Comparative Analysis of Complete Chloroplast Genome Sequences and Insertion-Deletion (Indel) Polymorphisms to Distinguish Five Vaccinium Species

Yonguk Kim 1, Jawon Shin 1, Dool-Ri Oh 1, Ah-Young Kim 2 and Chulyung Choi 1,*

1 Jeonnam Institute of Natural Resources Research, Jeollanam-do 59338, Korea; kyu9801@hanmail.net (Y.K.); sjo8127@naver.com (J.S.); pp0806@nate.com (D.-R.O.)
2 School of Biological Sciences and Technology, Chonnam National University, Gwangju 61186, Korea; kimay1243@hanmail.net

* Correspondence: blockstar@hanmail.net; Tel.: +82-61-860-2620

Received: 3 August 2020; Accepted: 24 August 2020; Published: 25 August 2020

Abstract: We report the identification of interspecific barcoding InDel regions in Vaccinium species. We compared five complete Vaccinium chloroplast (cp) genomes (V. bracteatum, V. vitis-idaea, V. uliginosum, V. macrocarpon, and V. oldhamii) to identify regions that can be used to distinguish them. Comparative analysis of nucleotide diversity from five cp genomes revealed 25 hotspot coding and noncoding regions, occurring in 65 of a total of 505 sliding windows, that exhibited nucleotide diversity (Pi) > 0.02. PCR validation of 12 hypervariable InDel regions identified seven candidate barcodes with high discriminatory powers: accD-trnT-GGU, rpoB-rpoA, ycf2-trnL-GAA, rps12-ycf15, trnV-GAC, and ndhE-ndhF. Among them, the rpoB-rpoA(2) and ycf2-trnL-CAA sequences clearly showed the intraspecific and interspecific distance among five Vaccinium species by using a K2P technique. In phylogenetic analysis, included five Vaccinium species (n = 19) in the Bayesian and Neighbor-Joining (NJ) analysis revered all species in two major clades and resolved taxonomic position within species groups. These two locus provide comprehensive information that aids the phylogenetics of this genus and increased discriminatory capacity during species authentication.

Keywords: chloroplast genome; InDel; nucleotide diversity; species authentication; Vaccinium

1. Introduction

Genus Vaccinium (Ericaceae) includes more than 450 species across Europe, North America, Central America, Central and South East Africa, Madagascar, South East Asia and Malaysia, Korea, China, and Japan [1,2]. Blueberry (Vaccinium ashei J.M. Reade, Vaccinium corymbosum L., Vaccinium angustifolium Aiton), bilberry (Vaccinium myrtillus L.), cranberry (Vaccinium macrocarpon Aiton, Vaccinium oxycoccos L., Vaccinium myrtillusoides Michx.), and lingonberry (Vaccinium vitis-idaea L.) are representative of this genus.

Blueberry is generally divided into three types: The highbush blueberry (V. corymbosum), the lowbush blueberry (V. angustifolium), and the rabbiteye blueberry (V. ashei) [3]. It is commercially cultivated worldwide, particularly in North America, Europe, and Asia [4]. Bilberry, also known as European blueberry, is native to northern and central Europe as well as North America. It is commercially cultivated mainly in North America, while France, Netherlands, Germany, Poland, and Spain are the leading European producers [5]. Cranberry is commonly separated into two types: The American cranberry (Vaccinium macrocarpon) and the European cranberry (Vaccinium oxycoccos). The American cranberry is native to Eastern North America and is commercially cultivated there and throughout Canada [6,7]. The European cranberry, commonly known as the small cranberry,
is cultivated in Central and Northeastern Europe [8]. Lingonberry is native to the alpine regions of North America, Northern Europe, Iceland, Greenland, and Northern Asia [9]. These edible fruits are popular worldwide as a “superfood” due to their bioactive properties, and high levels of antioxidant compounds (phenolics, flavonoids, and tannins), fruit colorants (anthocyanins and carotenoids), vitamins (ascorbic acid), and minerals [1]. Phenolic compounds are exceptionally strong antioxidants which prevent chronic and degenerative diseases, such as cancer and cardiovascular disorders [10].

The southern and northern alpine regions of Korea are native to five *Vaccinium* species: *V. bracteatum* Thunb., *V. oldhamii* Miq., *V. uliginosum* L., *V. Koreanum* Nakai., and *V. vitis-idaea* L. [11]. Although these wild berries have long been used as an edible fruit or traditional medicinal herb, breeding for cultivar improvement is still in the early stages. The commercial value of wild berries has reduced due to the dependence on imported blueberries. Also, most Korean blueberry farms grow the highbush blueberry cultivars, which is a major blueberry crop [12].

In August 2018, Korea enacted and enforced the Nagoya Protocol—an international agreement that aims to share the benefits arising from the use of genetic resources in a fair and equitable way [13]. Native wild berries have therefore attracted the attention of Korean researchers: They are commercially valuable, with potential bio-industrial and medicinal applications, and are also a good alternative to cultivated blueberries. Furthermore, with increasing demand for healthy foods worldwide, studies have shifted their focus to discovering wild relatives with higher concentrations of bioactive compounds [14–16].

In addition, we may be required to provide supporting the phylogenetic and taxonomical research for genetic diversity between *Vaccinium* species.

The multi-copy organellar genomes, including chloroplast (cp) and mitochondrion (mt) genomes, are valuable resources in molecular phylogenetic analysis [17]. The cp genome data can authenticate evolutionary relationships and confirm phylogenetic classifications for plants at the family and genus level. Higher plants contain cp genomes that range between 120 and 180 kb. They contain a pair of inverted repeat (IR) regions separated by a large single copy (LSC) and a small single copy (SSC) region [18]. This quadripartite structure is highly conserved in gene content and genome organization relative to the nuclear and mitochondrial genomes [19]. Comparative analyses between cp genomes of plant species reveal structural variations, such as IR or gene loss, as a result of environmental adaptation [17,18]. The cp genome contains informative genetic markers for phylogenetic and taxonomic analysis at the genus and species levels, due to its mostly uniparental inheritance, dense gene content, and slower evolutionary rate [17].

Recent studies have demonstrated the increased effectiveness of using complete chloroplast genomes over partial cpDNA sequences in land plant phylogenetics [20,21]. However, many phylogenetic studies continue to use DNA barcodes in chloroplast sequences, such as *trnH-psbA*, *rbcL*, and *matK*, which are limited in sequence divergence in several genera.

