Journal of Species Research 10(2):154-158, 2021

Chloroplast genome of the conserved *Aster altaicus* var. *uchiyamae* B2015-0044 as genetic barcode

Minjee Lee[1], Jae-Sun Yi[2], Jihye Park[1] and Jungho Lee[1,*]

[1]*Green Plant Institute, B-301, Heungdeok IT Valley, Yongin 16954, Republic of Korea*
[2]*Shingu Botanical Garden, Seongnam 13443, Republic of Korea*

**Correspondent: jlee@greenplant.re.kr*

An endemic endangered species, *Aster altaicus* var. *uchiyamae* (Danyang aster) B2015-0044, is cultivated at the Shingu Botanical Garden, which serves as the *ex situ* conservation institution for this species. In this work, we sequenced the chloroplast genome of *A. altaicus* var. *uchiyamae* B2015-0044. We found that the chloroplast (cp) genome of B2015-0044 was 152,457 base pairs (bps) in size: 84,247 bps of large single copy regions (LSC), 25,007 bps of inverted repeats (IRs), and 18,196 bps of small single copy regions. The B2015-0044 cp genome contains 79 protein-coding genes (PCGs), 4 RNA genes, 29 tRNA genes, and 3 pseudo-genes. These results were identical to a previously reported cp genome (Park *et al*., 2017), except for two sites in introns and three in intergenic spacer (IGS) regions. For the intronic differences, we found that *clpP.i1* had a 1-bp small simple repeat (SSR) (T) and *petD.i* had a 3-bp SSR (ATT). We found 1-bp SSRs in the IGSs of *trnT_ggu*~*psbD* and *psbZ*~*trnG_gcc*, C and A, respectively. The IGS of (*ndhF*)~*rpl32* had a SNP. Based on our results, the cp genome of the *A. altaicus* var. *uchiyamae* can be classified into two genotypes, $[C]^{1}$-$[A]^{12}$-$[T]^{12}$-$[ATT]^{4}$-C and $[C]^{2}$-$[A]^{11}$-$[T]^{11}$-$[ATT]^{2}$-A.

Keywords: *Aster altaicus*, chloroplast genome, conservation, genotype, SNP, SSR

DOI:10.12651/JSR.2021.10.2.154

Introduction

Endangered species conservation is a hot issue in botany. In Korea, *ex situ* Conservation Institutions play major roles in the propagation of and finding new habitats for rare species. *Aster altaicus* var. *uchiyamae*, an endemic endangered species, has potential horticultural and agricultural uses, suggesting that there are strong merits to conserving this species. The Danyang aster (*Aster altaicus* var. *uchiyamae*) is a useful garden plant with pink ray flowers. Recently, Mr. Lee of the Shingu Botanical Garden has developed a Danyang aster with white flowers. Unlike the other *Dendranthema* asterid species, which are propagated by cutting (Park *et al*., 2015), *A. altaicus* var. *uchiyamae* is a biennial and propagated by seeds.

Aster altaicus var. *uchiyamae* Kitam. was first collected by K. Uchiyama on October 7, 1902 in Suanbo, Chungcheongbuk Province (Kitamura, 1937). The plant grows in limited quantities at roadsides and sandy riverbanks along the South Han River (personal observation). In the last decade, much of this species' habitat was submerged by dam construction, thus its protection is urgent. *Aster altaicus* var. *uchiyamae* is protected by Korean law as an endangered species and is included in the Korean Red List of Threatened Species (Suh and Kim, 2014). Lines of *A. altaicus* var. *uchiyamae* have been propagated by the Korean Association for *ex situ* Conservation Institution (KAECI), supported by the Ministry of Environment, Republic of Korea.

