Characterization of Complete Chloroplast Genome of *Allium victorialis* and Its Application for Barcode Markers

Junki Lee, JaeKyung Chon, JongSung Lim, Eun-Kyoung Kim, Gyoungju Nah*

Genome Analysis Center at National Instrumentation Center for Environmental Management, Seoul National University, Seoul 08826, Korea

**ABSTRACT** Chloroplast genome sequencing has served as valuable source for developing DNA markers, including the authentication of plant material used for health supplement from its fraudulent materials. We sequenced and analyzed the chloroplast genome of *Allium victorialis*, a medicinal plant, to discover potential marker regions for the authentication from *Veratrum patulum*, an inedible toxic plant. Although we examined conventional barcode marker loci in chloroplast, *matK* and *rbcL*, there was a difficulty in aligning coding regions and determining PCR primer sequences in these two loci between *A. victorialis* and *V. patulum*, possibly due to the distant evolutionary relationship. Instead, we identified potential DNA markers that carry Insertion/Deletion (InDels) that are able to discriminate these two species around *clpP*, *petB*, *petD*, *rpl22*, and *ycf2* loci. In this analysis, we demonstrated the possibility of developing potential DNA markers in the chloroplast genome other than conventional barcode markers, such as *matK* and *rbcL*. The potential DNA markers identified in this analysis will serve as useful tools for future authentication of *Allium* and *Veratrum* species.

**Keywords** Chloroplast genome, DNA marker, InDel, authentication, *Allium victorialis*, *Veratrum patulum*

**INTRODUCTION**

*Allium victorialis*, also called a victory onion, is a broad-leaved Eurasian species of wild onion. It is a perennial of the *Amaryllis* family that occurs widely in mountainous regions of Europe and several places in Asia (GRIN 2015; Korean National Arboretum 2015). *A. victorialis* has been known as a medicinal plant in Korea. Previous studies reported that *A. victorialis* carries anti-cancer, anti-inflammatory, and anti-oxidant functions (Lee et al. 2001; Shirataki et al. 2001; Woo et al. 2012; Kim et al. 2014).

On the contrary, *Veratrum patulum*, whose vegetative leaves are indistinguishable from those of *A. victorialis*, is a toxic plant. It has been reported that intoxication of *V. patulum* was occurred after ingesting of its leaves (Lee et al. 2010). *V. patulum* contains many types of steroidal alkaloids which are known to be associated with various symptoms, for example, nausea, hypotension, bradycardia, and vomiting (Lee et al. 2010). *V. patulum* is a perennial plant of the *Melanthiaceae* Family which habitats across Europe, Asia, and North America (Do et al. 2013). *V. patulum* has been known to carry over 200 types of steroidal alkaloids and *Veratrum* alkaloids is known to affect sodium voltage gated channels, causing low blood pressure, obstructive sleep apnea, paresthesia, and numbness (Tezuka et al. 1998; Song et al. 2012). Therefore, it is important to develop DNA markers that are able to distinguish *A. victorialis* from *V. patulum* for future prevention of human consumption or being used as fraudulent ingredient in health supplement.

DNA barcode markers have been playing important roles in biodiversity research, breeding program, as well as the authentication of food material plants (Mishra et al. 2017; Cho et al. 2015). DNA barcode markers have been playing important roles in biodiversity research, breeding program, as well as the authentication of food material plants (Mishra et al. 2017; Cho et al. 2015).
Chloroplast genome serves as useful source for discovery of many different types of DNA markers with the advance of next generation sequencing technology. Although the Consortium for the Barcode of Life (CBOL) announced to use matK and rbcL as barcode DNA markers (CBOL 2009), sometimes matK or rbcL can be inappropriate as a marker, especially when two or more species in the comparison are either evolutionary distant or highly close in phylogenetic relationship. Thus, there has been high demand of appropriate marker development. In this analysis, we have sequenced, assembled, and compared the chloroplast genome of *A. victorialis* with those of *V. patulum*. We discovered five potential markers from both genic and intergenic regions that contain Insertion/Deletion (InDels), demonstrating the widespread availability of chloroplast genome sequence in identifying novel DNA markers, other than conventional barcode, *matK* or *rbcL*.

