Total number of genes</td>
<td>130</td>
<td>130</td>
<td>130</td>
<td>130</td>
<td>130</td>
<td>130</td>
<td>130</td>
</tr>
<tr>
<td>Protein-coding genes</td>
<td>85</td>
<td>85</td>
<td>85</td>
<td>85</td>
<td>85</td>
<td>85</td>
<td>85</td>
</tr>
<tr>
<td>Duplicated genes</td>
<td>17</td>
<td>17</td>
<td>17</td>
<td>17</td>
<td>17</td>
<td>17</td>
<td>17</td>
</tr>
<tr>
<td>tRNA genes</td>
<td>37</td>
<td>37</td>
<td>37</td>
<td>37</td>
<td>37</td>
<td>37</td>
<td>37</td>
</tr>
<tr>
<td>rRNA genes</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>Genes with introns</td>
<td>22</td>
<td>22</td>
<td>22</td>
<td>22</td>
<td>22</td>
<td>22</td>
<td>22</td>
</tr>
<tr>
<td>Pseudogenes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ndhB, matK, rpoB, ndhJ, rpoA, rpl16, ycf2, ndhH, rps15, and ycf1</td>
<td></td>
</tr>
</tbody>
</table>
Table 2. Simple sequence repeat analysis in the seven studied species.

<table>
<thead>
<tr>
<th>SSR Type</th>
<th>Repeat unit</th>
<th>Viburnum odoratissimum</th>
<th>Viburnum utile</th>
<th>Viburnum betulifolium</th>
<th>Viburnum japonicum</th>
<th>Viburnum erosum</th>
<th>Viburnum brachybotryum</th>
<th>Tetradoxa omeiensis</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mono</td>
<td>A/T</td>
<td>270</td>
<td>38</td>
<td>38</td>
<td>36</td>
<td>35</td>
<td>43</td>
<td>45</td>
<td>274</td>
</tr>
<tr>
<td>Di</td>
<td>AT/AT</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>4</td>
<td>12</td>
</tr>
<tr>
<td>Tri</td>
<td>AAT/ATT</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>270</td>
<td>42</td>
<td>42</td>
<td>40</td>
<td>39</td>
<td>39</td>
<td>44</td>
<td>296</td>
</tr>
</tbody>
</table>

The cpDNA of six Viburnum species has the same overall GC contents (38.1%) which is higher than that of T. omeiensis (37.7%). The coding size of V. odoratissimum (78,834 bp) is the second-largest genome among these complete cp genomes, while the smallest is from V. brachybotryum with the size of 47,316 bp. The reason for a significant difference in protein-coding size is a presence of 10 pseudogenes in V. brachybotryum, including ndhB, matK, rpoB, ndhJ, rpoA, rpl16, ycf2, ndhH, rps15, and ycf1. The abundance of pseudogenes in V. brachybotryum have few nucleotides that differ from these genes in other Viburnum species, that could be resulted from genomic mutations or sequencing errors [27].

3.2.2. Repeat Structure

In these studied chloroplast genomes, a total of 270 tandem repeat sequences is identified, with each accession containing 31-47 repeats (Figure 2A). The length of repeated sequences ranges mainly from 31-50 bp, consistent with reports in other angiosperms [28]. These repeats are primarily distributed in the intergenic spaces, but a few are located in the coding region (rpoCl, rps18, ycf2, ycf1, psaA), intron (clpP). In terms of quadripartite structure, tandem repeat equally distributed in the LSC and IR regions, accounting for around 42% each, while the SSC region has only 16.3%.

Simple sequence repeats (SSRs) are repeating short DNA motifs of 1-6 nucleotides that are excellent molecular markers in plant genetics and polymorphism research [29]. Herein, the types and quantity of SSRs are analyzed using MISA software. We found a total of 296 SSRs in the seven studied species, with each accession containing 39-50 SSRs. Most of these SSRs are distributed in the single-copy regions. Mononucleotide is the most frequent repeat, accounting for approximately 94.6% of all SSRs, followed by dinucleotide (4.0%) and trinucleotide (1.4%). A total of 274 mononucleotide repeats (97.9%) is A/T repeat, and all di- and tri-nucleotide repeats are AT/AT repeat and AAT/ATT repeat, respectively.

3.2.3. Ir Contraction and Expansion

The shrinkage and expansion of the IR/SC boundary regions of seven studied species are presented in Figure 3. The rps19 gene, located in the LSC region, extends into the IRb region by 245 bp – 247 bp in all Viburnum species, while the distance from rps19 gene to the border is 97 bp in T. omeiensis. The IRb/SSC boundary region is highly similar between these species. Briefly, the trnN gene and ndhF gene are located on either side of this boundary, separated by 1,372 bp (V. brachybotryum) to 1,977 bp (V. japonicum). The ycf1 gene spans to the regions at the junction of the SSC/IRa region in all seven species with 4,273 bp (V. betulifolium) to 4,733 bp (T. omeiensis) located in the SSC region. The IRa/LSC boundary is quite conserved between Viburnum species. The trnH gene is located in the LSC region and it is 0-80 bp apart from the IR/LSC junction in Viburnum, and 330 bp in Tetradoxa omeiensis.

