Journal of Species Research 4(2):152-158, 2015

Chloroplast genome of white wild chrysanthemum, *Dendranthema* sp. K247003, as genetic barcode

Sang Kun Park[1], Soo-Jin Kwon[2], Jihye Park[3], Minjee Lee[3], So Youn Won[2], Young Chul Kim[4,5], Yoon-Jung Hwang[6], Seong-Han Sohn[2] and Jungho Lee[3,*]

[1]*Floriculture Research Division, National Institute of Horticultural and Herbal Science (NIHHS), RDA, Jeonju 565-852, Korea*
[2]*Genomics Division, National Academy of Agricultural Science (NAAS), RDA, Jeonju 560-500, Korea*
[3]*Green Plant Institute, B-301, Heungdeok IT Valley, Giheung-gu, Yongin 446-908, Korea*
[4]*Korea Botanic Garden, Pyeongchang 232-953, Korea*
[5]*Department of Biology, Gangneung-Wonju National University, Gangneung 210-702, Korea*
[6]*Department of Life Science, Sahmyook University, Seoul 139-742, Korea*

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

Dendranthema boreale and *D. indicum* are easily distinguished from other Korean *Dendranthema* spp. by having yellow flowers. We have found a putative new taxon of *Dendranthema* having white flowers, except for sharing most characters with *Dendranthema boreale*. The chloroplast (cp) genome of the putative new taxon of *Dendranthema*, *Dendranthema* sp. K247003, registered in National Agro-Biodiversity Center (ABC), was completely characterized as a genetic barcode. The cp-genome of *Dendranthema* sp. K247003 was 151,175-bp in size: LSC was 82,886-bp, IR 24,971-bp, SSC 18,347-bp. The cp-genome of *Dendranthema* sp. K247003 contains 113 genes and 21 introns consisted of 79 protein coding genes, 4 RNA genes, and 30 tRNA genes, with 20 group II introns and one group I intron. Some of the genes and there introns were duplicated in IR. The cp-DNA of *Dendranthema* sp. K247003 is distinguished from that of *D. boreale* IT121002 by 67 SNPs in genic regions of 24 protein coding genes and by a 9-bp INDEL in *ycf1*. Further cp-DNA study will give us better information on genetic markers of *Dendranthema* species.

Keywords: Asteraceae, chloroplast genome, Compositae, *Dendranthema*, INDEL, SNP

DOI:10.12651/JSR.2015.4.2.152

Introduction

Dendranthema, commonly called as chrysanthemum, one of the most popular and economically important ornamental plants due to its huge diversity in growing habits and wide range of colors. Five native *Dendranthema* species are reported from Korea (Park, 2007). Three species, *D. zawadskii*, *D. sichotense* and *D. coreanum*, are known to have ray flowers from white, pale pink to pale violet, while two species, *D. boreale* and *D. indicum* have yellow ray flowers. *Dendranthema boreale* is characterized by having smaller floral heads, 1.2~2 cm in diameter, and globose involucre. During the last several years, we have collected wild *D. boreale* individuals from Korean peninsula and Jeju Island. Their chromosome numbers varied among the populations, i.e. 2n = 18, 18 + 2 (Hwang *et al*., 2014). During the field work, we have found individuals with white floral heads in the middle of yellow headed *D. boreale* population. These white flowered individuals shared other characteristics with the rest of the yellow-flowered population. We registered two white individuals of *D. boreale* (K247002 and K247003) from different localities to National Agro-Biodiversity Center, Rural Agricultural Administration (RDA), as genetic resources for vegetative clones, because chrysanthemum can easily be propagated by cutting.

As the National Agricultural Genome Center Project, Rural Agricultural Administration (RDA), the genome of *D. boreale* IT121002 (2n = 18) has been characterized as a reference genome for chrysanthemum molecular breeding, Marker Assisted Selection (MAS). Organellar genomes having maternal inheritance have been characterized in more than 16 plants of *Dendranthema* including all the five Korean species. Currently, chloroplast genomes from

more than six Asteraceae genera were reported (Dempewolf *et al*., 2010; Doorduin *et al*., 2011; Nie *et al*., 2012; Liu *et al*., 2013; Walker *et al*., 2014). Among them, 47 cp-genomes were reported only from agriculturally important *Helianthus* (Shaw *et al*., 2007; Timme *et al*., 2007; Bock *et al*., 2014), Asteraceae. For *Dendranthema*, while published as *Chrysanthemum*, some IGS regions of chloroplast were used for genetic diversity study (Liu *et al*., 2012). For MAS, genetic information on genic regions is valued than that of IGS regions. Thanks to the dramatic development of Next Generation Sequencing method (NGS) in recent years, it has become possible to complete chloroplast genome sequencing at low cost. Complete cp-genomic sequences have become more useful as genetic barcode of plants (Nock *et al*., 2011; Li *et al*., 2015). Here, we report the complete genome of *Dendranthema* sp. K247003 as genetic barcode.