In this study, we sequenced the complete cp genomes of three *Vaccinium* species using the paired-end sequencing method on an Illumina HiSeq X Ten platform. We compared the gene content of five *Vaccinium* cp genomes and identified hyper-variable insertion-deletion (InDel) regions. We further confirmed that candidate DNA barcodes can be used to authenticate these species. These results provide valuable sequence information for molecular phylogenetics and aid the development of molecular markers in *Vaccinium* species.

2. Materials and Methods

2.1. Comparison of Cp Genomes and Identification of InDel Loci

All five cp genome sequences of the *Vaccinium* genus with complete genome sequence information were downloaded from Genbank (*V. bracteatum*; LC521967, *V. uliginosum*; LC521968, *V. vitis-idaea*; LC521969, *V. macrocarpon*; NC_019616, and *V. oldhamii*; NC_042713). The sequences were aligned using the Clustal W algorithm from MEGA 7.0 [22]. The mVISTA program (http://genome.lbl.gov/vista/
were denatured at 95 °C (Takara, Japan), 10 ng genomic DNA, and 1 µL (10 pM) forward and reverse primers. The sequencing results were analyzed by BLAST searches of the GenBank database. Sequence alignment and data visualization were carried out using the CLC sequence viewer (http://irscope.shinapps.io/irapp/), based on the annotations of their available cp genomes in Genbank.

2.2. Development and Validation of the Candidate DNA Barcodes

To validate interspecies polymorphisms within the chloroplast genomes, specific primers were designed using Primer 3Plus, based on hotspot regions with high nucleotide diversity identified in these Vaccinium cp genomes [25]. All DNA and fresh leaf samples were showed in Table 1 and identified based on our previous study [26]. For the extraction of total genomic DNA, fresh leaf samples of all species (80 mg wet weight) were added to a tube filled with stainless steel beads (2.38 mm in diameter) from a PowerPlantPro DNA Isolation Kit according to the manufacturer’s instructions. PCR amplifications were performed in a reaction volume of 50 µL containing 5 µL 10x Ex Taq buffer (with MgCl2), 4 µL dNTP mixture (each 2.5 mM), Ex Taq (5 U/µL), (Takara, Japan), 10 ng genomic DNA, and 1 µL (10 pM) forward and reverse primers. The mixtures were denatured at 95 °C for 5 min and amplified for 40 cycles at 95 °C for 30 s, 55 °C for 20 s, and 72 °C for 30 s, with a final extension at 72 °C for 5 min. The target DNA was extracted and purified using a MinElute PCR Purification Kit (Qiagen). PCR products were visualized on 1.5% agarose gels with ethidium bromide. Purified PCR products were sequences by CosmoGenetech (Seoul Korea) using the rRNA/rrna sequences of V. macrocarpon and V. oldhamii. The comparison of the LSC/IRB/SSC/IRA junctions among these related species was visualized by IRScope (http://irscope.shinapps.io/irapp/), based on the annotations of their available cp genomes in Genbank.

Table 1. Vaccinium samples used in this study.

| No. | Scientific Name (n) | Common Name | Collection Site | Specimen Code |
|-----|---------------------|-------------|-----------------|---------------|
| 1   | Vaccinium bracteatum Thunb. (n = 3) | Sea blueberry | 33°31′03.5″ N 126°43′00.1″ E | NIBRGR0000424873 |
| 2   | Vaccinium vitis-idaea L. (n = 3) | Cowberry | 38°10′14.5″ N 128°27′39.3″ E | JINR000002321 |
| 3   | Vaccinium uliginosum L. (n = 3) | Bog bilberry, Moonberry | 33°19′37.6″ N 126°34′09.3″ E | NIBRGR0000424897 |
| 4   | Vaccinium macrocarpon Ait. (n = 3) | Cranberry | 34°40′54.3″ N 126°54′38.1″ E | JINR000002326 |
| 5   | Vaccinium oldhamii Miq. (n = 7) | Oldham blueberry | 36°20′45.2″ N 128°01′54.5″ E | NIBRGR0000424923 |
| 6   | Vaccinium vitis-idaea L. (n = 3) | Cowberry | 38°10′14.5″ N 128°27′39.3″ E | JINR000002322 |
| 7   | Vaccinium uliginosum L. (n = 3) | Bog bilberry, Moonberry | 38°10′14.5″ N 128°27′39.3″ E | JINR000002323 |
| 8   | Vaccinium macrocarpon Ait. (n = 3) | Cranberry | 34°40′54.3″ N 126°54′38.1″ E | JINR000002327 |
| 9   | Vaccinium oldhamii Miq. (n = 7) | Oldham blueberry | 36°33′33.7″ N 126°20′20.6″ E | NIBRGR0000424969 |
| 10  | Vaccinium vitis-idaea L. (n = 3) | Cowberry | 34°45′16.8″ N 127°58′43.8″ E | NIBRGR0000424971 |
| 11  | Vaccinium uliginosum L. (n = 3) | Bog bilberry, Moonberry | 34°22′36.6″ N 126°17′26.6″ E | NIBRGR0000424895 |
| 12  | Vaccinium macrocarpon Ait. (n = 3) | Cranberry | 34°48′49.7″ N 126°20′10.2″ E | NIBRGR0000374901 |
2.3. Phylogenetic Analysis

N = 19 samples (3–7 samples per species) from each five species of genus *Vaccinium*, were collected from various localities in South Korea (Table 1). The specimens were deposited at the National Institute of Biological Resources (NIBR) and Jeollanamdo Institute of Natural Resources Research (JINR Korea). The seven loci sequences of each *Vaccinium* species obtained during this study were compared with *Vaccinium* chloroplast genomes in Genbank (accession numbers LC521967, LC521968, LC521969, NC_019616, and NC_042713) using the Basic Local Alignment Search Tool (BLAST, available at http://blast.ncbi.nlm.nih.gov/Blast.cgi). *Rhododendron delavayi* (MN711645) and *Rhododendron pulchrum* (MN182619) was used as the outgroup taxon. The sequences were aligned using the Clustal W algorithm implemented in MEGA ver. 7.0. The phylogenetic tree was constructed using the neighbor-joining (NJ) method in MEGA software. The Komura 2-parameter (K2P) model and bootstrap analysis with 1000 replicates were included. Genetic distances were calculated using the K2P model. Bayesian analysis was conducted with MrBayes ver. 3.2 using two replicates of 1 million generations with the nucleotide evolutionary model. The best-fit model GTR + I + G was implemented using the Akaike Information Criterion (AIC) in MrModeltest ver. 2.3.