**MATERIALS AND METHODS**

**Plant materials and genomic DNA isolation**

The leaves of *A. victorialis* were provided by Hantaek Botanical Garden (http://www.hantaek.co.kr), Republic of Korea. The leaves were ground in liquid nitrogen and total genomic DNA was extracted using GeneAll® ExgeneTM Plant SV Mini Kit (GeneAll Biotechnology LTD., Seoul, Korea) according to the manufacturer’s instructions. The DNA quantification was performed using Quant-iT™ Picogreen® dsDNA Assay Kit (Invitrogen, Eugene, OR, USA). Genomic DNA library was then constructed using NEXTflex® Rapid DNA Sequencing Kit (Bioo scientific, Austin, TX, USA) according to the manufacturer’s instructions.

**Illumina HiSeq sequencing and quality trimming**

A total of 10,737,912,600 bp were generated by Illumina HiSeq platform from *A. victorialis* genomic DNA library (Table 1). We trimmed raw reads based on the quality score with minimum quality score ≥ 30, using the CLC quality trim (version 4.010.83648, CLC Inc. Aarhus, Denmark).

**Chloroplast genome assembly and phylogenetic analysis**

After retrieving high quality sequences, the *A. victorialis* chloroplast genome assembly was performed, using the CLC Genomics Workbench (version 10.0.3, CLC Inc. Aarhus, Denmark). The assembled contig of *A. victorialis* was aligned against the complete chloroplast genome of *Allium cepa* (KM088013.1) from GenBank for further manual editing, using MAFFT (http://mafft.cbrc.jp/alignment/software/). The annotation of the final contig was performed using CHLOROBOX (https://www.mpimp-golm.mpg.de/chlorobox) and the circular map was generated using OGDRAW (http://ogdraw.mpimp-golm.mpg.de/). Phylogenetic analysis was performed using MEGA 7.0 (Kumar *et al*. 2016) with Maximum Likelihood Method.

**RESULTS**

**Chloroplast genome sequencing and assembly of *A. victorialis***

We performed chloroplast genome sequencing of *A. victorialis* using Illumina HiSeq platform and initially 10,737,912,600 bp of paired-end reads (2x150 bp) were obtained. After quality filtration with Q-value ≥ 30, a total of 6,742,143,548 bp of high quality reads was obtained (Table 1). These high quality reads were assembled into a total of three contigs and the contigs were aligned to the chloroplast genome of *A. cepa* as a reference for further completion into a contig of 154,074 bp (Fig. 1). This final contig, i.e. complete chloroplast genome, was consisted of four major parts, large single copy (LSC) region of 83,170 bp, small single copy (SSC) region of 17,855 bp, and two inverted repeat (IR) regions of 26,526 bp. After annotation, 82 protein-coding genes and 30 tRNA genes were
identified (Fig. 1). The annotated chloroplast genome sequence of *A. victorialis* has been deposited in the GenBank under accession number MF687749.

**Phylogenetic analysis of *A. victorialis***

We compared 154,074-bp chloroplast genome sequence of *A. victorialis* with those of closely related eight species (*A. cepa*: KM088013.1, *Allium sativum*: KY085913.1, *Asparagus officinalis*: KY364194.1, *Oziroe biflora*: KX931463.1, *Anemarrhena asphodeloides*: KX931449.1, *Hosta ventricosa*: KX931460.1, *Hesperoyucca whipplei*: KX931459.1, *Hesperaloe parviflora*: KX931457.1), as well as its counterpart, *V. patulum* (KF437397.2). Based on our phylogenetic tree construction by maximum likelihood, *A. victorialis* is found to be most closely clustered with the lineage of *A. cepa* and *A. sativum*, while most distantly related to *V. patulum* in this tree (Fig. 2).
Identification of chloroplast markers that discriminate *A. victorialis* and *V. patulum*