Materials and Methods

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

White flowered chrysanthemum, *Dendranthema* sp. K247003, was collected at the population of *D. boreale* in Geounri, Yeongwol of Gangwon province (N: 37°15′ 18.9″ E: 128°31′39.3″). The plant was registered at Agro-Biodiversity Center (ABC), Rural Agricultural Administration (RDA), as genetic resources for vegetative clones (IT number: K247003). The plant was propagated in Floriculture Research Division, National Institute of Horticultural and Herbal Science (NIHHS), RDA. Fresh leaves of *Dendranthema* sp. K247003 were collected from the Floriculture Research Institute, NIHHS in Rural Development Administration (RDA), Jeonju, and stored in liquid nitrogen until usage. Total DNA was extracted using the Qiagen DNeasy Plant Mini Kit (Qiagen, Germany), and 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 PCR and sequencing.

For NGS data production, purified DNA was fragmented and used to construct short-insert libraries (insert size, 200-bp), according to the manufacturer's instructions (Illumina, USA). The short fragments were paired-end sequenced using an Illumina Hi-Seq 2500 sequencing system at NICEM of Seoul National University. NGS data (7.63 Gb of 82.97 M reads) were analyzed using CLC Genomic Workbench ver. 7.5.1 (Qiagen, Hilden, Germany), as described by Jeong *et al*. (2014). For Sanger sequencing, the whole cp-genome of *Dendranthema* sp. K247003 was PCR-amplified in ~1-2 kb fragments, and cp-genome structure was verified using Long PCR, with ~5-28 kb fragments, as described by Lee and Manhart (2002a; 2002b). Only PCR products ranging from ~1-2 kb were sequenced using Bigdye (ver. 3.1) and ABI3730 at NICEM of Seoul National University. Assembled cp-sequences were verified using Sequencher ver. 5.0 (Gencode, USA) by combining Sanger data and the assembled NGS sequence.

Genome annotation, genome comparison and sequence analysis

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 comparison 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 National Agricultural Biotechnology Information Center (NABIC), Jeonju, with accession number NG-0482-000001. The mVISTA program in Shuffle-LAGAN mode (Fraser *et al*., 2004) was used to compare the cp-genome of *Dendranthema* sp. K247003 with that of *D. boreale* IT121002 (NABIC: NG-0478-000001; unpublished).

Results and Discussion

The cp-genome of *Dendranthema* sp. K247003, was determined (Fig. 1) and found to be 151,175 bp in length. It includes small and large single copy (SSC, LSC) regions of 18,347 bp and 82,886 bp, respectively, separated by a pair of 24,971 bp Inverted Repeats (IRs). A total of 113 genes were detected, including 79 protein coding genes, 30 tRNA genes, and four rRNA genes (Table 1). This cp-genome was also found to contain 20 different introns, including 19 group II introns and a group I intron with a cyanobacterial origin (Besendahl *et al*., 2000) found within the *trnL_uaa* gene. Three protein coding genes, including *clpP*, *rps12*, and *ycf3*, contain two group II introns (*clpP.i1*, *clpP.i2*, *rps12.i1*, *rps12.i2*, *ycf3.i1* and *ycf3.i2*), and 14 genes contain a single group II intron: *rpoC1.i*, *rpl2.i*, *rpl16.i*, *rps16.i*, *atpF.i*, *petB.i*, *petD.i*, *ndhA.i*, *ndhB.i*, *trnA_ugc.i*, *trnG_ucc.i*, *trnI_gau.i*, *trnK_uuu.i*, and *trnV_uac.i*. Among the 20 group II introns, the intron in *rps12*, between exons 1 and 2, is trans-splicing, while the other 19 group II introns are cis-splicing.