3. Results

3.1. Comparative Analyses of the Chloroplast Genome of Five Vaccinium Species

A previous study reported that the three *Vaccinium* cp genome sequences were deposited in NCBI Genbank and published [26]. We compared the features of the newly sequenced cp genomes with those of *V. macrocarpon* (NC_019616) and *V. oldhamii* (NC_042713), already available in NCBI Genbank. The five *Vaccinium* cp genomes contained a pair of inverted repeat regions (IRs: 30,637–34,242 bp) which were separated by a small single copy region (SSC: 2979–3518 bp) and a large single copy region (LSC: 104,552–106,565 bp). All five varied in the number of genes; total gene number ranged from 117 to 147, protein–coding genes from 75 to 85, and tRNA genes from 30 to 38. All five cp genomes contained 8 rRNA genes. The overall GC content in each cp genome was approximately 37.1% (Table 2).

| Name of Taxon | V. bracteatum | V. uliginosum | V. vitis-idea | V. macrocarpon | V. oldhamii |
|---------------|---------------|---------------|---------------|---------------|-------------|
| Accession number | LC521967 | LC521968 | LC521969 | NC_019616 | NC_042713 |
| Genome length | 174,404 | 173,356 | 173,967 | 176,045 | 173,245 |
| LSC length | 106,565 | 105,856 | 106,013 | 104,552 | 108,904 |
| SSC length | 2979 | 3114 | 3518 | 3009 | 3067 |
| IR length | 32,430 | 32,193 | 32,218 | 34,242 | 30,637 |
| Total gene number | 117 | 125 | 125 | 147 | 130 |
| No. of protein coding genes | 79 | 79 | 79 | 75 | 85 |
| No. of tRNA genes | 30 | 38 | 38 | 36 | 37 |
| No. of rRNA genes | 8 | 8 | 8 | 8 | 8 |
| GC content in genome (%) | 36.8 | 36.8 | 36.7 | 37.1 | 37.2 |

These genes can be classified into five categories based on their different roles in the chloroplast (Table 3). The *rpoA*, *rps7*, *rps12*, *rps16*, *petB*, and *petD* genes were present in *V. macrocarpon* and *V. oldhamii*, but not in the other three. The *trnG-GCC* gene was absent from *V. macrocarpon* and *V. oldhamii*. The *trnD-GUC*, *trnF-M-CAU*, *trnK-UUU*, *trnR-UCLU*, *trnV-UIAC*, *trnY-GLA*, and *rpl2* genes were absent from *V. bracteatum*. The *rpl120* and *psbZ* genes were absent from *V. macrocarpon*. The *lhbA*, *infA*, *ycf2*, *ycf15b*, and *ycf68* genes were present only in *V. macrocarpon.*
Table 3. List of genes encoded by the five *Vaccinium chloroplast* genomes. a Gene with two copies; * Gene with one intron; ** Gene with two intron; The symbol • indicate the presence of the gene; -gene loss; Vb, *Vaccinium bracteatum*; Vu, *Vaccinium uliginosum*; Vv, *Vaccinium vitis-idaea*; Vm, *Vaccinium macrocarpon*; Vo, *Vaccinium oldhamii*.