Genetic information for native plants of different countries is provided in the Access to Genetic Resources and Benefit-Sharing development (ABS) under the Convention on Biological Diversity (CBD). Chloroplast (cp) genetic information has been used for more than three decades because of its maternal inheritance and sequence stability. In *Aster*, the first plastid genome was documented in *A. spathulifolius* (Choi and Park, 2015), which grows at seasides of Korea and Japan. The cp genome of *A. altaicus* var. *uchiyamae* was documented and the cp-DNA was compared to that of *A. spathulifolius* by Park *et al*. (2017). In the study, we reported that *A. altaicus* var. *uchiyamae* cp genome contains 112 genes and 21 introns and consisted of 79 protein-coding genes (PCGs), 4 RNA genes, 29 tRNA genes, and 3 pseudo-genes. Pseudo-genes include

ψ-ycf1, *ψ-rps19*, and *ψ-trnT_GGU*.

The reported plastid genome of *A. altaicus* var. *uchiyamae* (Park *et al*., 2017) was studied from plant material collected in 2010 before dam construction from the population at Jocheon-ri, Chungju city of Chungcheongbuk (CB) Province. Here, we characterize the cp genome of the species from a different collection site, Gyeonggi (GG) Province, to investigate genetic diversity in the cp genome. Complete cp-genomic sequences have been utilized as genetic barcodes in plants (Nock *et al*., 2011; Li *et al*., 2015). Here, we report the complete cp genome of *A. altaicus* var. *uchiyamae* as a genetic barcode.

Materials and Methods

Chloroplast DNA extraction, genome sequencing, assembly, and PCR-based validation

Aster altaicus var. *uchiyamae* was collected under the sample name B2015-0044 as an F4 generation from seeds collected from the Gangcheon Island of Yeoju City, Gyeonggi Province, before the island was submerged by dam construction. Fresh leaves of *A. altaicus* var. *uchiyamae* were collected from the Shingu Botanical Garden and stored in liquid nitrogen until DNA extraction. Total DNA was extracted using the Qiagen DNeasy Plant Mini Kit (Qiagen, Germany). DNA concentration and quality were determined using a Scandrop Nano-volume spectrophotometer (Analytik Jena, Germany). High quality DNA (concentration = 300 ng/μL, A260/280 ratio = 1.8−2.0, and A260/230 ratio = 1.7) was used for polymerase chain reaction (PCR) and sequencing.

The Illumina paired-end genomic library of 200 bps was constructed and sequenced using an Illumina HiSeq 2000 platform. The plastid sequence was obtained using CLC Genomics Workbench version 7.05 as described by Jeong *et al*. (2014). Circular structures of each replicon were confirmed by PCR amplification at their ends and by the joining of Sanger sequencing reads derived from the amplicons. The assemblies were further verified by examining paired-end distance and depth after re-mapping reads on the contig sequences. The BLAST searches of a large contig were verified to be plastid genomes.

Genome annotation, Genome comparison, and Sequence Analysis

For gene annotation of organelle genomes, protein-coding and ribosomal RNA genes were annotated using DOGMA (http://dogma.ccbb.utexas.edu/; Wyman *et al*., 2004). The boundaries of each annotated gene were manually determined by comparing with orthologous genes from other known cp genomes. Genes encoding tRNAs were first predicted using tRNAscan (http://lowelab.ucsc.edu/tRNAscan-SE; Lowe and Eddy, 1997) and ARAGORN version 1.2 (http://130.235.46.10/ARAGORN/; Laslett and Canback, 2004), and were manually verified by predicting the tRNA secondary structure. Circular genome maps were drawn using GenomeVx (Conant and Wolfe, 2008), followed by manual modification. The sequencing data and gene annotation were submitted to GenBank with accession number MK860967.

Results and Discussion

We performed a detailed comparison of two cp genomes of *A. altaicus* var. *uchiyamae*, such that the Sanger-sequenced KX35265 (CB) was revised at four sites and the B2015-0044 cp genome (GG) was determined (Fig. 1). A total of 112 genes, including 79 protein-coding genes, 29 tRNA genes, and 4 rRNA genes reported in Park *et al*. (2017), were identical in DNA sequence. Both cp genomes of *A. altaicus* var. *uchiyamae* contained identical copies of *ψ-ycf1* and *ψ-rps19* in the borders of inverted repeat (IR) and single copy regions. In addition, *ψ-trnT_GGU* copies were identical in length and sequence. The cp genomes were 152,450 bps and 152,457 bps in length (Table 1). The length was variable in large single copy (LSC) regions of 82,240 (CB population) and 82,247 (GG population) bps. However, the lengths of small single copy (SSC) region and IRs were identical, 18,196 bps and 25,007 bps (Table 1).