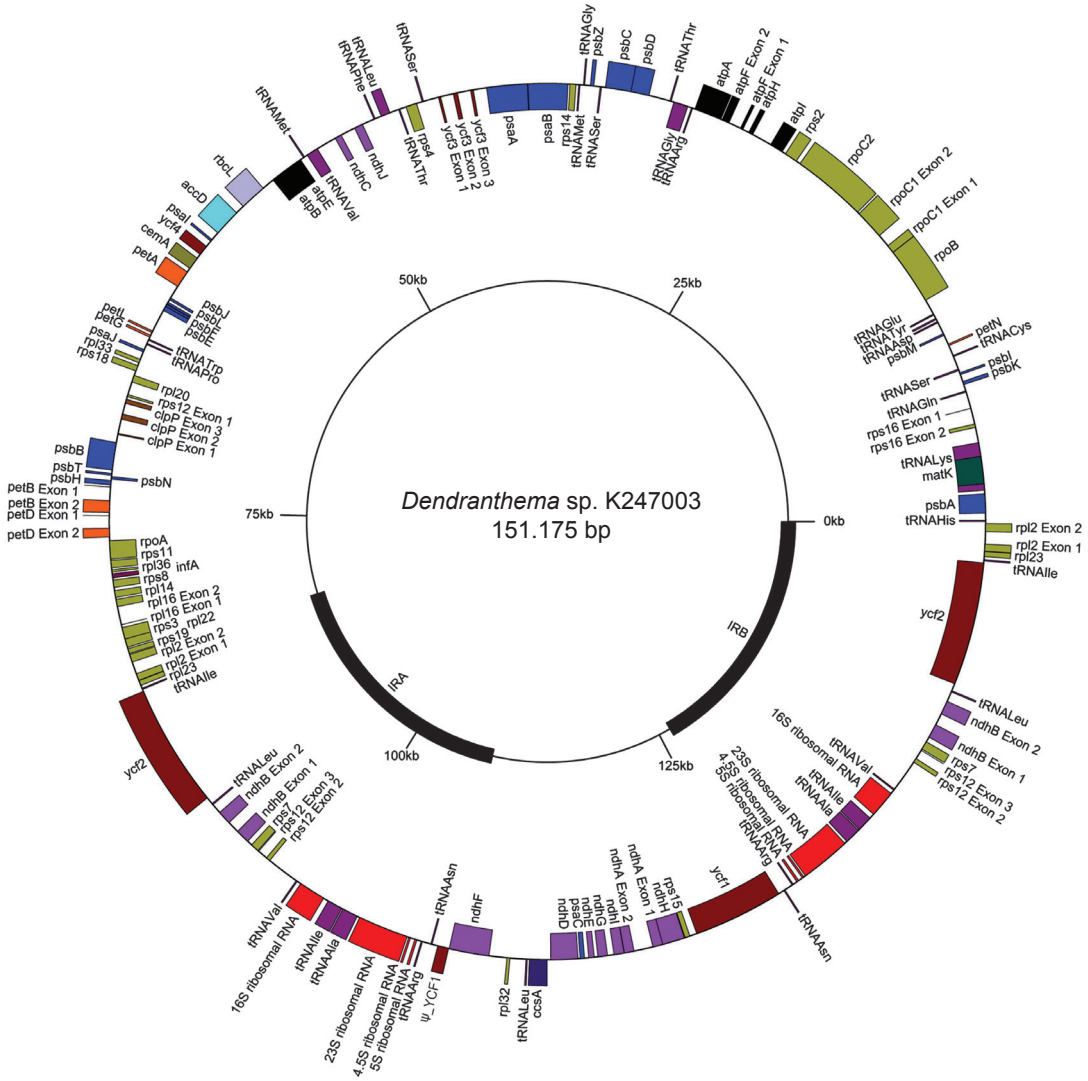

Fig. 1. Plastid genomic map of *Dendranthema* sp. K247003.