| Gene Category | Gene Group | Gene Names | Vb | Vu | Vv | Vm | Vo |
|---------------|------------|------------|----|----|----|----|----|
| Self-replication | RNA polymerase | rpoA | - | - | - | - | - |
| | | rpoB | • | • | • | • | • |
| | | rpoC1 | • | • | • | • | • |
| | | rpoC2 | • | • | • | • | • |
| | rRNA genes | rrm16 | • | • | • | • | - |
| | | rrm23 | • | • | • | • | • |
| | | rrm4.5 | • | • | • | • | • |
| | | rrm5 | • | • | • | • | • |
| | tRNA genes | trnA-UGC | • | • | • | • | • |
| | | trnC-GCA | • | • | • | • | • |
| | | trnD-GUC | - | • | • | • | • |
| | | trnE-UUC | • | • | • | • | • |
| | | trnF-MAU | - | • | • | • | • |
| | | trnG-GCC | • | • | • | - | - |
| | | trnG-UCC | • | • | • | • | • |
| | | trnH-GUG | • | • | • | • | • |
| | | trnL-CAU | • | • | • | • | • |
| | | trnL-GAU | • | • | • | • | • |
| | | trnG-CAA | - | • | • | • | • |
| | | trnL-CAA | - | • | • | • | • |
| | Self-replication | trnM-CAU | • | • | • | • | • |
| | | trnN-GUI | • | • | • | • | • |
| | | trnP-GGG | - | - | - | • | - |
| | | trnQ-UIU | • | • | • | • | • |
| | | trnR-ACG | • | • | • | • | • |
| | | trnR-UUC | - | • | • | • | • |
| | Small subunit of ribosome | trnS-GCU | • | • | • | • | • |
| | | trnS-GGA | • | • | • | • | • |
| | | trnS-UUA | • | • | • | • | • |
| | | trnT-UGU | • | • | • | • | • |
| | | trnV-GGU | • | • | • | • | • |
| | | trnV-GAC | • | • | • | • | • |
| | | trnV-UAC | - | • | • | • | • |
| | | trnW-CCA | • | • | • | • | • |
| | | trnY-GUA | - | • | • | • | • |
| Gene Category | Gene Group     | Gene Names | Vb | Vu | Vv | Vm | Vo |
|---------------|----------------|------------|----|----|----|----|----|
| Large subunit of ribosome | rpl2     | -          | *  | *  |     | *  | *  |
|               | rpl14     | *          | *  |     | *  | *  | *  |
|               | rpl16     | -          | -  | -  | -  | *  | *  |
|               | rpl20     | -          | *  |     |     | -  | -  |
|               | rpl22     | *          |     | -  | *  | *  | *  |
|               | rpl23     | *          | *  |     | *  |     | *  |
|               | rpl32     | *          |     |     | *  |     | *  |
|               | rpl33     | *          | *  | *  |     | *  |     |
|               | rpl36     | *          | *  | *  | *  | *  | *  |
| Photosynthesis | psaA      | *          | *  | *  | *  | *  | *  |
|               | psaB      | *          | *  | *  | *  | *  | *  |
|               | psaC      | *          | *  | *  | *  | *  | *  |
|               | psaI      | *          | *  | *  | *  | *  | *  |
|               | psaJ      | *          | *  | *  | *  | *  | *  |
|               | ycf3      | *          | *  | *  | *  | *  | *  |
| Photosystem I | lhxA      | -          | -  | -  | -  | *  | -  |
|               | psbA      | -          | -  | *  | *  | -  | -  |
|               | psbB      | *          | *  | *  | *  | *  | *  |
|               | psbC      | *          | *  | *  | *  | *  | *  |
|               | psbD      | *          | *  | *  | *  | *  | *  |
|               | psbE      | *          | *  | *  | *  | *  | *  |
|               | psbF      | *          | *  | *  | *  | *  | *  |
| Photosystem II | psbH      | *          | *  | *  | *  | *  | *  |
|               | psbl      | *          | *  | *  | *  | *  | *  |
|               | psbJ      | *          | *  | *  | *  | *  | *  |
|               | psbK      | *          | *  | *  | *  | *  | *  |
|               | psbL      | *          | *  | *  | *  | *  | *  |
|               | psbM      | *          | *  | *  | *  | *  | *  |
|               | psbN      | *          | *  | *  | *  | *  | *  |
|               | psbT      | *          | *  | *  | *  | *  | *  |
|               | psbZ      | *          | *  | *  | *  | *  | *  |
| NADH dehydrogenase | ndhA      | *          | *  | *  | *  | *  | *  |
|               | ndhB      | *          | *  | *  | *  | *  | *  |
|               | ndhC      | *          | *  | *  | *  | *  | *  |
|               | ndhD      | *          | *  | *  | *  | *  | *  |
|               | ndhE      | *          | *  | *  | *  | *  | *  |
|               | ndhF      | *          | *  | *  | *  | *  | *  |
|               | ndhG      | *          | *  | *  | *  | *  | *  |
|               | ndhH      | *          | *  | *  | *  | *  | *  |
|               | ndhI      | *          | *  | *  | *  | *  | *  |
|               | ndhK      | *          | *  | *  | *  | *  | *  |
|               | ndhJ      | *          | *  | *  | *  | *  | *  |
| Cytochrome b/f complex | petA      | *          | *  | *  | *  | *  | *  |
|               | petB      | -          | -  | -  | -  | *  | *  |
|               | petD      | -          | -  | -  | -  | *  | *  |
|               | petG      | *          | *  | *  | *  | *  | *  |
|               | petL      | *          | *  | *  | *  | *  | *  |
|               | petN      | *          | *  | *  | *  | *  | *  |
|               | rhcL      | *          | *  | *  | *  | *  | *  |
| Rubisco       | accD      | *          | *  | *  | *  | *  | *  |
| Other genes   | ccsA       | *          | *  | *  | *  | *  | *  |
|               | infA       | -          | -  | -  | -  | *  | *  |
|               | ccmA       | *          | *  | *  | *  | *  | *  |
|               | matK       | *          | *  | *  | *  | *  | *  |
In addition, we observed variation in the copy numbers and intron numbers of several genes. Six protein-coding genes, four rRNA genes, and two tRNA genes were present in two copies. Furthermore, rpoA, rps3, rps18, and rpl22 had two copies only in V. macrocarpon. Moreover, two copies of the rps12 gene were identified in both V. macrocarpon and V. oldhamii. Fourteen genes contained introns: These included the rpoC1 RNA polymerase gene, seven tRNA genes, and five protein-coding genes. rpoC1, trnA-UGC, trnL-GAU, ndhA, and ndhB genes contained one intron in all five cp genomes. trnU-UCC and rps16 genes with one intron were identified in the four cp genomes other than V. macrocarpon. The trnK-UUU gene in V. bracteatum and V. macrocarpon, and the trnV-UAC gene in V. bracteatum, contained no introns, while rps3 had one intron in the three cp genomes other than V. macrocarpon and V. oldhamii. The psbA, petB, and petD genes in V. macrocarpon and V. oldhamii contained one intron. Only the ycf3 genes had two introns in each of the five cp genomes. All of the above divergences are shown in Table 3.

We compared the border structure of the five cp genomes in detail (Figure 1). IR regions contained rpl32 and the IRA/LSC border contained a part of the psbA gene. V. vitis-idaea contained two copies of the psbA gene. One in the IRA/LSC border and the other in the IRB region. The ndhF gene was located in the SSC region, between 87 and 203 bp away from the borders. rpl32 resided in IRB, 616-669 bp away from the SSC/IRA border. In V. uliginosum and V. vitis-idaea, the 38 bp trnV-UAC gene was located in the LSC region, while V. macrocarpon had a 661 bp variant, and V. bracteatum lacked this gene.

### Table 3. Cont.

| Gene Name | Vb | Vu | Vv | Vm | Vo |
|-----------|----|----|----|----|----|
| ycf3      | ●  | ●  | ●  | ●  | ●  |
| ycf2      | -  | -  | -  | -  | ●  |
| ycf15b    | -  | -  | -  | -  | ●  |
| ycf68     | -  | -  | -  | -  | ●  |

