

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:

Le Thi Yen, Joonho Park. The Complete Chloroplast Genome Sequence of *Viburnum odoratissimum* and Phylogenetic Relationship with Other Close Species in the Adoxaceae Family. *Plant*. Vol. 9, No. 2, 2021, pp. 28-35. doi: 10.11648/j.plant.20210902.12

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: monoterpenes, sesquiterpenes, diterpenes [6], diterpenoids, triterpenoids, iridoids [7, 8], chlorogenic [5], amyrisin, 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. erosum* (MN641480.1) was downloaded from NCBI for a comparison.

2.2. Chloroplast Genome Assembly and Annotation

The filtered subreads were mapped to the reference 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 erosum* (MN218778), *Viburnum brachybotryum* (MN524624) 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 mono-, 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 erosum* (MN218778), *Viburnum brachybotryum* (MN524624), 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).

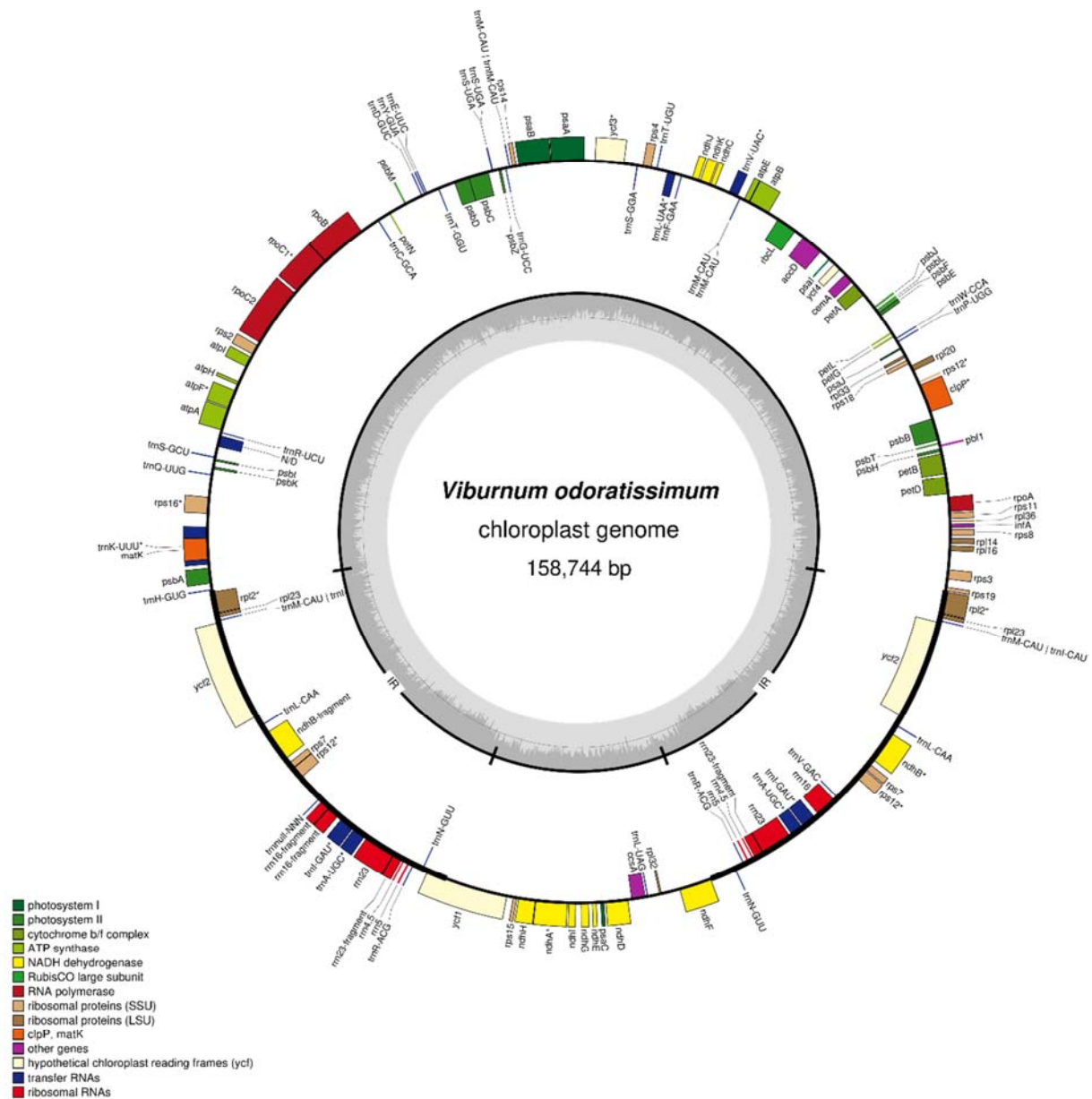


Figure 1. Gene map of *V. odoratissimum* chloroplast genome. Genes drawn inside and outside the circle are transcribed anti-clockwise and clockwise, respectively. Genes are differently colored by the functional groups which genes are affiliated with.

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

Features	<i>Viburnum odoratissimum</i>	<i>Viburnum utile</i>	<i>Viburnum betulifolium</i>	<i>Viburnum japonicum</i>	<i>Viburnum erosum</i>	<i>Viburnum brachybotryum</i>	<i>Tetradoxa omeiensis</i>
Genome size	158,744	157,620	158,023	158,614	158,624	157,433	157,502
