



Comparative analyses of *Agrimonia* complete chloroplast genomes with newly assembled chloroplast genomes of *A. coreana* and *A. nipponica* (Rosaceae)

Jongsun PARK^{1,2*}, Hwa-Jung SUH², and Hong XI¹

¹Infoboss Inc. and Infoboss Research Center, Seoul 07766, Korea

²Department of Biology, Daejeon University, Daejeon 34520, Korea

Corresponding author

Jongsun PARK

E-mail: starflr@infoboss.co.kr

ABSTRACT: The genus *Agrimonia* L. (1753) is a small group consisting of 19 species and three varieties. We completed two *Agrimonia* chloroplast genomes, *Agrimonia coreana* Nakai and *Agrimonia nipponica* Koidz, collected in Korea and conducted comprehensive analyses of six *Agrimonia* chloroplast genomes. Two of the *Agrimonia* chloroplast genomes have a typical quadripartite structure with lengths ranging from 155,161 bp to 151,362 bp, similar to the remaining *Agrimonia* chloroplast genomes. High nucleotide diversity was detected in the *yefl* gene, which can serve as a molecular marker. Intraspecific variations of four *A. pilosa* chloroplast genomes and interspecific variations of three *Agrimonia* species revealed that KY419942 is distinct to the other three *A. pilosa* chloroplast genomes. A phylogenetic analysis demonstrated that KY419942 was clustered with *A. nipponica*, requiring further analyses to understand *Agrimonia* species in East Asia. A comparison of simple sequence repeats identified from the six *Agrimonia* chloroplast genomes suggests potential molecular markers to distinguish species or populations of the same species. Our results define the phylogenetic relationship of three *Agrimonia* species and provide insight into the intraspecific features of *Agrimonia* chloroplast genomes with which to gain a better understanding of the genetic structure of *Agrimonia* species.

KEYWORDS: interspecific variations, intraspecific variations, nucleotide diversity, phylogenetic analysis, Rosaceae, simple sequence repeats

RECEIVED 28 February 2023; **REVISED** 27 March 2024; **ACCEPTED** 28 March 2024

OPEN ACCESS

© 2024 The Author(s)

This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

INTRODUCTION

The subtribe Agrimoniinae (Rosaceae, Rosoideae), one of two members of the tribe Sanguisorbeae, displays polyploidy and morphological variations, both of which are good sources of information when attempting to understand the corresponding biogeographical patterns and endemism (Chung, 2008). This subtribe consists of the five genera, *Agrimonia* (19 species in Eurasia and South America), *Hagenia* (one species in east Africa), *Leucosidea* (one species in southern Africa), *Spenceria* (one species in central Asia), and *Aremonia* (one species in southern Europe) (Potter et al., 2007; Chung 2008), with all genera except *Agrimonia* being monotypic. Given this biased distribution of species, partial

or complete chloroplast genomes for four out of the five genera except *Aremonia* have been determined thus far: two complete and one partial determination for *Agrimonia pilosa* Ledeb. (Zhang et al., 2017; Heo et al., 2020; Liu et al., 2020b), one complete determination for *Agrimonia pilosa* var. *nepalensis* (D. Don) Nakai (Yang et al., 2021), one complete and one partial determination for *Hagenia abyssinica* Willd (Gichira et al., 2017; Zhang et al., 2017), one partial determination for *Leucosidea sericea* Eckl. & Zeyh. (Zhang et al., 2017), and one partial determination for *Spenceria ramalana* Trimen (Zhang et al., 2017).

Agrimonia L. (1753), consisting of 19 species and three varieties, is characterized by two major morphological characteristics: interrupted pinnately compound leaves and

bristly indumentum, within Rosaceae (Chung, 2008). *Agrimonia* species are usually distributed in temperate regions, including North America, Central America, the West Indies, southern South America, Europe, Asia, and southern Africa (Hutchinson, 1964; Skalický, 1973; Robertson, 1974; Murata and Umemoto, 1983; Chung and Kim, 2000; Kalkman, 2004; Chung, 2008). The base chromosome number is $x = 7$ (Darlington and Wylie, 1955) and the polyploidy levels of *Agrimonia* species range from tetraploid ($2n = 4x = 28$) to octaploid ($2n = 8x = 56$) (Chung, 2008). Various natural hybrids observed in *Agrimonia* hinder a clear delimitation of its species boundary (Wallroth, 1842; Skalický, 1962; Chung, 2008), indicating that alternative methods, such as the use of molecular markers, are required to understand the species boundary.