3.2. Divergence Hotspots of Five Vaccinium Cp Genomes

We compared the sequence divergence among the five Vaccinium cp genomes using mVISTA, with V. macrocarpon annotation as the reference (Figure 2). In general, non-coding regions were more divergent than coding regions. Seventeen non-coding regions—rps4-ndhj, ndhC-rbcl, atpE-psaI, petA-psbM, petN-rps18, psbL-psbE, petD-psbD, lhbA-rps14, psaA-psbK, rpoB-rpoA, rpl16-rps3, rpl23-ndhB, rps7-rps15, ndhl-ndhE, rpl32-ndhF, and ndhF-rpl32—were highly variable among the five cp genomes. Coding regions were more conserved, with the exception of ndhF. To determine the level of sequence divergence, we calculated Pi value for regions spanning 300 bp on either end of coding regions in five
Vaccinium cp genomes with DnaSP 6.0 software. The average value of Pi for InDel diversity for all cp genome sites was 0.01032. Among the 505 windows, 65 windows showed much higher Pi values than the cp genome average (>0.02); this included 17 noncoding and their associated intergenic space regions (trnM-CAU-psal, 0.028; psbM-petN, 0.02; trnC-GCA-rps12, 0.02; psbF-psbB, 0.047; accD-trnT-GGU, 0.027; psaA-trnQ-UUG, 0.021; rpoB-rpoA, 0.055; rpl16-rps18, 0.027; ycf2-trnL-CAA, 0.02; trnN-GIUU-rps15, 0.104; psbA-ndhI, 0.03; rps15-rps32, 0.055; ndhI-psbaA, 0.133; ndhG-ndhl, 0.032; ndhl-psbaA, 0.029; rps32-rps15, 0.0515; ndhI-psbaA, 0.133), and 8 coding and their associated intergenic space regions (ycf3-trnS-GGA, 0.022; trnT-UGU, 0.021; rps12-ycf15, 0.025; trnV-GAC, 0.028; ndhE-ndhf, 0.138; ndhf, 0.0312; ndhf-ndhG, 0.032; psbA, 0.037; (Figure 3). Of the 25 candidates, 13 lie in the LSC region, 4 in both LSC and SSC regions, 4 in the SSC region, 2 in both SSC and IR regions, and 2 in the IR region. The regions containing ycf3-trnS-GGA and ndhf had slightly less polymorphism, and trnM-CAl-psal contained species-nonspecific polymorphisms and multicopy regions. Furthermore, psbF-psbB, rpl16-rps18, trnN-GIUU-rps15, psbaA-ndhI, rps15-rps32, ndhI-psbaA (3), ndhf-ndhfG, ndhG-ndhl, rpl32-rps15, and psbA had multicopy sequences. Therefore, we did not consider these regions for further analyses.

![Figure 2](image)

**Figure 2.** Comparison of five Vaccinium chloroplast genome using mVISTA. The cp genomes of four Vaccinium species were compared with that of *V. macrocarpon*. Blue blocks: Conserved genes, red blocks: Conserved non-coding sequences (CNS). White represents regions with sequence variation among the five Vaccinium species.

We validated the remaining 12 InDel-variable loci by PCR in the Vaccinium species samples (n = 19), to test their suitability as DNA barcodes (Table S1). The eight specific primer sets for six loci (accD-trnT-GGU, rpoB-rpoA, ycf2-trnL-CAA, rps12-ycf15, trnV-GAC, and ndhE-ndhf) successfully amplified their targets in all five species. For the other loci, we failed to amplify the following: trnT-UGU of *V. bracteatum*, *V. uliginosum*, and *V. oldhamii*; psbM-petN of *V. uliginosum* and *V. vitis-idaea*;
trnC-GCA–rps12 of *V. vitis-idaea*; and *psaA-trnQ-UUG* of *V. bracteatum* and *V. oldhamii* (Table 4). The *accD-trnT-GGU* primers specific to *V. uliginosum*, *V. vitis-idaea*, and *V. macrocarpon*, were derived from 8 indels with PCR products of 639, 636, and 630 bp, respectively, whereas the amplicons of *V. bracteatum* and *V. oldhamii* were identical in size (634bp). The *rpoB-rpoA* (2) and (3) primers had unique amplicon sizes that were specific to each species. *V. uliginosum*, *V. vitis-idaea*, *V. macrocarpon*, and *V. oldhamii* yielded band sizes of: 611, 603, 605, 706, and 656 bp spanning 5 indels; 552, 556, 558, 659, and 597 bp spanning 13 indels; and 770, 729, 725, 762, and 733 bp spanning 8 indels, respectively. The *ycf2-trnL-CAA* primers were specific to five species and spanned 11 indels that yielded amplicons of 908, 924, 905, 643, and 911 bp, respectively. The *ndhE-ndhF* primers were specific to *V. uliginosum*, *V. vitis-idaea*, and *V. macrocarpon* and spanned 7 indels with PCR products of 627, 601, and 643 bp, respectively, while the amplicons of *V. bracteatum* and *V. oldhamii* were identical in size (587 bp). To evaluate the sequence divergence in the seven hypervariable InDel regions with successful PCR amplification, we calculated the average pairwise distance for each marker using MEGA 7.0. The *rpoB-rpoA* (2) locus was the most divergent with a maximum pairwise distance of 0.05182, followed by *ndhE-ndhF* (0.04406). Although the lowest genetic distance was observed at *ycf2-trnL-CAA* (0.01768), there are still differences between the five species at that locus. The phylogenetic tree generated from *rpoB-rpoA* (2) and *ycf2-trnL-CAA* sequences datasets by NJ approach tree were identical except for *ndhE-ndhF* sequences. In tree topology of *rpoB-rpoA* (2) and *ycf2-trnL-CAA*, all nineteen *Vaccinium* species recovered in two major clades (A and B) and was placed in the basal position as the sister to the rest of the clades of the five *Vaccinium* species. In clade A, three species (*V. oldhamii*, *V. bracteatum*, and *V. uliginosum*) were present with bootstrap support (76–82%), while clade B covered two species (*V. macrocarpon* and *V. vitis-idaea*) with bootstrap support (77–87%). For *ndhE-ndhF*, in clade A, four species (*V. oldhamii*, *V. vitis-idaea*, *V. bracteatum*, and *V. uliginosum*) were present with a bootstrap support of 85%, while clade B covered only one species (*V. macrocarpon*) with a bootstrap support of 79% (Figure S1). K2P genetic distances within and between the different species of *Vaccinium* for each marker are given in Table S2. Among them, the obtained sequences from *rpoB-rpoA* (2) and *ycf2-trnL-CAA* locus clearly showed the intraspecific and interspecific distance among five *Vaccinium* species. Intraspecific variation of the *rpoB-rpoA* (2) and *ycf2-trnL-CAA* sequences in the *V. oldhamii* was as high as 1.0% and 0.9%, respectively, whereas in *V. bracteatum*, *V. uliginosum*, *V. vitis-idaea*, and *V. macrocarpon* the variation was lower (maximum 0.7%). The *rpoB-rpoA* (2) and *ycf2-trnL-CAA* sequences of *V. macrocarpon* are clearly distinct from those of *V. bracteatum*, *V. uliginosum*, and *V. oldhamii* (K2P genetic distances 3.3–5.8% and 3.7–5.8%, respectively). Furthermore, the *rpoB-rpoA* (2) and *ycf2-trnL-CAA* of *V. vitis-idaea* are clearly distinct from those of *V. bracteatum*, *V. uliginosum*, and *V. oldhamii* (K2P genetic distances 4.4–5.6% and 3.2–5.2%, respectively). However, the *rpoB-rpoA* (2) and *ycf2-trnL-CAA* sequences of *V. oldhamii*, with intraspecific variation of 0.0–2.0% and 0.0–1.9%, respectively, are not clearly distinct from the sequences of *V. bracteatum* (K2P genetic distances 1.2–2.0% and 1.6–2.7%, respectively) and *V. uliginosum* (K2P genetic distances 2.2–3.4% and 1.4–2.8%, respectively). In addition, the *rpoB-rpoA* (2) and *ycf2-trnL-CAA* sequences of *V. macrocarpon* are not clearly distinct from the sequences of *V. vitis-idaea* (K2P genetic distance 1.9–2.6% and 1.1–2.2%, respectively).
still differences between the five species at that locus. The phylogenetic tree generated from rpoB-rpoA (2) and ycf2-trnL-CAA sequences datasets by NJ approach tree were identical except for ndhE-ndhF sequences. In tree topology of rpoB-rpoA (2) and ycf2-trnL-CAA, all nineteen Vaccinium species recovered in two major clades (A and B) and was placed in the basal position as the sister to the rest of the clades of the five Vaccinium species. In clade A, three species (V. oldhamii, V. bracteatum, and V. uliginosum) were present with bootstrap support (76–82%), while clade B covered two species (V. macrocarpon and V. vitis-idaea) with bootstrap support (77–87%). For ndhE-ndhF, in clade A, four species (V. oldhamii, V. vitis-idaea, V. bracteatum, and V. uliginosum) were present with a bootstrap support of 85%, while clade B covered only one species (V. macrocarpon) with a bootstrap support of 79% (Figure S1).