Eighteen genes, five introns, and parts of two genes and an intron are found within the IR, which has two copies. These 18 genes include seven protein-coding genes (*ndhB*, *rpl2*, *rpl22*, *rpl23*, *rps7*, *rps19*, *ycf2*), all four rRNA genes (*16S*, *23S*, *4.5S*, *5S*), and seven tRNA genes (*trnA_ugc*, *trnI_cau*, *trnI_gau*, *trnL_caa*, *trnN_guu*, *trnR_acg*, *trnV_gau*). The five introns are *ndhB.i*, *rpl2.i*, *trnA_ugc.i*, *trnI_gau.i*, and *rps12.i2*. The IR also contains the 5′ end of *ycf1* at the border with the SSC, resulting in one intact *ycf1* and a 612-bp *ψ-ycf1* in the cp-genome. In addition, the IR contains parts of the *rps12* gene. This *rps12* gene consists of three exons, *rps12.e1*, *rps12.e2*, and *rps12.e3* (Lee, 1997); *rps12.e1* is in the LSC, but *rps12.e2* and *rps12.e3* are in the IR. Thus, the genome contains a single copy of *rps12.e1* but has two copies of *rps12.e2* and *rps12.e3*. A cis-splicing group II intron, *rps12.i2*, intervenes between *rps12.e2* and *rps12.e3*, but a trans-splicing intron, *rps12.i1t*, occurs between *rps12.e1* and *rps12.e2*. The *rps12.i1t* is split into two pieces, *rps12.i1t1* and *rps12.i1t2*, because the *rps12* gene is transcribed in two separate operons, the *clpP* operon (*clpP-rps12.e1-rps12.i1t1-rpl20*) and the 3′ *rps12* operon (*rps12.i1t2-rps12.e2-rps12.i2-rps12.e3-rps7-ndhB*).

Direct comparison of chloroplast genomes of *Dendranthema* sp. K247003 and *D. boreale* IT121002 using mVISTA program is shown in Figure 2. In the analysis, 97 IGS regions showed variation. In genes, 24 genes have variations. There were no variation in 30 tRNA and 4 rRNA genes. Among 79 protein coding genes, 24 genes had variation. There were two kinds of variations (Table 2). One was Single Nucleotide Polymorphism (SNP) and the other was INDEL. SNPs were found at 67 sites in 24 protein coding genes. The protein coding genes include 6 genetic system genes and 18 photosynthesis

Table 1. Gene list in chloroplast genome of *Dendranthema* sp. K247003.

Gene									
Genetic system genes									
Conserved orf	ycf1	ycf2 ×2	ycf3**	ycf4					
maturase K	matK								
RNA polymerase	rpoA	rpoB	rpoC1*	rpoC2					
Ribosomal protein									
Large subunits	rpl2* ×2	rpl14	rpl16*	rpl20	rpl22	rpl23 ×2	rpl32	rpl33	rpl36
Small subunits	rps2	rps3	rps4	rps7 ×2	rps8	rps11	rps12**α ×2	rps14	rps15
	rps16*	rps18	rps19						
Photosynthesis genes									
Acetyl-CoA carboxylase	accD								
ATP-dependent Clp protease	clpP**								
ATP synthase	atpA	atpB	atpE	atpF*	atpH	atpI			
Cytochrome b	petB*								
Cytochrome b/f	petD*	petG	petL	petN					
Cytochrome f	petA								
Cytochrome C biogenesis	ccsA								
Membrane protein	cemA								
NADH dehydrogenase	ndhA*	ndhB* ×2	ndhC	ndhD	ndhE	ndhF	ndhG	ndhH	ndhI
	ndhJ	ndhK							
Photosystem I	psaA	psaB	psaJ	psaC	psaI				
Photosystem II	psbA	psbB	psbC	psbD	psbE	psbF	psbH	psbI	psbJ
	psbK	psbL	psbM	psbN	psbT	psbZ			
Rubisco	rbcL								
Translation initiation factor 1	InfA								
Ribosomal RNA	rrn16S ×2	rrn23S ×2	rrn4.5S ×2	rrn5S ×2					
Transfer RNA	trnA_UGC* ×2	trnC_GCA	trnD_GUC	trnE_UUC	trnF_GAA	trnfM_CAU	trnG_GCC	trnG_UCC*	trnH_GUG
	trnI_CAU ×2	trnI_GAU* ×2	trnK_UUU*	trnL_CAA ×2	trnL_UAA*	trnL_UAG	trnM_CAU	trnN_GUU ×2	trnP_UGG
	trnQ_UUG	trnR_ACG ×2	trnR_UCU	trnS_GCU	trnS_GGA	trnS_UGA	trnT_GGU	trnT_UGU	trnV_GAC ×2
	trnV_UAC*	trnW_CCA	trnY_GUA						
Pseudo gene	ψ-ycf1	ψ-rps19							