In the search of potential marker regions, we investigated *matK* and *rbcL* regions which have been known for standard barcodes. However, we failed to properly align the regions of *matK* and *rbcL* in *A. victorialis* and *V. patulum* because these two species were highly diverged in evolutionary relationship. Moreover, it was difficult to search conserved flanking sites for generating universal PCR primers. Thus, we investigated several loci that are able to make proper alignment, exhibiting polymorphisms, as well as possessing conserved flanking regions for PCR primers. Five candidate regions carrying InDels were identified across the chloroplast genomes between *A. victorialis* and *V. patulum* (Fig. 3) and named as AvVp_InDel01–05 (Table 2). The estimated PCR product size pair of *A. victorialis*/ *V. patulum* at individual locus is 530/487 bp at AvVp_InDel01, 292/301 bp at AvVp_InDel02, 110/116 bp at AvVp_InDel03, 173/247 bp at AvVp_InDel04, and 210/189 bp at AvVp_InDel05, respectively (Table 2). The PCR primer set of individual loci was determined and designated in Table 2. The relative positions of multiple InDel sites at individual loci were indicated in Fig. 3. There are three deletions in *A. victorialis* and six deletions in *V. patulum* at AvVp_InDel01, four in *A. victorialis* and two in *V. patulum* at AvVp_InDel02, one in *A. victorialis* at AvVp_InDel03, six in *A. victorialis* at AvVp_InDel04, and one in *V. patulum* at AvVp_InDel05. The size of one InDel event was ranged from 1 bp to 63 bp.

DISCUSSION

Chloroplast genome has greatly contributed to DNA marker development due to the rapid and efficient way of sequencing, as shown in previous studies (Guan *et al.* 2017; Nguyen *et al.* 2017; Wang *et al.* 2017; Yang *et al.* 2017). Two chloroplast barcode genes, *matK* and *rbcL*, have been playing important roles in many studies since CBOL proposed these two loci as the standard barcode regions for land plants (CBOL 2009). However, several studies showed that *matK* and *rbcL* are not enough for correct identification of species under many circumstances. In higher plants, only 72% of 907 samples from 440 species were identified by *matK* and *rbcL*, remaining 28% unclear (CBOL 2009). Furthermore, the development of novel barcode markers may carry unexpected difficulties in proper alignment between the regions of interests, iden-
Chloroplast Genome and Barcode Markers for *A. victorialis*

**Fig. 3.** Schematic diagram of designed five primers between *A. victorialis* and *V. patulum*. The grey colored boxes indicate genes. Blue and red triangles describe deletion of *A. victorialis* and *V. patulum*, respectively. The numbers above or below triangles indicate the length of deleted nucleotides. Purple dashed arrows represent designed primers.

**Table 2.** Estimated PCR product size and primer sequences for five candidate loci in *A. victorialis* (Av) ans *V. patulum* (Vp).

| Marker ID       | Estimated product size (bp) | Primers                      | Location               |
|-----------------|-----------------------------|------------------------------|------------------------|
| AvVp_InDel01    | 503 487                     | F AGGACAAATGATCTCAGTACCACT R TGCCCATTGGTGTTCCAAAAG | *clpP* genic-intergenic region |
| AvVp_InDel02    | 292 301                     | F ACACACTTTTTTGATATGCTCTTT  R TCTTCGAGAATCCACTTCAACT         | *petB* genic-intergenic region |
| AvVp_InDel03    | 110 116                     | F GCTCGAGCCGGATGATGAAA       R AATACAGGATCTCACTTCACTT       | *petD* genic-intergenic region |
| AvVp_InDel04    | 173 247                     | F GGGTTGTACCAAGTCTGAACC     R CGATAAAAAGACCCACTTGTGAT         | *rpl22* genic-intergenic region |
| AvVp_InDel05    | 210 189                     | F CTAAGTACCTCGTTTCTTTTGTC  R TCAAATGAACGATTTGAACCCACTT      | *ycf2* genic region          |