LSC Length	87,348	86,576	86,761	87,060	87,060	86,552	86,526
SSC Length	18,408	18,726	18,338	18,523	18,530	18,615	18,682
IR length	26,494	26,159	26,462	26,516	26,517	26,133	26,147
Coding Size	78,834	75,411	77,061	78,852	77,124	47,316	75,525
GC content (%)	38.1	38.1	38.1	38.1	38.1	38.1	37.7
Total number of genes	130	130	130	130	130	130	130
Protein-coding genes	85	85	85	85	85	85	85
Duplicated genes	17	17	17	17	17	17	17
tRNA genes	37	37	37	37	37	37	37
rRNA genes	8	8	8	8	8	8	8
Genes with introns	22	22	22	22	22	22	22
Pseudogenes						<i>ndhB</i> , <i>matK</i> , <i>rpoB</i> , <i>ndhJ</i> , <i>rpoA</i> , <i>rpl16</i> , <i>ycf2</i> , <i>ndhH</i> , <i>rps15</i> , and <i>ycf1</i>	

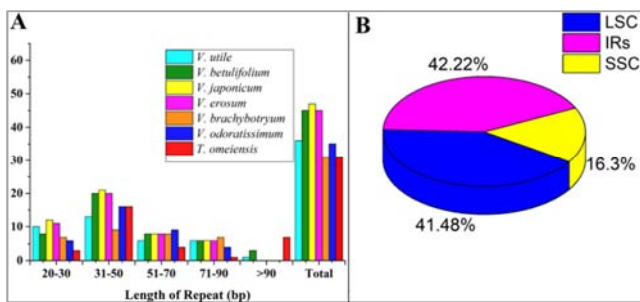
Table 2. Simple sequence repeat analysis in the seven studied species.

SSR Type	Repeat unit	<i>Viburnum odoratissimum</i>	<i>Viburnum utile</i>	<i>Viburnum betulifolium</i>	<i>Viburnum japonicum</i>	<i>Viburnum erosum</i>	<i>Viburnum brachybotryum</i>	<i>Tetradoxa omeiensis</i>	Total
Mono	A/T	39	38	38	36	35	43	45	274
	C/G	2	1	0	1	1	0	1	6
Di	AT/AT	1	2	1	1	2	1	4	12
Tri	AAT/ATT	0	1	1	1	1	0	0	4
Total		42	42	40	39	39	44	50	296

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 (*rpoC1*, *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%.

**Figure 2.** Length and distribution of tandem repeats.

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 P_i value ranges from 0 to 0.277, indicating a partial difference among these plastomes. As expected, the LSC and the SSC regions are higher divergences than the IR regions. The region *trnC-petN-psbM* is the most divergent region with a P_i 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.