Table 1. Structural variation in *Aster* chloroplast genomes.

Species	Genome size (bp)	LSC size (bp)	IR size (bp)	SSC size (bp)	Genbank ACC #	Coll. Site/Reference
Aster altaicus var. *uchiyamae*	152,457	84,247	25,007	18,196	MK860967	Yeoju, Gyeonggi [GG], This study
	152,450	84,240	25,007	18,196	NC_034996	Cheongju, Chungbuk [CB], Park *et al*., 2017
Aster indicus	152,885	84,610	25,003	18,269	NC_040126	Liu *et al*. 2018
Aster scaber	152,780	84,521	24,992	18,275	GPI-ASTR01	Jeju Island, Lee *et al*. unpublished
Aster spathulifolius	149,555	81,999	24,796	17,972	NC_027434	Dokdo, KyungBuk, Choi and Park, 2015

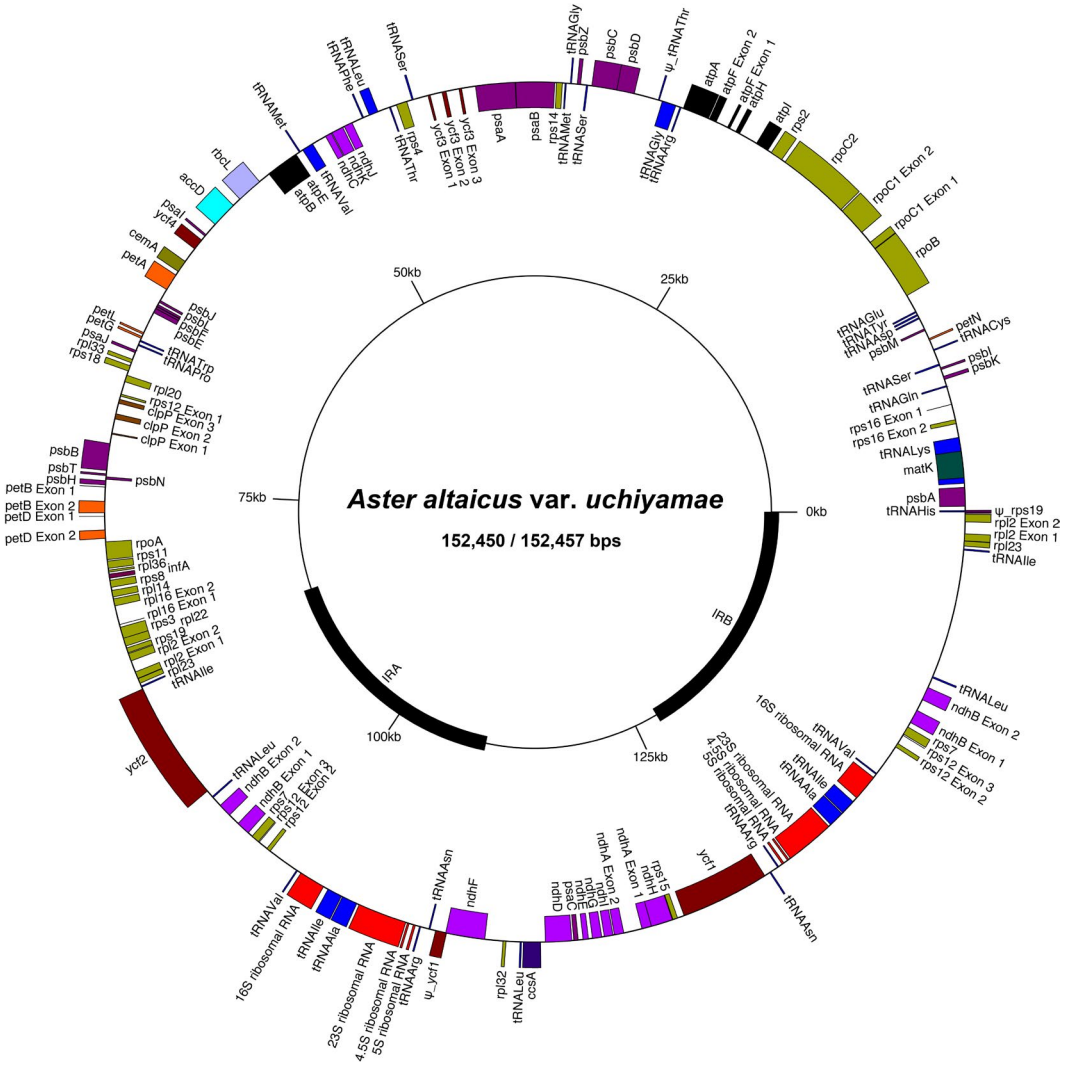

Fig. 1. Plastid genomic map of *Aster altaicus* var. *uchiyamae*.

Table 2. Variable sites between two cp-genomes of *Aster altaicus* var. *uchiyamae*.

	Region	Site	Size (GG/CB)	Type of variation	Number of variation	Memo [GG/CB]
Intron	LSC	clpP_intron1	813/812	INDEL	1	1-bp SSR [$(T)^{12}/(T)^{11}$]
		petD_intron	766/760	INDEL	1	3-bp SSR [$(ATT)^4/(ATT)^2$]
IGS	LSC	trnT_ggu~psbD	1396/1397	INDEL	1	1-bp SSR [$(C)/(C)^2$]
		psbZ~trnG_gcc	322/321	INDEL	1	1-bp SSR [$(A)^{12}/(A)^{11}$]
	SSC	(ndhF)~rpl32	1037	SNP	1	C/A

There are five variable sites in the cp genomes (Table 2). The detailed sequence alignment of the five variable sites is shown in Figure 2. IRs of both genomes are identical in DNA sequence, but SSCs of the two genomes have one SNP in the IGS between *ndhF* and *rpl32*. As well as the single SNP in SSC, small simple repeats (SSR) were found at four variable sites in LSC. The four SSR occur in two introns and two IGSs (Table 2). The introns include the first intron of *clpP*, *clpP.i1*, and the intron of *petD*, *petD.i*. The IGSs include *trnT_ggu~psbD* and *psb-*