tification of polymorphism at target loci, and investigation of conserved region for PCR primer designing (Kress and Erickson 2008). Therefore, *matK* and *rbcL* loci may not serve as universal barcodes in following cases: (1) when two or more species are highly closely in evolutionary relationship, so that *matK* and *rbcL* loci possess the lack of polymorphism, or (2) when two or more species are distantly related, so that *matK* or *rbcL* locus is not aligned well or difficult to identify conserved PCR primer region due to the sequence diversification. In the study of the *Vicia* species, no allelic polymorphism were detected in *matK* region (Raveendar et al. 2015), suggesting the high level of conservation of standard barcode locus among the *Vicia* species. In the chloroplast genome analysis of 20 genotypes belonging to genus *Cynara*, a total of 73 InDels were identified with 39 polymorphic simple sequence repeats (SSRs) and 34 other InDels (Curci et al. 2016), exhibiting the level of quantity of InDels within genus. In this
analysis, we compared *Allium victorialis* chloroplast genome with those of *Veratrum patulum* and identified five potential regions for future DNA markers. These five regions carry multiple InDels ranged from 1-63 bp in size. InDels in chloroplast genome serve as a useful marker system and outperforms SNPs because PCR products are easily discriminated by size (Melodelima *et al.* 2013; Chaney *et al.* 2016; Curci *et al.* 2016; Daniell *et al.* 2016). Intergenic regions are useful sources for InDels, because these regions underwent relatively rapid mutation due to the less evolutionary constraint. Chloroplast DNA markers for authentication have demonstrated their roles in many aspects as shown in previous studies (Little and Jeanson 2013; Purushothaman *et al.* 2014; Moon *et al.* 2016; Vassou *et al.* 2016). The availability of chloroplast genome sequence to develop the novel loci for DNA markers will become greater in many fields of biology. Therefore, the chloroplast genome sequence and potential DNA markers identified in this study will be good genetic resource for molecular study as well as authentication of *Allium* species.

ACKNOWLEDGEMENTS

This research was supported by a grant (1716MFDS065) from Ministry of Food and Drug Safety in 2017. We highly appreciate Plants Research Center at Hantaek Botanical Garden, Korea for providing *A. victorialis* plant materials.

REFERENCES

CBOL Plant Working Group. 2009. A DNA barcode for land plants. Proc. Natl. Acad. Sci. U.S.A. 106: 12794-12797.
Chaney L, Mangelson R, Ramaraj T, Jellen EN, Maughan PJ. 2016. The complete chloroplast genome sequences for four *Amaranthus* species (Amaranthaceae). Appl. Plant Sci. 4: 1600063.
Curci PL, De Paola D, Sonnante G. 2016. Development of chloroplast genomic resources for *Cynara*. Mol. Ecol. Resour. 16: 562-573.
Daniell H, Lin CS, Yu M, Chang WJ. 2016. Chloroplast genomes: diversity, evolution, and applications in genetic engineering. Genome Biol. 17: 134.
Do HD, Kim JS, Kim JH. 2013. Comparative genomics of four Liliales families inferred from the complete chloroplast genome sequence of *Veratrum patulum* O. Loes. (Melanthiaceae). Gene 530: 229-235.
GRIN (Germplasm Resources Information Network). 2015. *Allium victorialis* L. (https://npgsweb.ars-grin.gov/gringlobal/taxonomydetail.aspx?id=2415). United States Department of Agriculture, Agricultural Research Service. Beltsville (MD). Available from: http://www.ars-grin.gov/.
Guan X, Yuyama N, Stewart A, Ding C, Xu N, Kiyoshi T, *et al.* 2017. Genetic diversity and structure of *Lolium* species surveyed on nuclear simple sequence repeat and cytoplasmic markers. Front. Plant Sci. 8: 584.
Kim HJ, Park MJ, Park HJ, Chung WY, Kim KR, Park KK. 2014. Chemopreventive and anticancer activities of *Allium victorialis* var. *platyphyllum* extracts. J. Cancer Prev. 19: 179-186.
Korea National Arboretum. 2015. English names for Korean native plants. Pocheon: National Arboretum. p. 348.
Kress WJ, Erickson DL. 2008. DNA barcodes: genes, genomics, and bioinformatics. Proc. Natl. Acad. Sci. U.S.A. 105: 2761-2762.
Kumar S, Stecher G, Tamura K. 2016. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. Mol. Biol. Evol. 33: 1870-1874.
Lee KT, Choi JH, Kim DH, Son KH, Kim WB, Kwon SH, *et al.* 2001. Constituents and the antitumor principle of *Allium victorialis* var. *platyphyllum*. Arch. Pharm. Res. 24: 44-50.
Lee SH, Kim DY, Ha DC, Han KH, Jeong WJ, Kang HD, *et al.* 2010. *Veratrum patulum* intoxication developed in a group of twenty-three patients. 79: 417-421.
Little DP, Jeanson ML. 2013. DNA barcode authentication of saw palmetto herbal dietary supplements. Sci. Rep. 3: 3518.
Melodelima C, Lobréaux S. 2013. Complete *Arabis alpina* chloroplast genome sequence and insight into its polymorphism. Meta Gene. 1: 65-75.
Mishra P, Kumar A, Nagireddy A, Mani DN, Shukla AK, Tiwari R, *et al.* 2016. DNA barcoding: an efficient tool to overcome authentication challenges in the herbal market. 14: 8-21.
Moon BC, Kim WJ, Ji Y, Lee YM, Kang YM, Choi G. 2016. Molecular identification of the traditional herbal
Chloroplast Genome and Barcode Markers for *A. victorialis*