K2P genetic distances within and between the different species of Vaccinium for each marker are given in Table S2. Among them, the obtained sequences from rpoB-rpoA (2) and ycf2-trnL-CAA locus clearly showed the intraspecific and interspecific distance among five Vaccinium species.

Intraspecific variation of the rpoB-rpoA (2) and ycf2-trnL-CAA sequences in the V. oldhamii was as high as 1.0% and 0.9%, respectively, whereas in V. bracteatum, V. uliginosum, V. vitis-idaea, and V. macrocarpon the variation was lower (maximum 0.7%). The rpoB-rpoA (2) and ycf2-trnL-CAA sequences of V. macrocarpon are clearly distinct from those of V. bracteatum, V. uliginosum, and V. oldhamii (K2P genetic distances 3.3–5.8% and 3.7–5.8%, respectively). Furthermore, the rpoB-rpoA (2) and ycf2-trnL-CAA of V. vitis-idaea are clearly distinct from those of V. bracteatum, V. uliginosum, and V. oldhamii (K2P genetic distances 4.4–5.6% and 3.2–5.2%, respectively). However, the rpoB-rpoA (2) and ycf2-trnL-CAA sequences of V. oldhamii, with intraspecific variation of 0.0–2.0% and 0.0–1.9%, respectively, are not clearly distinct from the sequences of V. bracteatum (K2P genetic distances 1.2–2.0% and 1.6–2.7%, respectively) and V. uliginosum (K2P genetic distances 2.2–3.4% and 1.4–2.8%, respectively). In addition, the rpoB-rpoA (2) and ycf2-trnL-CAA sequences of V. macrocarpon are not clearly distinct from the sequences of V. vitis-idaea (K2P genetic distance 1.9–2.6% and 1.1–2.2%, respectively).

Figure 3. Comparison of nucleotide diversity values among the five Vaccinium cp genomes.

Table 4. Validation of 12 molecular markers derived from InDel regions of five Vaccinium cp genomes.

| No. | Locus     | Forward Primer (Sequence 5’ to 3’) | Reverse Primer (Sequence 5’ to 3’) | Product Range (bp) | AS | No. of InDels | Mean Pairwise Distance |
|-----|-----------|------------------------------------|------------------------------------|--------------------|----|---------------|------------------------|
| 1   | trnT-UGU  | CAGTAATCTTTTGCAAAAGGAAAAAAC        | TTCGTCGAACCTTTCACCCCTTTATGA        | 617–653            | 40 | 4             | 0.01683                |
| 2   | psbM-petN | ATGAGAGCTTCTCTGAAATAATTTTG         | CATTCTCTTTCACTCGIAATATGG           | 650–665            | 60 | 2             | 0.01467                |
| 3   | trnC-GCA-rps12 | AATTCGATTGAAATAAAATGGAGGA | GGAATTTGCCAACGTCAAA       | 611–630            | 80 | 3             | 0.02496                |
| 4   | accD-trnT-GGU | GGATCTAAATAGGACCTGTTGATGTTGTAAGAATAGCTGAAA | 630–639            | 100 | 8             | 0.02504                |
| 5   | psaA-trnQ-UUG | ATCCCTGGTATCTTTATCTACTATT    | TTGCTGAATATCAAGTCACACAGA          | 682–718            | 60 | 6             | 0.01364                |
Table 4. Cont.

| No. | Locus     | Forward Primer (Sequence 5′ to 3′) | Reverse Primer (Sequence 5′ to 3′) | Product Range (bp) | AS | No. of InDels | Mean Pairwise Distance |
|-----|-----------|-----------------------------------|-----------------------------------|--------------------|----|---------------|-----------------------|
| 6   | rpoB-rpoA (1) | AAAAGCCCAAT TACGACCAAAAT           | ATCCACGCAAA TGCATCTCTAT          | 603–706            | 20 | 5            | 0.02594               |
| 7   | rpoB-rpoA (2) | GCACTGAGATC TGGCATTTAAT           | GTCACTGACGAG ATTTTGGAC           | 552–659            | 100 | 14           | 0.05182               |
| 8   | rpoB-rpoA (3) | CTTTCTCCTT TCATGCTCTCA             | TCCCCTCTATGTA TGTATTGGC          | 725–770            | 100 | 7            | 0.02646               |
| 9   | ycf2-trnL-CAA | ATTTTTTGGG CTGGATTTGGG            | CTAGGAGGAC AATGATGTGATG          | 643–924            | 100 | 11           | 0.01768               |
| 10  | rps12-ycf15 | CTCTACCTCGG TCCAAAAGAAGC          | CTTCCTGGTTG ACAATGCTCT           | 698–709            | 100 | 6            | 0.01991               |
| 11  | trnV-GAC    | GAGCTCTAAATG GAAATGGAAAGC         | GCCATTGATAC CATTCACTAC           | 625–647            | 100 | 11           | 0.03506               |
| 12  | ndhE-ndhF  | AACATGTTGAGG CACGTGTTGAC          | GAGATTTTTCG TTGCTGCTG            | 587–643            | 100 | 7            | 0.04406               |

AS: Amplified Success rate (%).