*intron $^{\alpha}$cp genome contains a copy of rps12 exon 1 in LSC and two copies of rps12 exon 2 and 3 in IR

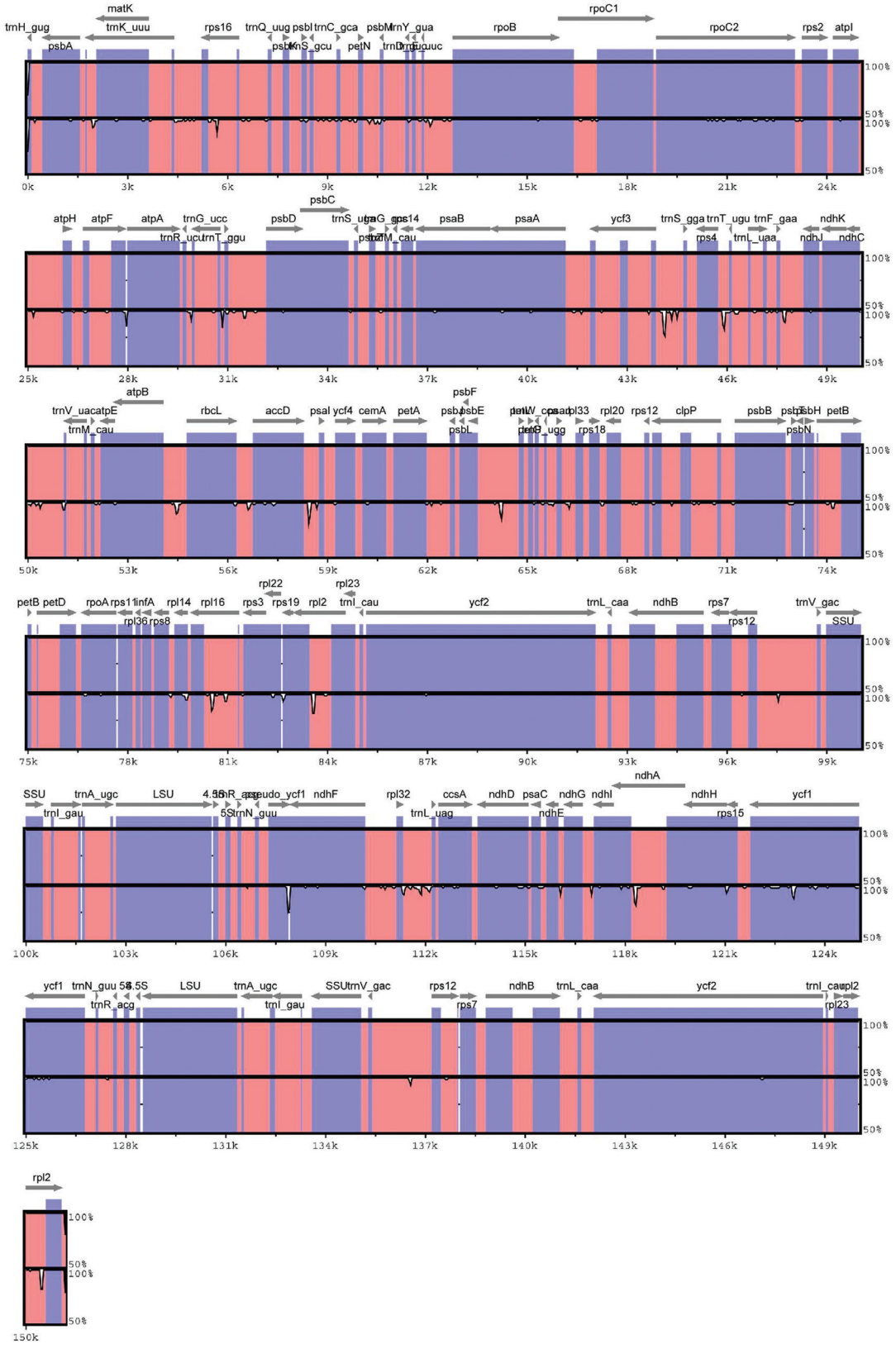

Fig. 2. Comparison of chloroplast genomes of *Dendranthema* sp. K247003 and *D. boreale* IT121002 using mVISTA program. Grey arrows and thick black lines above the alignment indicate genes with their orientation and the position of the IRs, respectively. The Y-scale represents the percent identity between 50-100%. Genome regions are color-coded: Coding regions in blue; noncoding sequences (CNS) in red.