3.2.4. Divergence Hotspot Regions

To determine the divergent regions that could be applied to the phylogenetic study, the seven chloroplast genomes were aligned with mVISTA (Figure 4). The comparison shows that the IR regions are less divergent than the single-copy regions and the non-coding regions contain more hypervariable regions than the coding regions. The significant difference between these species includes trnH-psaA, atpH-atpI, trnC-petN-psbM, rbcL-accD, psbE-petL and ndhF-rpl32-trnL.

The nucleotide variability values in all seven accessions were detected with DnaSP software to quantify the diversity at the sequence level (Figure 5). The Pi value ranges from 0 to 0.277, indicating a partial difference among these plastomes. As expected, the LSC and the SSC regions than the coding regions. The significant difference between these regions than the IR regions. The region trnC-petN-psbM is the most divergent region with a Pi value of 0.277. We also detect some regions that differ among seven studied species, including intergenic spacers trnH-psbA, ndhC-trnV, trnE-trnT, ndhF-rpl32-trnL and coding regions rpl16 and rpl22. These regions with a high degree of nucleotide variation could be used as potential molecular markers to reconstruct a phylogenetic tree in the Viburnum genus. Choi et al. used the regions of trnK, matK, and rbcL to distinguish Viburnum species [13] but our study shows that these sequences in the Viburnum chloroplast genomes exhibit low divergence.
Figure 3. Comparison of IR/SC boundary of chloroplast genomes in seven species.

Figure 4. The alignment of the seven complete chloroplast genomes.
3.2.5. Phylogenetic Analysis

To identify the phylogenetic relationship between *V. odoratissimum* and other species within the *Viburnum* genus, a Randomized Axelerated maximum likelihood (RAxML) method was performed based on plastid genomes of 10 species, with *T. omeiensis* and *S. nigra* used as outgroups. The resulting phylogenetic tree is shown in Figure 6. The eight *Viburnum* species are divided into three clades. *V. odoratissimum* and *V. brachybotryum* form a clade *Solenotinus*. *V. betulifolium*, *V. japonicum*, *V. erosum* and *V. dilatatum* classify in the clade *Succotinus*. *Solenotius* along with *Succotinus* are members of a large clade *Pluriviburnum*. *V. carcephalum* and *V. utile* belong to the clade *Euviburnum*.

**Figure 6.** Phylogenetic reconstruction based on 10 complete chloroplast genomes

DNA barcodes have proven to possess an expanding range of application in taxonomical studies. In plants, most DNA barcoding regions are located in the chloroplast genome and a few are in the ITS regions of nuclear ribosomal genes [30]. Several chloroplast-derived barcodes were identified and recommended for species discrimination, including coding regions (*matK, rpoB, ycf1, accD, rbcL*, and *ndhJ*) and non-coding regions (*trnH-psbA, apF-atpH*) [31-33]. However, no single DNA region is able to be a promising candidate for all plants. As a result, the combination of a DNA barcode sequence of more than one barcode should be typically used to provide more accurate species identification [34]. In the *Viburnum* genus, a single barcode *trnK*, a combination of
provides a combination of barcodes (by the Seoul National University of Science and Technology. psbA, trnC-petN reconstruction.

In this study, a combination of three hotspot regions (trnH-psbA, trnC-petN, and ndhC-trnV) that exhibit high divergence by a sliding window analysis in the Viburnum species were used to construct a phylogenetic tree of above 10 species. The result (Figure 7) shows a similar pattern with the data based on the complete chloroplast genomes, revealing a high discriminatory power of this combination that could be a promising genetic marker for phylogenetic relationship studies.

4. Conclusion

In this study, using PacBio RS II system sequencing technology, we report the complete chloroplast genome of V. odoratissimum. Compare with the other Adoxaceae genomes, the size of V. Odoratissimum plastid genome and coding regions is largest, but the gene content and organization are highly similar, except for the abundance of pseudogenes in V. brachybotryum. The divergence hotspot region analysis provides a combination of barcodes (trnH-psbA, trnC-petN, and ndhC-trnV) that can be used as a potential genetic marker for discrimination of Viburnum species and phylogenetic tree reconstruction.

Acknowledgements

This study was supported by the Research Project funded by the Seoul National University of Science and Technology.

References