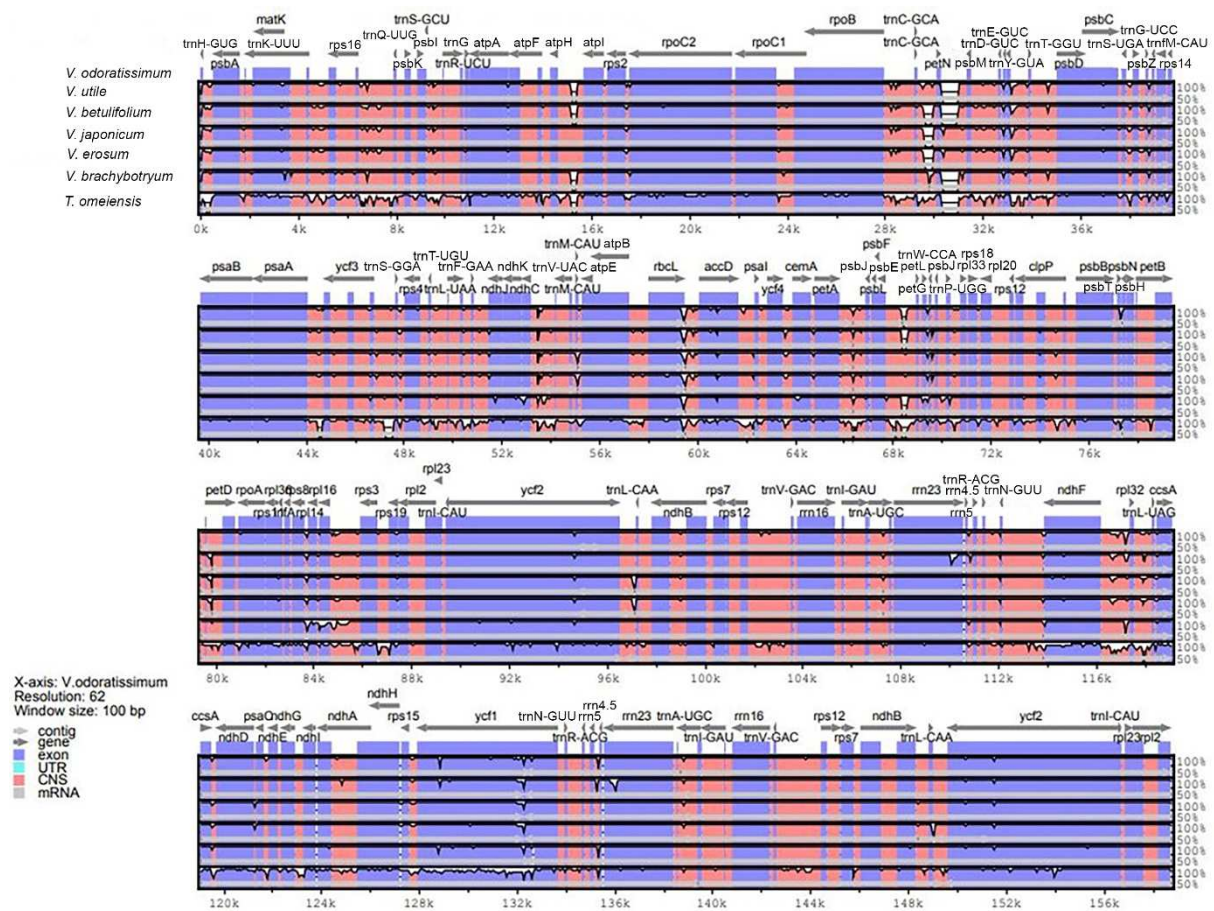


Figure 4. The alignment of the seven complete chloroplast genomes.