The *Agrimonia pilosa* complex is an excellent case for investigating molecular variation at the genomic level. The *A. pilosa* complex, with three morphologically similar species *A. pilosa*, *A. coreana* Nakai, and *A. nipponica* Koidz, is characterized by having relatively small fruits with spreading, erect, or connivent prickles (Chung and Kim, 2000; Chung, 2008). All of these species are distributed in East Asia with extensions to Eastern Europe. An examination of the intraspecific variations of chloroplast genomes together with their morphological characteristics would provide fundamental information about the levels and patterns of variations within and among closely related species, especially considering that data pertaining to four chloroplast genomes of *A. pilosa* are available in China and Korea (Zhang et al., 2017; Heo et al., 2020; Liu et al., 2020b). In addition, plants of the *A. pilosa* complex have been used to treat boils, eczema, taeniasis, abdominal pain, sore throat, headaches, and heat stroke in traditional medicine in China, Nepal, and Korea (World Health Organization, 1998). Careful and thorough investigations have also found that *A. pilosa* extracts have various useful effects, among them antiviral (Shin et al., 2010), whitening (Kim et al., 2011a; Kim et al., 2011b), antinociceptive (Park et al., 2012), antioxidant (Seo et al., 2008; Kim et al., 2011b; Chen and Kang, 2014), anticancer (Seo et al., 2008), anti-aging (Yoon et al., 2012), and anti-inflammatory (Jung et al., 2010) effects, suggesting the potential value of investigating the genetic background of *Agrimonia* species in Korea.

Owing to next-generation sequencing and third-generation sequencing technologies (Zhou et al., 2010; Roberts et al., 2013; Van Dijk et al., 2014; Deamer et al., 2016; Goodwin et al., 2016), nearly ten thousand chloroplast genomes have been sequenced thus far (Park et al., 2021d). These chloroplast genomes have been used for phylogenetic or phylogenomic analyses (Hassemer et al., 2019; Alzahrani et al., 2020; Liang

et al., 2020; Chang et al., 2021), to identify the phylogenetic positions of new species candidates (Kim et al., 2019f; Oh et al., 2019a; Park et al., 2021b), to investigate intraspecific variations on chloroplast genomes (Li et al., 2018; Bum et al., 2020; Baek et al., 2021), and to develop useful molecular markers (Li et al., 2020a; Wang et al., 2020). Complete chloroplast genomes can be utilized to understand the relationships between morphological features and their phylogenetic positions. Examples of this possibility include *Chenopodium album* L. (Park et al., 2021d), *Suaeda japonica* Makino (Kim et al., 2020), the *Allium* genus (Xie et al., 2019), and the *Fagopyrum* genus (Wang et al., 2017). Due to the morphological complexity and the presence of intermediate forms in the *A. pilosa* complex, additional chloroplast genomes of *A. coreana*, and *A. nipponica* beyond the three chloroplast genomes of *A. pilosa* already reported (Zhang et al., 2017; Heo et al., 2020; Liu et al., 2020b) can be helpful to understand the relationships, morphological diversity and phylogenetic positions of this complex.

Here, we complete the two chloroplast genomes of *A. coreana* and *A. nipponica* and conduct comparative analyses of the six *Agrimonia* chloroplast genomes available to date. Our comprehensive analyses of nucleotide diversity, intraspecific and interspecific variations, and phylogeneticity in these genomes and comparisons of simple sequence repeats (SSRs) demonstrate that two groups of *A. pilosa*, one a sister to *A. nipponica* and the other a sister to *A. coreana*, reflect that *A. coreana*, *A. nipponica*, and *A. pilosa* have complex evolutionary histories. Also presented are the chromosome configurations of the three species. Further analyses of additional chloroplast genomes and the continuing whole genome sequencing of other *Agrimonia* species will be required to understand the precise evolutionary history of the *Agrimonia* genus.

MATERIALS AND METHODS

DNA extraction of two *Agrimonia* samples isolated in Korea

Two *Agrimonia* species, *A. coreana* and *A. nipponica*, were collected on the Korean Peninsula (Table 1). Both are native Korean plant species (Park et al., 2020a) collectable without special permission in Korea. The sampling process for *Agrimonia* was conducted while also adhering to all local, national, and international guidelines and laws. All vouchers of the two *Agrimonia* samples were deposited into the InfoBoss Cyber Herbarium (IN) (Table 1). Their total DNA was extracted from fresh leaves of the two samples using a DNeasy Plant Mini Kit (QIAGEN, Hilden, Germany).

Table 1. List of six *Agrimonia* chloroplast genomes analyzed in this study.

Species	Voucher	GenBank accession	Length (bp)			GC contents				No. of genes		Ref.			
			Whole	LSC	SSC	IR	Gap	Whole (%)	LSC (%)	SSC (%)	IR (%)		No. of PCGs	No. of tRNAs	
<i>A. coreana</i>	IB-01066	MZ604439	155,362	84,597	18,807	25,979	0	36.9	34.8	30.3	42.6	84	37	8	This study
<i>A. nipponica</i>	IB-01069	MZ604440	155,161	84,514	19,825	25,411	0	36.9	34.8	30.7	42.8	84	37	8	This study
<i>A. pilosa</i>	IB-01030	MT415946 MT415946	155,125	84,458	18,737	25,965	0	36.9	34.9	30.4	42.6	84	37	8	Heo et al. (2020)
<i>A. pilosa</i>	N/A	MT040192	155,188	84,521	18,737	25,965	5	36.9	34.9	30.4	42.6	84	37	8	Liu et al. (2020b)
<i>A. pilosa</i> ^a	N/A	KY419942	129,106	84,399	18,743	25,964	0	35.8	34.9	30.4	42.6	78	30	4	Zhang et al. (2017)
<i>A. pilosa</i> var. <i>nepalensis</i>	2020Jul SH08	MW387437	155,147	84,480	18,737	25,965	0	36.9	34.8	30.4	42.6	85	37	8	Yang et al. (2021)

LSC, large single-copy; SSC, small single-copy; IR, inverted repeat; PCG, protein-coding gene; N/A, not available.