4. Discussion

In our previous work, we sequenced the cp genomes of *V. bracteatum*, *V. vitis-idaea*, and *V. uliginosum* using Illumina Hiseq platform, which provided resources for evolutionary and genetic studies of *Vaccinium* [26]. Although a recent study submitted sequences of a few *Vaccinium* species, such as *V. macrocarpon* and *V. oldhamii*, to the NCBI Genbank database, most research has been limited to “core” DNA barcodes, with resolution limited to the species level. By comparing the gene structure, content, and arrangement of five *Vaccinium* cp genomes, we have detected valuable variations in intergenic spacer lengths, which could serve as interspecific DNA barcodes.

Of these five species, *V. macrocarpon* had the largest cp genome and IR length; other species exhibited minor differences in genome and IR length, whereas *V. vitis-idaea* and *V. oldhamii* had the largest SSC and LSC length, respectively. All five cp genomes contained variation in protein coding and tRNA genes, with the exception of *V. uliginosum* and *V. vitis-idaea*, which were identical to the reference. All cp genomes had identical rRNA genes. The *ycf15* and *ycf68* genes were lost in three of the cp genomes, and was inferred as a pseudogene in *V. macrocarpon*. Their function as hypothetical genes is ambiguous in various land plants [28]. The *infA* gene, which codes for a translation initiation factor, was missing in all species other than *V. macrocarpon*. Millen et al. 2001 [29] demonstrated at least 24 independent losses of *infA* in angiosperms, with a transfer into the nucleus in four lineages.

DNA barcodes are universal DNA sequences, such as *rbcL*, *trnH-psbA*, and *matK*, that have a high mutation rate. Their use allows researchers to distinguish a species within a given taxon, and to reliably identify plant species. Because the core DNA barcodes lack sufficient variation between closely related taxa, none of them work across all plant species [17]. With advances in NGS technologies, recent barcoding studies have focused on the use of whole-chloroplast genome-based barcodes. Because they are more efficient at detecting gene loss and determining gene order than the established DNA barcoding, they are better able to distinguish between closely related taxa [30]. The continuing advances in NGS technology may make these the method of choice for plant identification. In contrast to SNPs and SSRs, INDELs have received more attention recently; they are relatively abundant, spread throughout the genome, contribute to both intra- and inter-specific variation, and are suitable for fast and cost-effective genotyping.

In our assessment of nucleotide diversity among five *Vaccinium* cp genomes, we observed sequence divergence in noncoding regions. Among 12 hot spot regions derived from cp genome sequences, we validated seven regions in terms of amplification success and a large number of interspecific indels (accD-trnT-GGU, rpoB-rpoA(2),(3), ycf2-trnL-CAA, rps12-ycf15, trnV-GAC, and ndhE-ndhF).
Our phylogenetic analysis of rpoB-rpoA(2) and ycf2-trnL-CAA gene sequences, the two major clade (A and B) were confirmed, where the clade A represent the species with the fruit color of red type (a closely related species to cranberry) and the clade B represent species with the fruit color of dark blue type (a closely related species to blueberry). Many researchers reported that V. macrocarpon is closely related to V. vitis-idaea based on a phylogenetic analysis of nrITS sequence data [31,32]. We suggest that these intergenic spacers are suitable for use as DNA barcodes; they have good priming sites, and exhibit length variation and interspecific variation. This study analyzed a limited number of Vaccinium species based on available cp genomes; more complete cp genome sequences are needed to resolve the comprehensive phylogenies and genetic divergence within the Vaccinium genus.

Supplementary Materials: The following are available online at http://www.mdpi.com/1999-4907/11/9/927/s1, Table S1: Hypervariable InDels of 12 intergenic spacer regions among five Vaccinium cp genomes. Vb, Vaccinium bracteatum cp genome; Vu, Vaccinium uliginosum cp genome; Vv, Vaccinium vitis-idaea cp genome; Vm, Vaccinium macrocarpon cp genome; Vo, Vaccinium oldhamii cp genome. Table S2: Intra- and interspecific variation of K2P genetic distances of rpoB-rpoA (2), ycf2-trnL-CAA, and ndhE-ndhF sequences for five Vaccinium species; Vb, Vaccinium bracteatum; Vu, Vaccinium uliginosum; Vv, Vaccinium vitis-idaea; Vm, Vaccinium macrocarpon; Vo, Vaccinium oldhamii. Figure S1. Phylogenetic tree derived from (A) rpoB-rpoA (2), (B) ycf2-trnL-CAA, and (C) ndhE-ndhF sequences of specimens of the V. bracteatum, V. uliginosum, V. vitis-idaea, V. macrocarpon, V. oldhamii and outgroup Rh. delavayi and Rh. pulchrum. See details in Table 1.

Author Contributions: Conceptualization, Y.K. and C.C.; methodology, Y.K. and A.-Y.K.; investigation, Y.K. and J.S.; resources, D.-R.O. and J.S.; validation, J.S. and A.-Y.K.; software, Y.K.; writing—original draft preparation, Y.K.; writing—review and editing, Y.K. and C.C.; project administration, C.C.; funding acquisition, D.-R.O. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Bio & Medical Technology Development Program of the National Research Foundation (NRF) and funded by the Korean government (MSIT) [No. 2019M3A913084518].

Acknowledgments: This paper is dedicated to the memory of Heungsu Kim and Jaeman Kim for inspiring and mentoring our work and teaching us as a curious, inquisitive inventor and scientist with a passion for plant biology.