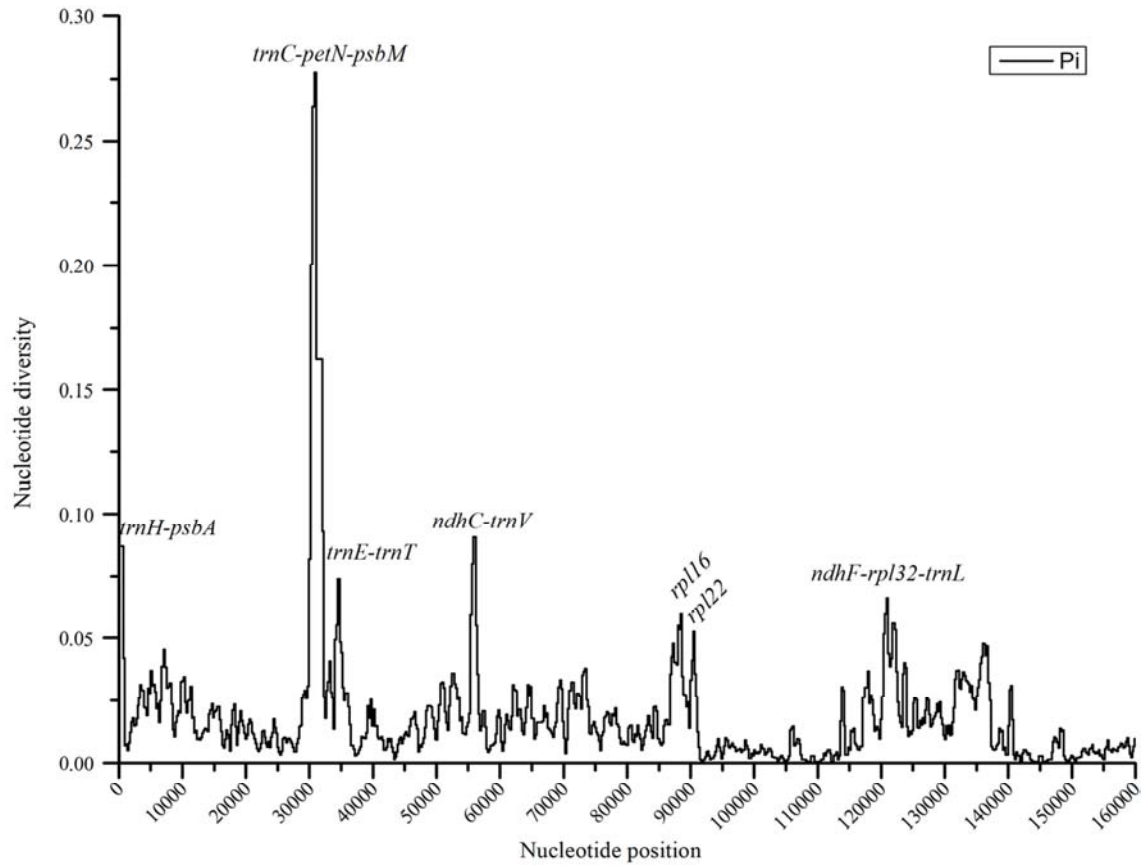


Figure 5. Nucleotide variability (Pi) values of the seven chloroplast genomes using DnaSP analysis.

3.2.5. Phylogenetic Analysis

To identify the phylogenetic relationship between *V. odoratissimum* and other species within the *Viburnum* genus, 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*.