^aPartial chloroplast genome with only one IR region.

Genome sequencing and de novo assembly of the two *Agrimonia* chloroplast genomes

Genome sequencing was performed with the NovaSeq 6000 system at Macrogen Inc. in Korea using the DNA extracted from the two *Agrimonia* species. *De novo* assembly was then done, with confirmation accomplished with Velvet v1.2.10 (Zerbino and Birney, 2008) after filtering raw reads using Trimmomatic v0.33 (Bolger et al., 2014). After obtaining the first draft of the chloroplast genome sequences, gaps were filled with GapCloser v1.12 (Zhao et al., 2011) and all bases from the assembled sequences were confirmed by checking each base in the alignment (tview mode in SAMtools v1.9) (Li et al., 2009) against the assembled chloroplast genome generated with BWA v0.7.17 (Li, 2013). All of these processes were conducted under the environment of Genome Information System (GeIS; <http://geis.infoboss.co.kr/>), which has been utilized in a range of studies (Kim et al., 2021b, 2021c; Park et al., 2021c, 2021h), including work on plant organelle genomes (Suh et al., 2021; Park et al., 2022a, 2022c; Yoo et al., 2023).

Chloroplast genome annotation

Geneious Prime 2020.2.4 (Biomatters Ltd., Auckland, New Zealand) was used for genome annotation of the two *Agrimonia* species based on the *A. pilosa* chloroplast genome (GenBank accession number: MT415946) (Heo et al., 2020) by transferring annotations while correcting exceptional cases, including missing start or stop codons. tRNA was predicted and confirmed based on predictions by tRNAscan-SE v2.0.6 (Lowe and Chan, 2016). Circular maps of the two *Agrimonia* chloroplasts were drawn using the OGDRAW v1.3.1 (Greiner et al., 2019).

Identification of sequence variations from complete chloroplast genomes

Single nucleotide polymorphisms (SNPs) and insertions and deletions (INDELs) were identified by means of a pair-wise alignment of the two selected chloroplast genomes done using MAFFT v7.450 (Kato and Standley, 2013). When the numbers of INDELs were calculated, continuous INDEL bases were considered as one INDEL. When partial chloroplast genomes without one of the inverted repeat (IR) regions were used to identify sequence variations, SNPs and INDELs located in the IR regions were counted once more with the assumption that the IR region missing in the partial chloroplast genome is identical to the IR region in the partial chloroplast genomes. These methods have been utilized in previous studies (Kim et al., 2021a, 2021e; Park et al., 2021g).

Identification of SSRs

SSRs were identified on the chloroplast genome sequence

using the pipeline of the SSR database (SSRDB; <http://ssrdb.infoboss.co.kr/>; Park et al., in preparation), as utilized in previous studies (Hong et al., 2022). Based on the conventional definition of an SSR on a chloroplast genome, monoSSR (1 bp) to hexaSSR (6 bp), the total length of the SSRs on the chloroplast genome exceeds 10 bp. Owing to the different criteria pertaining to SSRs on chloroplast genomes, we adopted the criteria used with the organelle genomes of *Dysphania* species (Kim et al., 2019d), *Arabidopsis thaliana* L. (Park et al., 2020c), *Chenopodium album* (Park et al., 2021d), *Diarthron linifolium* Turcz. (Kim et al., 2021d), *Campsis grandiflora* (Park and Xi 2022), and *Rosa rugosa* Thunb. (Park et al., 2020d), as follows; the monoSSR (unit sequence length of 1 bp) to hexaSSR (6 bp) are used as normal SSRs, and the heptaSSR (7 bp) to decaSSR (10 bp) are defined as the extended SSRs. Among the normal SSRs, pentaSSRs, and hexaSSRs with a unit sequence repeat number of 2 are classified as potential SSRs. Regional classifications of the chloroplast genome were conducted using the method described in the section above.

Comparison of SSRs identified from the six *Agrimonia* chloroplast genomes

SSRs identified from ten *Agrimonia* chloroplast genomes were compared based on their flanking sequences under the SSRDB environment (<http://ssrdb.infoboss.co.kr/>) (Park et al., in preparation). The pipeline of the SSR comparison implemented in the SSRDB relies on various organelle genome studies (Park et al., 2021i, 2022b; Kim et al., 2023) and was used with the following conditions: a cut-off e-value of $1e-10$ and a maximum flanking sequence for the comparison of 60 bp.