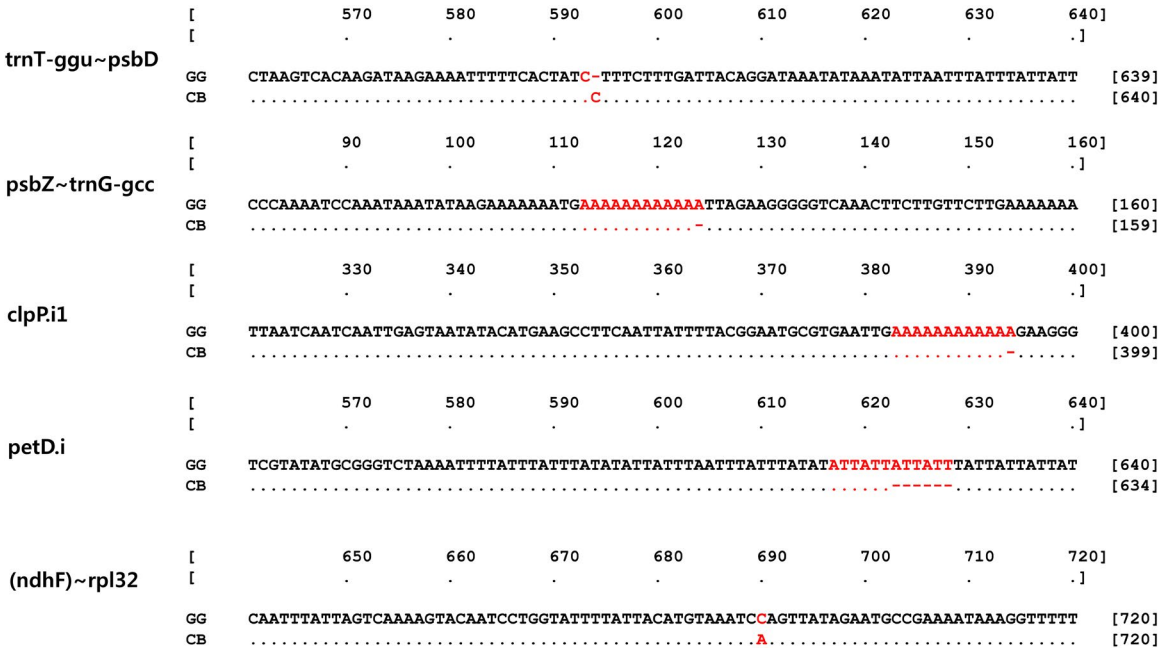

Fig. 2. The sequence alignment of variable sites in the chloroplast genomes of *Aster altaicus* var. *uchiyamae*. Variable sequences are marked in red. GG: Yeoju, Gyeonggi Province, CB: Cheongju, Chungcheongbuk Province.

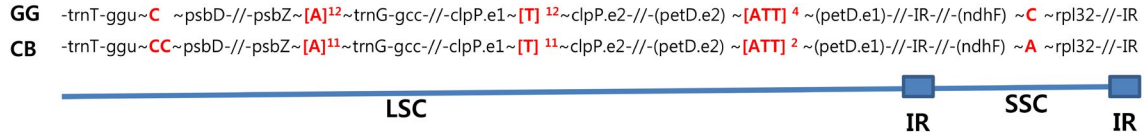

Fig. 3. The variable sites in the chloroplast genomes of *Aster altaicus* var. *uchiyamae*. Variable sequences are marked in red. GG: Yeoju, Gyeonggi Province, CB: Cheongju, Chungcheongbuk Province.

Z~trnG_gcc. A 1-bp SSR of $[T]^{12}$ and $[T]^{11}$ occurs in *clpP.i1*, and 3-bp SSRs of $[ATT]^4$ and $[ATT]^2$ occur in *petD.i*. An IGS of *trnT_ggu~psbD* has [C] and $[C]^2$, and IGS of *psbZ~trnG_gcc* has $[A]^{12}$ and $[A]^{11}$. The position of the variable sites in chloroplast genome structure is shown in Figure 3.

Based on these variations, the chloroplast genome of *A. altaicus* var. *uchiyamae* can be simplified as two genotypes. One genotype, GG, is $[C]^1$-$[A]^{12}$-$[T]^{12}$-$[ATT]^4$-C and the other genotype, CB, is $[C]^2$-$[A]^{11}$-$[T]^{11}$-$[ATT]^2$-A. The relative positions of the variable sites in the cp-map are shown in Figure 3. Though a detailed characterization of *Aster* cp genomes is currently in process, it is clear that four rRNA genes (*16S*, *23S*, *4.5S*, *5S*) are identical in three species of *Aster* (Table 1).

All 112 gene sequences and 3 pseudo-genes were identical in plant materials from two population localities that were 25 km apart. We found minor variation in five sites in introns and IGS regions, which indicates that the species is stable with regard to cp genome evolution.

Conclusion

We investigated two cp genomes of *A. altaicus* var. *uchiyamae*, an endangered species. One genome (CB, KX35265) was from the extinguished population of Chungcheongbuk Province. The other genome (GG, MK860967) was the conserved line B2015-0044 from the Shingu Botanical Garden, which was propagated by seeds from Gyeonggi Province. The chloroplast genomes were 152,450 and 152,457 bps, respectively. GG cp-DNA contains 79 protein-coding genes (PCGs), 4 RNA genes, 29 tRNA genes, and 3 pseudo-genes that were identical to CB cp-DNA. We found differences in five sites at introns and IGS. The introns are *clpP.i1* and *petD.i*, which had a 1-bp SSR (T) and a 3-bp SSR (ATT), respectively. The IGSs include *trnT_ggu~psbD*, *psbZ~trnG_gcc*, and (*ndhF*)~*rpl32*. IGSs of *trnT_ggu~psbD* and *psbZ~trnG_gcc* have 1-bp SSRs, C and A, respectively. The IGS of (*ndhF*)~*rpl32* has a SNP. Based on results, two genotypes of Danyang aster were identified, GG type $[C]^1$-$[A]^{12}$-$[T]^{12}$-