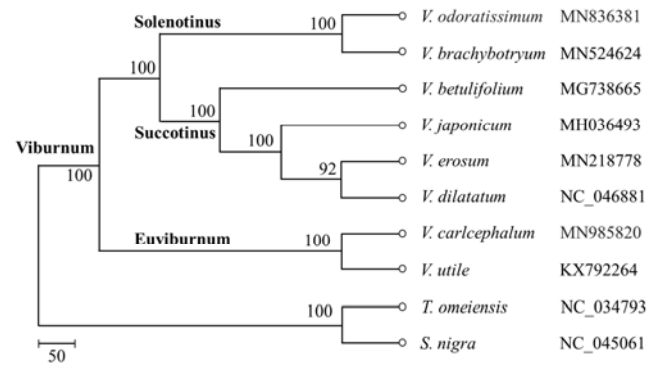


Figure 6. Phylogenetic reconstruction based on 10 complete chloroplast genomes

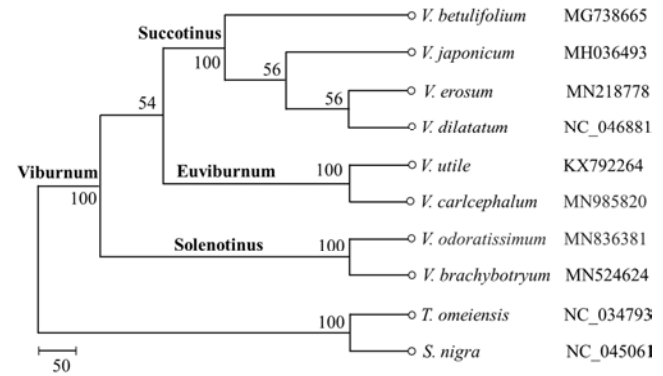


Figure 7. Phylogenetic reconstruction based on the combination of *trnH-psbA*, *trnC-petN*, and *ndhC-trnV* regions from eight *Viburnum* species and two outgroups.

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

rbcL, *matK* and *trnH-psbA*, and the ITS region [1, 13] were used to discriminate *Viburnum* species, but these regions still remain some unresolved issues and conflicting relationships. 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

- [1] Donoghue, M. J., Baldwin, B. G., Li, J. and Winkworth, R. C. (2004) *Viburnum* Phylogeny Based on Chloroplast TrnK Intron and Nuclear Ribosomal ITS DNA Sequences. *Systematic Botany*, 29, 188–198. <https://doi.org/10.1600/036364404772974095>.
- [2] Killip, E. P. and Smith, A. C. (1929) The Genus *Viburnum* in Northwestern South America. *Bulletin of the Torrey Botanical Club*, 56, 265. <https://doi.org/10.2307/2480649>.
- [3] The Angiosperm Phylogeny Group. (2016) An Update of the Angiosperm Phylogeny Group Classification for the Orders and Families of Flowering Plants: APG IV. *Botanical Journal of the Linnean Society*, 181, 1–20. <https://doi.org/10.1111/boj.12385>.
- [4] Ge, Y.-C., Zhang, H.-J., Lei, J.-X. and Wang, K.-W. (2018) Chemical Constituents of *Viburnum Odoratissimum* and Their Cytotoxic Activities. *Chemistry of Natural Compounds*, 54, 600–602. <https://doi.org/10.1007/s10600-018-2422-z>.
- [5] Velioglu, Y. S., Ekici, L. and Poyrazoglu, E. S. (2006) Phenolic Composition of European Cranberrybush (*Viburnum Opulus* L.) Berries and Astringency Removal of Its Commercial Juice. *International Journal of Food Science & Technology*, 41, 1011–1015. <https://doi.org/10.1111/j.1365-2621.2006.01142.x>.
- [6] Wang, L.-Q., Chen, Y.-G., Xu, J.-J., Liu, Y., Li, X.-M. and Zhao, Y. (2008) Compounds from *Viburnum* Species and Their Biological Activities. *Chemistry & Biodiversity*, 5, 1879–1899. <https://doi.org/10.1002/cbdv.200890175>.
- [7] Bock, K., Jensen, S. R., Nielsen, B. J. and Norn, V. (1978) Iridoid Allosides from *Viburnum Opulus*. *Phytochemistry*, 17, 753–757. [https://doi.org/10.1016/S0031-9422\(00\)94220-1](https://doi.org/10.1016/S0031-9422(00)94220-1).
- [8] Nguyen, T. T., Truong, B. N., Doan Thi Mai, H., Litaudon, M., Nguyen, V. H., Do Thi, T., Chau, V. M. and Pham, V. C. (2017) Cytotoxic Dammarane-Type Triterpenoids from the Leaves of *Viburnum Sambucinum*. *Bioorganic & Medicinal Chemistry Letters*, 27, 1665–1669. <https://doi.org/10.1016/j.bmcl.2017.03.014>.