Nucleotide diversity analysis

Nucleotide diversity was calculated using the method proposed by Nei and Li (Nei and Li, 1979) based on the multiple sequence alignment of *Agrimonia* chloroplast genomes using a Perl script used in previous studies (Lee et al., 2020; Choi et al., 2021). The window size was set to 500 bp and the step size was 200 bp when using the sliding-window method. The genomic coordination of each window was compared to the gene annotation of the chloroplast genome under the GenomeArchive (<http://www.genomearchive.net/>) environment for further analyses.

Construction of phylogenetic trees

Thirty-three whole chloroplast genomes of tribe Sanguisorbeae and two outgroup species, *Rosa* and *Potentilla* (Table 2), were aligned by MAFFT v7.450 (Katoh and Standley, 2013) and the alignment quality was checked manually. The maximum

likelihood (ML) tree was reconstructed in IQ-TREE v1.6.6 (Nguyen et al., 2015). In the ML analysis, a heuristic search was used with nearest-neighbor interchange branch swapping, the GTR + F + R4 model, and uniform rates among sites. All other options used default settings. Bootstrap analyses with 1,000 pseudoreplicates were conducted with the same options. The posterior probability of each node was estimated by means of Bayesian inference (BI) using the MrBayes v3.2.6 (Huelsenbeck and Ronquist, 2001) plug-in implemented in Geneious Prime 2020.2.4 (Biomatters Ltd.). The HKY85 model with gamma rates was used as a molecular model. A Markov-chain Monte Carlo (MCMC) algorithm was employed for 1,100,000 generations, sampling trees every 200 generations, with four chains running simultaneously. Trees from the first 100,000 generations were discarded as burn-in.

RESULTS AND DISCUSSION

Two complete *Agrimonia* chloroplast genomes, *A. coreana* and *A. nipponica*

Two *Agrimonia* species, *A. coreana* and *A. nipponica*, were sampled on the Korean Peninsula (Table 1) (Chung 2017; Park et al., 2020e). Chloroplast genomes of the two *Agrimonia* species were successfully completed, with corresponding coverage rates of 133.88x and 150.31x. It was thus found that the *A. coreana* chloroplast genome was 151,362 bp long (Fig. 1A), the longest among the six *Agrimonia* chloroplast genomes, while that of *A. nipponica* was 155,161 bp long (Fig. 1B), similar to the other *Agrimonia* chloroplast genomes (Table 1). A small single-copy (SSC) region of *A. nipponica* was the longest (19,825 bp) among the six *Agrimonia* chloroplast genomes, while the IR region of *A. nipponica* was the shortest (25,411 bp). A large single-copy (LSC) region of *A. coreana* was longest (84,597 bp) (Table 1). The two *Agrimonia* chloroplast genome sequences sequenced in this study can be accessed via accession numbers MZ604439 and MZ604440 in the NCBI GenBank.

The overall GC contents of *A. coreana* and *A. nipponica* amounted to 36.9%, identical to the remaining *Agrimonia* chloroplast genomes, except for the partial chloroplast genome (GenBank accession number: KY419942) (Table 1). The GC contents of the LSC in both *A. coreana* and *A. nipponica* amounted to 34.8%, whereas the corresponding SSC and IR rates of *A. coreana* were 30.3% and 42.6%, lower than those of *A. nipponica* (Table 1), caused by different range of the IR region.

Both of the *Agrimonia* chloroplast genomes contained 130 genes, consisting of 84 protein-coding genes (PCGs), eight ribosomal RNAs (rRNAs), 37 transfer RNAs (tRNAs), and one pseudogene (*ycf1*) (Online Supplementary Material S1).

Table 2. List of intraspecific variations identified from various plant species isolated from Korea and China using pair-wise comparison method.