[ATT]^{4}-C and CB type [C]^{2}-[A]^{11}-[T]^{11}-[ATT]^{2}-A. The descendants of B2015-0044 at the Shingu Botanical Garden can be classified as the GG type [C]^{1}-[A]^{12}-[T]^{12}-[ATT]^{4}-C. Additionally, more lines of *A. altaicus* var. *uchiyamae* have been propagated in other *ex situ* conservation institutions. The characterization of these additional lines of the species would be useful for tracing the origin of their respective habitats and the *ex situ* conservation institution, before the species are propagated by the general public.

Acknowledgements

This research was funded by the National Research Foundation of Korea (NRF), grant number NRF-2019R1F1 A1062617, and was also supported by a grant from the National Institute of Biological Resources under the Ministry of Environment (NIBR201703201), Republic of Korea.

References

Choi, K.S. and S. Park. 2015. The complete chloroplast genome sequence of *Aster spathulifolius* (Asteraceae); genomic features and relationship with Asteraceae. Gene 572: 214-221

Conant, G.C. and K.H. Wolfe. 2008. GenomeVx: simple web-based creation of editable circular chromosome maps. Bioinformatics 24:861-862.

Jeong, H., J.M. Lim, J. Park, Y.M. Sim, H.G. Choi, J. Lee and W.J. Jeong. 2014. Plastid and mitochondrion genomic sequences from Arctic Chlorella sp. ArM0029B. BMC Genomics 15:286. https://doi.org/10.1186/1471-2164-15-286

Kitamura, S. 1937. *Aster altaicus* Willd., var. *Uchiyamae* Kitamura var. nov. Memoirs of the College of Science, Kyoto Imperial University, Series B. pp. 367-368.

Laslett, D. and B. Canback. 2004. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. Nucleic Acids Res. 32:11-16.

Li, X., Y. Yang, R.J. Henry, M. Rosseto, Y. Wang and S. Chen. 2015. Plant DNA barcoding: from gene to genome. Biol. Rev. 90:157-166.

Liu, X., B. Zhou, H. Yang, Y. Li, Q. Yang, Y. Lu and Y. Gao. 2018. Sequencing and Analysis of *Chrysanthemum carinatum* Schousb and *Kalimeris indica*. The Complete Chloroplast Genomes Reveal Two Inversions and rbcL as Barcoding of the Vegetable. Molecules 23:1358-1379.

Lowe, T.M. and S.R. Eddy. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. 25:955-964.

Nock, C.J., D.L.E. Waters, M.A. Edwards, S.G. Bowen, N. Rice, G.M. Cordriro and R.J. Henry. 2011. Chloroplast genome sequences from total DNA for plant identification. Plant. Biotechnology Journal 9:328-333.

Park, J., J. Shim, H. Won and J. Lee. 2017. Plastid genome of *Aster altaicus* var. *uchiyamae* Kitam., an endanger species of Korean asterids. Journal of Species Research 6(1):76-90.

Park, S.K., S.-J. Kwon, J. Park, M. Lee, S.Y. Won, Y.C. Kim, Y.-J. Hwang, S.-H. Sohn and J. Lee. 2015. Chloroplast genome of white wild chrysanthemum, *Dendranthema* sp. K247003, as genetic barcode. Journal of Species Research 4(2):152-158.

Suh, Y.B. and Y.J. Kim. 2014. Korean Red List of Threatened Species. 2nd ed. National Institute of Biological Resources, Korea.

Wyman, S.K., R.K. Jansen and J.L. Boore. 2004. Automatic annotation of organellar genomes with DOGMA. Bioinformatics 20:3252-3255.

Submitted: January 3, 2021
Revised: February 2, 2021
Accepted: February 3, 2021