Table 2. Variation in cp-genic regions between *Dendranthema* sp. K247003 and *D. boreale* IT121002. SNP: Single Nucleotide Polymorphism.

	Gene name	Size (*D.* sp / *D. boreale*)	No of SNP	INDEL		Gene name	Size (D. sp / *D. boreale*)	No of SNP	INDEL
1	*psbA*	1062	1	–	13	*accD*	1503	4	–
2	*matK*	1518	2	–	14	*cemA*	690	1	–
3	*rpoB*	3183	1	–	15	*rpoA*	1008	2	–
4	*rpoC2*	4152	10	–	16	*rpl22*	468	2	–
5	*atpI*	744	1	–	17	*ycf2*	6849	1	–
6	*atpF*	555	2	–	18	*ndhF*	2226	1	–
7	*psbD*	1062	1	–	19	*ccsA*	975	3	–
8	*psaB*	2205	1	–	20	*ndhD*	1503	3	–
9	*psaA*	2253	2	–	21	*ndhI*	561	1	–
10	*ndhJ*	477	1	–	22	*ndhA*	1092	3	–
11	*atpB*	1479	1	–	23	*ndhH*	1236	2	–
12	*rbcL*	1479	1	–	24	*ycf1*	5016 / 5007	19	9-bp INDEL

genes. As the genetic system genes, two conserved open reading frames (*ycf1* and *ycf2*), maturase K gene (*matK*), 3 RNA polymerase genes (*rpoA*, *rpoB* and *rpoC2*) and ribosomal protein large subunit gene (*rpl22*) are included. Among photosynthesis genes, six genes of NADH dehydrogenase (*ndhA*, *ndhD*, *ndhF*, *ndhF*, *ndhH*, *ndhI* and *ndhJ*), three genes of ATP synthase (*atpB*, *atpF* and *atpI*), two photosystem I genes (*psaA* and *psaB*), two photosystem II genes (*psbA* and *psbD*), a rubisco gene (*rbcL*), a membrane protein gene (*cemA*), cytochrome C biogenesis gene (*ccsA*), and a Acetyl-CoA carboxylase gene (*accD*) were variable. In addition to SNP, 9-bp INDEL was found in *ycf1* gene containing 19 SNPs.

Conclusion

This is the first report of chloroplast genome in *Dendranthema*, Asteraceae. As genetic barcode of *Dendranthema* sp. K247003, a possible new species, 151,175-bp of chloroplast genomic sequence was registered to NABIC (NG-0482-000001). The chloroplast genome is distinguished from that of *D. boreale* IT121002, by 67 SNPs and an INDEL in coding regions, in addition to 97 variable IGS sites. As suggested by Dong *et al.* (2015) in land plants, *ycf1* would be useful for plant identification as having 19 SNPs and an INDEL in the comparison of *Dendranthema* sp. K247003 and *D. boreale* IT121002. In addition, as suggested by Li *et al.* (2015), we show that chloroplast genomic information is useful for genetic barcode in *Dendranthema*. Further characterization of organellar genomes using NGS data would facilitate our phylogenomic study and molecular marker developments in *Dendranthema* at low cost. Finally, further morphological and cytological studies on *Dendranthema* sp. K247003 are remained for the taxonomic treatment of this taxon.

Acknowledgements

This work was supported by a grant from the National Agricultural Genome Center (project No. PJ0104562015), Rural Development Administration, Republic of Korea.

References

Besendahl, A., Y.L. Qiu, J. Lee, J.D. Palmer and D. Bhattacharya. 2000. The endosymbiotic origin and vertical evolution of the plastid tRNA Leu group I intron. Curr. Genet. 37:12-23.

Bock, D.G., N.C. Kane, D.P. Ebert and L.H. Rieseberg. 2014. Genome skimming reveals the origin of the Jerusalem Artichoke tuber crop species: neither from Jerusalem nor an artichoke. New Phytol. 201(3):1021-1030.

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

Dempewolf, H., N.C. Kane, K.L. Ostevik, M. Geleta, M.S. Barker, Z. Lai, M.L. Stewart, E. Bekele, J.M.M. Engels, Q. Cronk and L.H. Rieseberrg. 2010. Establishing genomic tools and resources for *Guizotia abyssinica* (L.f.) Cass.-the development of a library of expressed sequence tags, microsatellite loci, and the sequencing of its chloroplast genome. Mol. Ecol. Resour. 10(6):1048-1058.