Family	Species name	Source	Target	No. of SNPs	Proportion of SNPs	No. of INDEL regions	INDEL length (bp)	Proportion of INDELs	Alignment length (bp)	Reference
Selaginellaceae	<i>Selaginella tamariscina</i>	Korean isolate (MN894555)	Chinese isolate (NC_041646)	1,223	0.009603	376	1,635	0.012839	127,350	Xu et al. (2018), Park et al. (2020b)
Rosaceae	<i>Hagenia abyssinica</i>	Chinese isolate (KX008604)	Chinese isolate (KY420026)	82	0.000529	28	262	0.001689	155,127	Gichira et al. (2017), Zhang et al. (2017)
		<i>Sanguisorba officinalis</i>	Chinese isolate (NC_044694)	Korean isolate (MZ145058)	85	0.000546	61	301	0.001934	155,596
	Chinese isolate (NC_044694)		Korean isolate (MK696193)	75	0.000482	60	290	0.001864	155,581	
	Korean isolate (MZ145058)		Korean isolate (MK696193)	10	0.000064	17	105	0.000675	155,455	
	<i>Sanguisorba tenuifolia</i>	Chinese isolate (NC_044692)	Chinese isolate (NC_042223)	27	0.000174	32	138	0.000887	155,499	Meng et al., (2018), Park et al. (2018), Unpublished
		Chinese isolate (NC_044692)	Korean isolate (MK696194)	91	0.000585	64	338	0.002172	155,605	
		Chinese isolate (NC_042223)	Korean isolate (MK696194)	89	0.000572	66	344	0.002211	155,581	
	<i>Sanguisorba stipulata</i>	Korean isolate (MZ145059)	Korean isolate (MK696195)	78	0.000502	77	367	0.002360	155,540	Unpublished
	<i>Sanguisorba filiformis</i>	Chinese isolate (NC_044693)	Chinese isolate (KY419920)	45	0.000291	12	478	0.003090	154,707	Zhang et al. (2017), Meng et al. (2018)
	Amaranthaceae	<i>Chenopodium album</i>	CAGAP004 (MW446246)	CAGOH01 (MW446245)	28	0.000184	12	19	0.000125	152,207
CAGAP004 (MW446246)			CAJEJG05 (MW446243)	2	0.000013	2	2	0.000013	152,199	
CAGAP004 (MW446246)			CCANG01 (MW446241)	27	0.000177	17	315	0.002070	152,210	
CAGAP004 (MW446246)			CSJUK01 (MW446242)	21	0.000138	12	18	0.000118	152,207	
CAGAP004 (MW446246)			CVHUP01 (MW446244)	14	0.000092	13	33	0.000217	152,291	
CAGAP004 (MW446246)			NC_034950	17	0.000112	8	36	0.000237	152,201	
CAGOH01 (MW446245)			CAJEJG05 (MW446243)	30	0.000197	9	35	0.000230	152,207	
CAGOH01 (MW446245)			CCANG01 (MW446241)	0	0.000000	7	298	0.001958	152,200	
CAGOH01 (MW446245)			CSJUK01 (MW446242)	21	0.000138	7	15	0.000099	152,204	
CAGOH01 (MW446245)			CVHUP01 (MW446244)	30	0.000197	9	40	0.000263	152,293	

Table 2. Continued

Family	Species name	Source	Target	No. of SNPs	Proportion of SNPs	No. of INDEL regions	INDEL length (bp)	Proportion of INDELs	Alignment length (bp)	Reference
		CAGOH01 (MW446245)	NC_034950	33	0.000217	6	37	0.000243	152,200	
		CAJEJG05 (MW446243)	CCANG01 (MW446241)	29	0.000191	36	331	0.002175	152,210	
		CAJEJG05 (MW446243)	CSJUK01 (MW446242)	23	0.000151	14	34	0.000224	151,961	
		CAJEJG05 (MW446243)	CVHUP01 (MW446244)	16	0.000105	15	49	0.000322	152,045	
		CAJEJG05 (MW446243)	NC_034950	19	0.000125	10	52	0.000342	151,955	
		CCANG01 (MW446241)	CSJUK01 (MW446242)	21	0.000138	10	309	0.002030	152,206	
		CCANG01 (MW446241)	CVHUP01 (MW446244)	30	0.000197	17	336	0.002206	152,296	
		CCANG01 (MW446241)	NC_034950	33	0.000217	11	333	0.002188	152,203	
		CSJUK01 (MW446242)	CVHUP01 (MW446244)	23	0.000151	13	39	0.000256	152,293	
		CSJUK01 (MW446242)	NC_034950	26	0.000171	7	36	0.000237	152,200	
		CVHUP01 (MW446244)	NC_034950	18	0.000118	8	57	0.000374	152,287	
	<i>Dysphania pumilio</i>	Anyang city, Gyeonggi-do (MH936550)	Gangseo-gu, Seoul (MK541016)	25	0.000165	1	2	0.000013	151,962	Kim et al. (2019b), Park et al. (2019i)
	<i>Suaeda japonica</i>	Ganghwa type (MK764271)	Julpo type (MK558824)	3	0.000020	3	3	0.000020	152,112	Kim et al., (2019e), Kang et al. (2020)
Magnoliaceae	<i>Magnolia kobus</i>	Jejudo (NC_023237)	Jejudo (MN894553)	50	0.000314	28	63	0.000395	159,457	Unpublished
	<i>Liriodendron tulipifera</i>	Korean isolate (MK477550)	Chinese isolate (NC_008326)	12	0.000075	0	0	0.000000	159,886	Cai et al. (2006), Kwon et al. (2019d)
Theaceae	<i>Camellia japonica</i>	Soyeonpyeong-do (MK353210)	Seogwang-ri in Jejudo (MK353211)	25	0.000159	2	2	0.000013	156,976	Kim et al., (2017), Kwon et al. (2019c)
		Soyeonpyeong-do (MK353210)	Wimi-ri in Jejudo island (KU951523)	25	0.000159	2	2	0.000013	156,972	Li et al. (2019), Min et al. (2019c)
		Seogwang-ri in Jejudo (MK353211)	Wimi-ri in Jejudo (KU951523)	25	0.000159	0	0	0.000000	156,971	
		Cv. Huaheling (MW602996)	Soyeonpyeong-do (MK353210)	8	0.000051	7	36	0.000229	157,004	
		Cv. Huaheling (MW602996)	Seogwang-ri in Jejudo (MK353211)	8	0.000050	9	38	0.000242	157,005	