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

References
1. Luby, J.J.; Ballington, J.R.; Draper, A.D.; Pliszka, K.; Austin, M.E. Blueberries and cranberries (Vaccinium). Genet. Resour. Temp. Fruit Nut Crops 1991, 290, 393–458. [CrossRef]
2. Ballington, J.R. Collection, utilization, and preservation of genetic resources in Vaccinium. HortScience 2001, 36, 206–213. [CrossRef]
3. Abreu, O.A.; Barreto, G.; Prieto, S. Vaccinium (Ericaceae): Ethnobotany and pharmacological potentials. Emir. J. Food Agric. 2014, 26, 577. [CrossRef]
4. Strik, B.C. A review of optimal systems for organic production of blueberry and blackberry for fresh and processed markets in the northwestern United States. Sci. Horticul. 2016, 208, 92–103. [CrossRef]
5. Primetta, A.K.; Jaakola, L.; Ayaz, F.A.; Inceer, H.; Riihinen, K.R. Anthocyanin fingerprinting for authenticity studies of bilberry (Vaccinium myrtillus L.). Food Control 2013, 30, 662–667. [CrossRef]
6. Zhu, H.; Senalik, D.; McCown, B.H.; Zeldin, E.L.; Speers, J.; Hyman, J.; Zalapa, J.E. Mining and validation of pyrosequenced simple sequence repeats (SSRs) from American cranberry (Vaccinium macrocarpon Ait.). Theor. Appl. Genet. 2012, 124, 87–96. [CrossRef]
7. Jothi, G.A.; Sircom, J. Native pollinators alone provide full pollination on small-scale commercial cranberry (Ericaceae) farms. Can. Entomol. 2019, 151, 745–756. [CrossRef]
8. Jurikova, T.; Skrovankova, S.; Mlicek, J.; Balla, S.; Snopek, L. Bioactive compounds, antioxidant activity, and biological effects of European cranberry (Vaccinium oxyccocos). Molecules 2019, 24, 24. [CrossRef]
9. Gustavsson, B.A. Genetic variation in horticulturally important traits of fifteen wild lingonberry Vaccinium vitis-idaea L. populations. Euphytica 2001, 120, 173–182. [CrossRef]
10. Hokkanen, J.; Mattila, S.; Jaakola, L.; Pirttilä, A.M.; Tolonen, A. Identification of phenolic compounds from lingonberry (Vaccinium vitis-idaea L.), bilberry (Vaccinium myrtillus L.) and hybrid bilberry (Vaccinium × intermedium Ruthe L.) leaves. J. Agric. Food Chem. 2009, 57, 9437–9447. [CrossRef]
11. Lee, T.B. Coloured Flora of Korea; Hyangmunsa: Seoul, Korea, 2003; Volume 2, pp. 27–29. (In Korean)
12. Kim, C.Y.; Kim, Y.H.; Han, S.H.; Ko, H.C. Current Situations and Prospects on the Cultivation Program of Tropical and Subtropical Crops in Korea. *Korean J. Plant Res.* 2019, 32, 45–52. [CrossRef] [PubMed]

13. Sherman, B.; Henry, R.J. The Nagoya Protocol and historical collections of plants. *Nat. Plants* 2020, 6, 430–432. [CrossRef] [PubMed]

14. Lee, B.L.; Kang, J.H.; Kim, H.M.; Jeong, S.H.; Jang, D.S.; Jang, Y.P.; Choung, S.Y. Polyphenol-enriched *Vaccinium uliginosum* L. fractions reduce retinal damage induced by blue light in A2E-laden ARPE19 cell cultures and mice. *Nutr. Res.* 2016, 36, 1402–1414. [CrossRef] [PubMed]

15. Oh, B.T.; Jeong, S.Y.; Velmurugan, P.; Park, J.H.; Jeong, D.Y. Probiotic-mediated blueberry (*Vaccinium corymbosum* L.) fruit fermentation to yield functionalized products for augmented antimicrobial and antioxidant activity. *J. Biosci. Bioeng.* 2017, 124, 542–550. [CrossRef]

16. Oh, D.R.; Yoo, J.S.; Kim, Y.; Kang, H.; Lee, H.; Lm, S.J.; Hong, J. *Vaccinium bracteatum* leaf extract reverses chronic restraint stress-induced depression-like behavior in mice: Regulation of hypothalamic-pituitary-adrenal axis, serotonin turnover systems, and ERK/Akt phosphorylation. *Front. Pharmacol.* 2018, 9, 604. [CrossRef]

17. Li, X.; Yang, Y.; Henry, R.J.; Rossetto, M.; Wang, Y.; Chen, S. Plant DNA barcoding: From gene to genome. *Biolog. Rev.* 2015, 90, 157–166. [CrossRef]

18. Ravi, V.; Khurana, J.P.; Tyagi, A.K.; Khurana, P. An update on chloroplast genomes. *Plant Syst. Evol.* 2008, 271, 101–122. [CrossRef]

19. Guallberto, J.M.; Mileschina, D.; Wallet, C.; Niazi, A.K.; Weber-Lotfi, F.; Dietrich, A. The plant mitochondrial genome: Dynamics and maintenance. *Biochimie* 2014, 100, 107–120. [CrossRef]

20. Bi, Y.; Zhang, M.F.; Xue, J.; Dong, R.; Du, Y.P.; Zhang, X.H. Chloroplast genomic resources for phylogeny and chloroplast genomics: Analyses including new sequences from the angiosperms *Nuphar advena* Thunb. *V. vitis-idaea* L. and *V. uliginosum* L. (Ericaceae). *Mitochondrion DNA B* 2020, 5, 1843–1844. [CrossRef]

21. Knudsen, B.; Knudsen, T.; Flensborg, M.; Sandmann, H.; Heltzen, M.; Andersen, A.; Mønsted, S. Comparative chloroplast genomics: Analyses including new sequences from the angiosperms *Nuphar advena* and *Ranunculus macranthus*. *BMC Genom.* 2007, 8, 174. [CrossRef]

22. Ruhe, C.A.; Peery, R.; Chumley, T.W.; Dziubek, C.; Fourcade, H.M.; Boore, J.L.; Jansen, R.K. Comparative chloroplast genomics: Analyses including new sequences from the angiosperms *Nuphar advena* and *Ranunculus macranthus*. *BMC Genom.* 2007, 8, 174. [CrossRef]

23. Sultana, N.; Pascual-Díaz, J.P.; Gers, A.; Ilga, K.; Serge, S.; Vitales, D.; García, S. Contribution to the knowledge of genome size evolution in edible blueberries (genus *Vaccinium*). *J. Berry Res.* 2019, 10, 243–257. [CrossRef]