- [9] Sagdic, O., Aksoy, A. and Ozkan, G. (2006) Evaluation of the Antibacterial and Antioxidant Potentials of Cranberry (*Gilaburu*, *Viburnum Opulus* L.) Fruit Extract. *Acta Alimentaria, Akadémiai Kiadó*, 35, 487–492. <https://doi.org/10.1556/AAlim.35.2006.4.12>.
- [10] Wang, L.-X., Fang, Y.-D., Zhang, R.-H., Ren, F.-C., Zhang, X.-J., Wang, F. and Xiao, W.-L. (2018) Hispanane-Type Diterpenoid and Secoiridoid Glucosides from *Viburnum Cylindricum*. *Chemistry & Biodiversity*, 15, e1700418. <https://doi.org/10.1002/cbdv.201700418>.
- [11] Donoghue, M. J. (1983) A Preliminary Analysis of Phylogenetic Relationships in *Viburnum* (Caprifoliaceae s.l.). *Systematic Botany, American Society of Plant Taxonomists*, 8, 45–58. <https://doi.org/10.2307/2418562>.
- [12] Winkworth, R. C. and Donoghue, M. J. (2004) *Viburnum* Phylogeny: Evidence from the Duplicated Nuclear Gene GBSSI. *Molecular Phylogenetics and Evolution*, 33, 109–126. <https://doi.org/10.1016/j.ympev.2004.05.006>.
- [13] Choi, Y. G., Youm, J. W., Lim, C. E. and Oh, S. H. (2018) Phylogenetic Analysis of *Viburnum* (Adoxaceae) in Korea Using DNA Sequences. *Korean Journal of Plant Taxonomy*, 48, 206–217. <https://doi.org/10.11110/kjpt.2018.48.3.206>.
- [14] Clement, W. L. and Donoghue, M. J. (2012) Barcoding Success as a Function of Phylogenetic Relatedness in *Viburnum*, a Clade of Woody Angiosperms. *BMC Evolutionary Biology*, 12, 73. <https://doi.org/10.1186/1471-2148-12-73>.
- [15] Soltis, D., Soltis, P., Endress, P., Chase, M. W., Manchester, S., Judd, W., Majure, L. and Mavrodiev, E. (2018) *Phylogeny and Evolution of the Angiosperms: Revised and Updated Edition*. University of Chicago Press.
- [16] Huo, Y., Gao, L., Liu, B., Yang, Y., Kong, S., Sun, Y., Yang, Y. and Wu, X. (2019) Complete Chloroplast Genome Sequences of Four Allium Species: Comparative and Phylogenetic Analyses. *Scientific Reports*, 9, 1–14. <https://doi.org/10.1038/s41598-019-48708-x>.
- [17] Sahu, S. K., Thangaraj, M. and Kathiresan, K. (2012) DNA Extraction Protocol for Plants with High Levels of Secondary Metabolites and Polysaccharides without Using Liquid Nitrogen and Phenol. *ISRN Molecular Biology*, 2012. <https://doi.org/10.5402/2012/205049>.
- [18] Li, H. and Durbin, R. (2009) Fast and Accurate Short Read Alignment with Burrows–Wheeler Transform. *Bioinformatics*, 25, 1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>.