Table 2. Continued

Family	Species name	Source	Target	No. of SNPs	Proportion of SNPs	No. of INDEL regions	INDEL length (bp)	Proportion of INDELS	Alignment length (bp)	Reference
		Cv. Huaheling (MW602996)	Wimi-ri in Jejudo (KU951523)	33	0.000210	9	38	0.000242	157,005	
		S288C (NC_036830)	Soyeonpyeong-do (MK353210)	78	0.000496	60	644	0.004099	157,110	
		S288C (NC_036830)	Seogwang-ri in Jejudo (MK353211)	78	0.000496	61	645	0.004105	157,111	
		S288C (NC_036830)	Wimi-ri in Jejudo (KU951523)	103	0.000656	61	645	0.004105	157,111	
		S288C (NC_036830)	Cv. Huaheling (MW602996)	80	0.000509	61	671	0.004270	157,139	
Caryophyllaceae	<i>Pseudostellaria palibiniana</i>	Mt. Taebaeksan (MK120981)	Mt. Gwangdeoksan (MK309611)	84	0.000561	43	175	0.001169	149,741	Kim et al. (2019c), Park et al. (2019d)
Salicaceae	<i>Salix koriyanagi</i>	Male, Seoul (MK541017)	Female, Seoul (MK120982)	0	0.000000	0	0	0.000000	155,548	Kim et al. (2019a), Park et al. (2019e)
Rosaceae	<i>Pyrus ussuriensis</i>	Bonghwa-gun	Mt. Hambeaksan	121	0.000753	61	781	0.004859	160,727	Cho et al. (2019), Gil et al. (2019)
Fagaceae	<i>Fagus multinervis</i>	Ulleungdo (MN894556)	Ulleungdo (MK518070)	2	0.000013	2	2	0.000013	158,350	Park et al. (2019a), Park and Oh (2020), Yang et al. (2020)
		Ulleungdo (MN894556)	Ulleungdo (MT762296)	2	0.000013	3	3	0.000019	158,350	
		Ulleungdo (MK518070)	Ulleungdo (MT762296)	0	0.000000	0	0	0.000000	158,348	
Asteraceae	<i>Artemisia fukudo</i>	Aphaedo, Shinan-gun (NC_044156)	Jeungdo, Shinan-gun (MG951488)	7	0.000046	5	12	0.000079	151,838	Lee et al. (2016), Min et al. (2019d)
Orchidaceae	<i>Goodyera schlechtendaliana</i>	Korean isolate (MK144665)	Korean isolate (MK134679)	200	0.001294	59	511	0.003305	154,607	Niu et al. (2017), Oh et al. (2019a, 2019b)
		Chinese isolate (NC_029364)	Korean isolate (MK144665)	842	0.005434	130	1,779	0.011480	154,964	
		Chinese isolate (NC_029364)	Korean isolate (MK134679)	740	0.004780	95	1,470	0.009495	154,820	
		Chinese isolate (LC085346)	Chinese isolate (NC_029364)	514	0.003310	110	2,133	0.013733	155,322	
		Chinese isolate (LC085346)	Korean isolate (MK144665)	700	0.004526	112	1,366	0.008832	154,665	
		Chinese isolate (LC085346)	Korean isolate (MK134679)	597	0.003863	72	1,065	0.006892	154,525	
		Chinese isolate (AB893949)	Chinese isolate (LC085346)	445	0.002887	29	415	0.002693	154,119	
		Chinese isolate (AB893949)	Chinese isolate (NC_029364)	864	0.005573	123	2,045	0.131912	155,027	