Dong, W., C. Xu, C. Li, J. Sun, Y. Zuo, S. Shi, T. Chang, J. Guo and S. Zhou. 2015. ycf1, the most promising plastid DNA barcode of land plants. Sientific Report 5:8348.

Doorduin, L., B. Gravendeel, Y. Lammers, Y. Ariyurek, T. Chin-A-Woeng and K. Vrieling. 2011. The Complete Chloroplast Genome of 17 Individuals of Pest Species *Jacobaea vulgaris*: SNPs, Microsatellites and Barcoding Markers for Population and Phylogenetic Studies. DNA Res. 18(2):93-105.

Fraser, K.A., L. Pachter, A. Poliakov, E.M. Rubin and I. Dubchak. 2004. VISTA: computational tools for comparative genomics. Nucleic Acids Research 32:W273-W279.

Hwang, Y.-J., A. Younis, K.B. Ryu, K.-B. Lim, C.-H. Eun, J. Lee, S.-H. Sohn and S.-J. Kwon. 2014. Karyomorphological Analysis of Wild *Chrysanthemum boreale* Collected from Four Natural Habitats in Korea. Flower Res. J. 21(4):182-189.

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 15:286-299.

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.

Lee, J. 1997. Gene clusters and introns of *Spirogyra maxima* chloroplast and other charophytes and their phylogenetic implications in green plants. Texas A&M University Ph.D. Thesis.

Lee, J. and J.R. Manhart. 2002a. Four embryophyte introns and *psbB* operon explains *Chlorokybus* as a basal lineage of streptophytes. Algae 17(1):53-58.

Lee, J. and J.R. Manhart. 2002b. The Chloroplast *rpl23* gene cluster of *Spirogyra maxima* (Charophyceae) shares many similarities with the angiosperm *rpl23* operon. Algae 17(1):59-68.

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, P.-L., Q. Wan, Y.-P. Guo, J. Yang and G.-Y. Rao. 2012. Phylogeny of the Genus *Chrysanthemum* L.: Evidence from Single-Copy Nuclear Gene and Chloroplast DNA Sequences. PLoS One 7(11):e48970.

Liu, Y., N. Huo, L. Dong, Y. Wang, S. Zhang, H.A. Young, X. Feng and Y.Q. Gu. 2013. Complete Chloroplast Genome Sequences of Mongolia Medicine *Artemisia frigida* and Phylogenetic Relationships with Other Plants. PLoS One 8(2):e57533.

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.

Nie, X., S. Lv, Y. Zhang, X. Du, L. Wang, S.S. Biradar, X. Tan, F. Wan and S. Weining. 2012. Complete chloroplast genome sequence of a major invasive species, crofton weed (*Ageratina adenophora*). PLoS One 7(5):e36869.

Nock, C.J., D.L. Waters, M.A. Edward, S.G. Bowen, N. Rice, G.M. Cordeiro and R.J. Henry. 2011. Chloroplast genome sequences from total DNA for plant identification. Plant Biotechnol. J. 9(3):328-333.

Park, C.-W. 2007. The Genera of Vascular Plants of Korea. Academy Publishing Co., Seoul.

Park, J.-H. 2007. *Dendranthema* (DC.) Des Moul., Actes. Linn. Bordeaux 20: 561, 1855. In: C.-W. Park (ed.), The Genera of Vascular Plants of Korea. Academy Publishing Co., Seoul. pp. 1024-1026.

Shaw, J., B.L. Edgar, E.S. Edward and L.S. Randall. 2007. Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: The tortoise and the hare III. American Journal of Botany 94(3):275-288.

Timme, R.E., V.K. Jennifer, L.B. Jefferey and K.J. Robert. 2007. A comparative analysis of the *Lactuca* and *Helianthus* (Asteraceae) plastid genomes: Identification of divergent regions and categorization of shared repeats. American Journal of Botany 94(3):302-312.

Walker, J.F., M.J. Zanis and N.C. Emery. 2014. Comparative analysis of complete chloroplast genome sequence and inversion variation in *Lasthenia burkei* (Madieae, Asteraceae). Am. J. Bot. 101(4):722-729.

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

Submitted: July 1, 2015
Revised: July 24, 2015
Accepted: July 27, 2015