- [19] Koren, S., Walenz, B. P., Berlin, K., Miller, J. R., Bergman, N. H. and Phillippy, A. M. (2017) Canu: Scalable and Accurate Long-Read Assembly via Adaptive k-Mer Weighting and Repeat Separation. *Genome Research*, 27, 722–736. <https://doi.org/10.1101/gr.215087.116>.
- [20] Tillich, M., Lehwark, P., Pellizzer, T., Ulbricht-Jones, E. S., Fischer, A., Bock, R. and Greiner, S. (2017) GeSeq – Versatile and Accurate Annotation of Organelle Genomes. *Nucleic Acids Research*, 45, W6–W11. <https://doi.org/10.1093/nar/gkx391>.
- [21] Greiner, S., Lehwark, P. and Bock, R. (2019) OrganellarGenomeDRAW (OGDRAW) Version 1.3.1: Expanded Toolkit for the Graphical Visualization of Organellar Genomes. *Nucleic Acids Research*, 47, W59–W64. <https://doi.org/10.1093/nar/gkz238>.
- [22] Benson, G. (1999) Tandem Repeats Finder: A Program to Analyze DNA Sequences. *Nucleic Acids Research*, 27, 573–580. <https://doi.org/10.1093/nar/27.2.573>.
- [23] Beier, S., Thiel, T., Münch, T., Scholz, U. and Mascher, M. (2017) MISA-Web: A Web Server for Microsatellite Prediction. *Bioinformatics*, Oxford Academic, 33, 2583–2585. <https://doi.org/10.1093/bioinformatics/btx198>.
- [24] Frazer, K. A., Pachter, L., Poliakov, A., Rubin, E. M. and Dubchak, I. (2004) VISTA: Computational Tools for Comparative Genomics. *Nucleic Acids Research*, 32, W273–W279. <https://doi.org/10.1093/nar/gkh458>.
- [25] Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. and Higgins, D. G. (1997) The CLUSTAL_X Windows Interface: Flexible Strategies for Multiple Sequence Alignment Aided by Quality Analysis Tools. *Nucleic Acids Research*, 25, 4876–4882.
- [26] Rozas, J., Ferrer-Mata, A., Sánchez-DelBarrio, J. C., Guirao-Rico, S., Librado, P., Ramos-Onsins, S. E. and Sánchez-Gracia, A. (2017) DnaSP 6: DNA Sequence Polymorphism Analysis of Large Data Sets. *Molecular Biology and Evolution*, 34, 3299–3302. <https://doi.org/10.1093/molbev/msx248>.
- [27] Li, W., Yang, W. and Wang, X.-J. (2013) Pseudogenes: Pseudo or Real Functional Elements? *Journal of Genetics and Genomics*, 40, 171–177. <https://doi.org/10.1016/j.jgg.2013.03.003>.
- [28] Li, X., Tan, W., Sun, J., Du, J., Zheng, C., Tian, X., Zheng, M., Xiang, B. and Wang, Y. (2019) Comparison of Four Complete Chloroplast Genomes of Medicinal and Ornamental Meconopsis Species: Genome Organization and Species Discrimination. *Scientific Reports*, 9. <https://doi.org/10.1038/s41598-019-47008-8>.
- [29] Powell, W., Morgante, M., McDevitt, R., Vendramin, G. G. and Rafalski, J. A. (1995) Polymorphic Simple Sequence Repeat Regions in Chloroplast Genomes: Applications to the Population Genetics of Pines. *Proceedings of the National Academy of Sciences*, 92, 7759–7763. <https://doi.org/10.1073/pnas.92.17.7759>.
- [30] Chen, S., Yao, H., Han, J., Liu, C., Song, J., Shi, L., Zhu, Y., Ma, X., Gao, T., Pang, X., Luo, K., Li, Y., Li, X., Jia, X., Lin, Y. and Leon, C. (2010) Validation of the ITS2 Region as a Novel DNA Barcode for Identifying Medicinal Plant Species. *PLOS ONE*, Public Library of Science, 5, e8613. <https://doi.org/10.1371/journal.pone.0008613>.
- [31] Ford, C. S., Ayres, K. L., Toomey, N., Haider, N., Van Alphen Stahl, J., Kelly, L. J., Wikström, N., Hollingsworth, P. M., Duff, R. J., Hoot, S. B., Cowan, R. S., Chase, M. W. and Wilkinson, M. J. (2009) Selection of Candidate Coding DNA Barcoding Regions for Use on Land Plants. *Botanical Journal of the Linnean Society*, Oxford Academic, 159, 1–11. <https://doi.org/10.1111/j.1095-8339.2008.00938.x>.
- [32] Johnson, M. and Trott, T. (2017) DNA Barcoding of *Quercus Falcata*, *Quercus Palustris*, *Quercus Rubra*, and Their Hybrids Using RbcL, MatK, and Ycf1. 26.
- [33] Kress, W. J. and Erickson, D. L. (2007) A Two-Locus Global DNA Barcode for Land Plants: The Coding RbcL Gene Complements the Non-Coding TrnH-PsbA Spacer Region. *PLOS ONE*, Public Library of Science, 2, e508. <https://doi.org/10.1371/journal.pone.0000508>.
- [34] Shneyer, V. S. and Rodionov, A. V. (2019) Plant DNA Barcodes. *Biology Bulletin Reviews*, 9, 295–300. <https://doi.org/10.1134/S207908641904008X>.