Table 2. Continued

Family	Species name	Source	Target	No. of SNPs	Proportion of SNPs	No. of INDEL regions	INDEL length (bp)	Proportion of INDELS	Alignment length (bp)	Reference
Oleaceae	<i>Gastrodia elata</i>	Chinese isolate (AB893949)	Korean isolate (MK144665)	282	0.001828	86	1,060	0.006871	154,261	Yuan et al. (2018), Kang et al. (2020), Park et al. (2020b)
		Chinese isolate (AB893949)	Korean isolate (MK134679)	163	0.001058	43	652	0.004232	154,067	
		Korean isolate (MN026874)	Korean isolate (MN296709)	324	0.009119	76	630	0.017731	35,530	
		Chinese isolate (NC_037409)	Korean isolate (MN026874)	493	0.013851	79	651	0.018291	35,592	
	<i>Abeliophyllum distichum</i>	Chinese isolate (NC_037409)	Korean isolate (MN296709)	457	0.012842	75	671	0.018855	35,587	Min et al. (2019a), Min et al. (2019b), Park et al. (2019g), Park et al. (2021e)
		NC_031445	MN127986	93	0.000596	15	57	0.000365	156,023	
		NC_031445	MK616470	93	0.000596	19	64	0.000410	156,023	
		NC_031445	MF407183	93	0.000594	15	57	0.000364	156,567	
		NC_031445	MN116559	102	0.000654	19	64	0.000410	156,032	
		NC_031445	MW426545	99	0.000634	18	72	0.000461	156,034	
		MN127986	MK616470	0	0.000000	0	0	0.000000	156,008	
		MN127986	MF407183	0	0.000000	1	1	0.000006	156,553	
		MN127986	MN116559	9	0.000058	10	12	0.000077	156,019	
		MN127986	MW426545	6	0.000038	7	20	0.000128	156,021	
		MK616470	MF407183	0	0.000000	1	1	0.000006	156,553	
		MK616470	MN116559	9	0.000058	10	11	0.000071	156,019	
		MK616470	MW426545	6	0.000038	7	19	0.000122	156,021	
		MF407183	MN116559	9	0.000057	10	11	0.000070	156,564	
		MF407183	MW426545	6	0.000038	8	21	0.000134	156,566	
MN116559	MW426545	7	0.000045	9	23	0.000147	156,028			
Adoxaceae	<i>Viburnum erosum</i>	MN641480	MN218778	16	0.000101	8	50	0.000315	158,630	Park et al. (2019b), Choi et al. (2020)
Ranunculaceae	<i>Aconitum coreanum</i>	NC_031421	KU318669	29	0.000185	26	61	0.000388	157,075	Park et al. (2017), Kim et al. (2019f)
		NC_031421	Namhansan Fotress (MN400660)	5	0.000032	9	52	0.000331	157,052	
		Namhansan Fotress (MN400660)	(KU318669)	19	0.000121	25	92	0.000586	157,088	
Thymelaeaceae	<i>Daphne genkwa</i>	Jindo island (MT754180)	Buan	59	0.000444	90	404	0.003039	132,923	Yoo et al. (2021), Unpublished
		Jindo island (MT754180)	Chinese isolate (NC_045891)	69	0.000518	225	772	0.005798	133,156	

Table 2. Continued

Family	Species name	Source	Target	No. of SNPs	Proportion of SNPs	No. of INDEL regions	INDEL length (bp)	Proportion of INDELS	Alignment length (bp)	Reference
		Buan	Chinese isolate (NC_045891)	85	0.000639	100	543	0.004080	133,083	
Campanulaceae	<i>Campanula takesimana</i>	Dok-do island (MW013763)	Ulleung-do island (NC_026203)	33	0.000194	39	662	0.003900	169,737	Cheon et al. (2016), Park and Xi (2021)
Poaceae	<i>Zoysia matrella</i>	Korean isolate (MT983887)	Chiba Fair Green (AP014937)	28	0.000206	56	91	0.000670	135,894	Lee and Park (2021b)
	<i>Zoysia japonica</i>	Jeonju city (MW690657)	Unknown (NC_036827)	68	0.000500	24	50	0.000405	135,894	Lee and Park (2021a)
Asteraceae	<i>Erigeron canadensis</i>	Korean isolate (MT806101)	Chinese isolate (NC_046789)	103	0.000674	47	208	0.001360	152,926	Zhang et al. (2019b) Park et al. (in preparation)
Staphyleaceae	<i>Euscaphis japonica</i>	Chinese isolate (MN159078)	Korean isolate (NC_052922)	424	0.002635	152	809	0.005067	160,941	Xiang et al. (2019), Oh and Park (2020)

SNP, single nucleotide polymorphism; INDEL, insertion and deletion.