medicines, *Arisaematis Rhizoma* and *Pinelliae Tuber*, and common adulterants via universal DNA barcode sequences. Genet. Mol. Res. 15: gmr7064.

Nguyen VB, Park HS, Lee SC, Lee J, Park JY, Yang TJ. 2017. Authentication markers for five major *Panax* species developed via comparative analysis of complete chloroplast genome sequences. J. Agric. Food. 65: 6298-6306.

Purushothaman N, Newmaster SG, Ragupathy S, Stalin N, Suresh D, Arunraj DR, *et al.* 2014. A tiered barcode authentication tool to differentiate medicinal Cassia species in India. 13: 2959-2968.

Raveendar S, Jeon Y-A, Lee J-R, Lee G-A, Lee K-J, Cho G-T, *et al.* 2015. The complete chloroplast genome sequence of Korean landrace “Subicho” pepper (*Capsicum annuum var. annuum*). Plant Breed. Biotech. 3: 88-94.

Shirataki Y, Motohashi N, Tani S, Sunaga K, Sakagami H, Satoh K, *et al.* 2001. Antioxidative activity of *Allium victorialis* L. extracts. Anticancer Res. 21: 3331-3339.

Song Q, Wang S, Zhao W. 2012. Total steroidal alkaloids from *Veratrum patulum* L. Inhibit platelet aggregation, thrombi formation and decrease bleeding time in rats. J. Ethnopharmacol. 141: 183-186.

Tezuka Y, Kikuchi T, Zhao W, Chen J, Guo Y. 1998. Two new steroidal alkaloids, 20-isoveratramine and verapatuline, from the roots and rhizomes of *Veratrum patulum*. J. Nat. Prod. 61: 1078-1081.

Vassou SL, Nithaniyal S, Raju B, Parani M. 2016. Creation of reference DNA barcode library and authentication of medicinal plant raw drugs used in Ayurvedic medicine. BMC Complement Altern. Med. 1: 186.

Wang Y, Xie H, Yang Y, Huang Y, Wang J, Tan F. 2017. Chloroplast and mitochondrial microsatellites for *Millettia pinnata* (Fabaceae) and cross-amplification in related species. Appl. Plant Sci. 5: 1700034.

Woo KW, Moon E, Park SY, Kim SY, Lee KR. 2012. Flavonoid glycosides from the leaves of *Allium victorialis* var. *platyphyllum* and their anti-neuroinflammatory effects. 22: 7465-7470.

Yang J, Vázquez L, Chen X, Li H, Zhang H, Liu Z, *et al.* 2017. Development of chloroplast and nuclear DNA markers for Chinese oaks (*Quercus* Subgenus *Quercus*) and assessment of their utility as DNA barcodes. Front. Plant Sci. 8: 816.