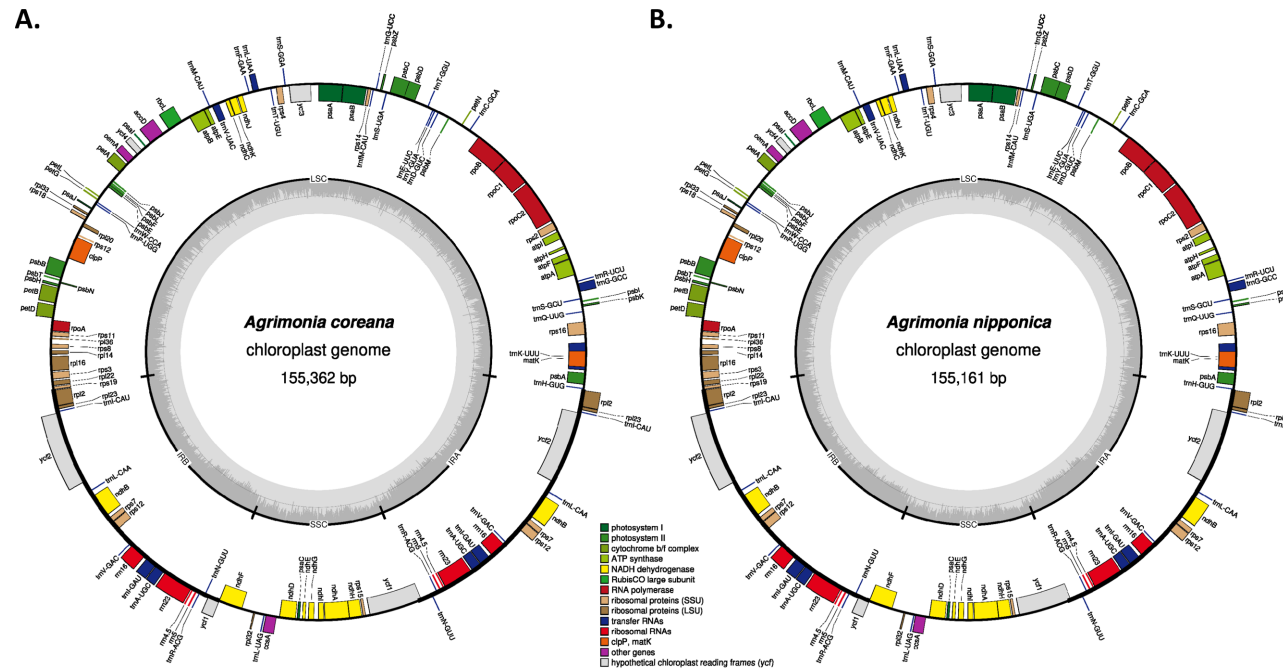


Fig. 1. Circular gene map of *Agrimonia coreana* and *Agrimonia nipponica* chloroplast genomes. **A.** *Agrimonia coreana*, **B.** *Agrimonia nipponica*. Genes shown outside are transcribed clockwise, and those inside the circle are transcribed counter-clockwise. Genes are color-coded to distinguish different functional groups. The dark grey and the light grey plots in the inner circle correspond to the GC content and AT content, respectively.

The 17 genes duplicated in IR regions consisted of seven PCGs (*rpl2*, *rpl23*, *ycf2*, *ndhB*, *rps7*, *rps12*, and *ycf1*), four rRNAs (*rrn16*, *rrn23*, *rrn4.5*, and *rrn5*), and seven tRNAs (*trnI-CAU*, *trnL-CAA*, *trnV-GAC*, *trnI-GAU*, *trnA-UGC*, *trnR-ACG*, and *trnN-GUU*). The *ycf1* gene in the IRb region was considered as a pseudogene because the putative start methionine was inside *ndhF*. However, *ycf1* in *A. pilosa* var. *nepalensis* was annotated as a normal gene by the start methionine located at 89-bp inside *ndhF*, addressing the need for additional research on the correct position of *ycf1* in the IR region. This caused one additional PCG in the *A. pilosa* var. *nepalensis* chloroplast genome (Table 1). In addition, the annotation of the three genes, *rpl16*, *petD*, and *petB*, in the *A. pilosa* var. *nepalensis* chloroplast indicated that they had no intron; however, those in the remaining *Agrimonia* chloroplast genomes consisted of two exons. The annotation of these genes should also be corrected for a more accurate analysis.

In the six *Agrimonia* chloroplast genomes, including the two *Agrimonia* chloroplast genomes sequenced in this study after correction of the corrected three genes in *A. pilosa* var. *nepalensis*, eight PCGs (*rps16*, *rpoC1*, *petB*, *petD*, *rpl16*, *ndhB*, *ndhA*, and *rpl2*) contain one intron and *clpP*, *ycf3*, and *rps12* have two introns, which are conserved across *Agrimonia* chloroplast genomes. The gene structure of the six *Agrimonia* chloroplast genomes is well conserved.

Nucleotide diversity of six *Agrimonia* chloroplast genomes

To investigate the landscape of nucleotide diversity in six *Agrimonia* chloroplast genomes, nucleotide diversity from the alignment of the six *Agrimonia* chloroplast genomes was assessed (see Materials and Methods). The LSC and SSC regions displayed a high level of nucleotide diversity, whereas two IR regions exhibited a low level, congruent to those of other chloroplast genomes (Du et al., 2017; Jiang et al., 2017; Li et al., 2018, 2019; Liu et al., 2020a; Loeuille et al., 2021). Most highly diverse regions were intergenic regions, and some regions contained tRNA genes, specifically *trnQ*, *trnS*, *trnT*, *trnS*, and *trnT* (Fig. 2), which is a reasonable observation in that most genes in chloroplast genomes are essential for photosynthesis and self-replication (Online Supplementary Material S1). One exceptional case is *ycf1*, displaying high nucleotide diversity throughout the coding genes in comparison to other PCGs (Online Supplementary Material S2). In detail, 29 non-synonymous SNPs, 15 synonymous SNPs, and 9-bp insertions were found in *ycf1* (Online Supplementary Material S2). The ratio of non-synonymous to synonymous SNPs (1.93) was high enough in comparison to those of all intraspecific variations of the PCGs of the *C. album* chloroplast genome (15:11 = 1.36), which is also considered as a high ratio in comparison to other plant species (Park et al., 2021d). In addition, *ycf1* displays high nucleotide diversity within the

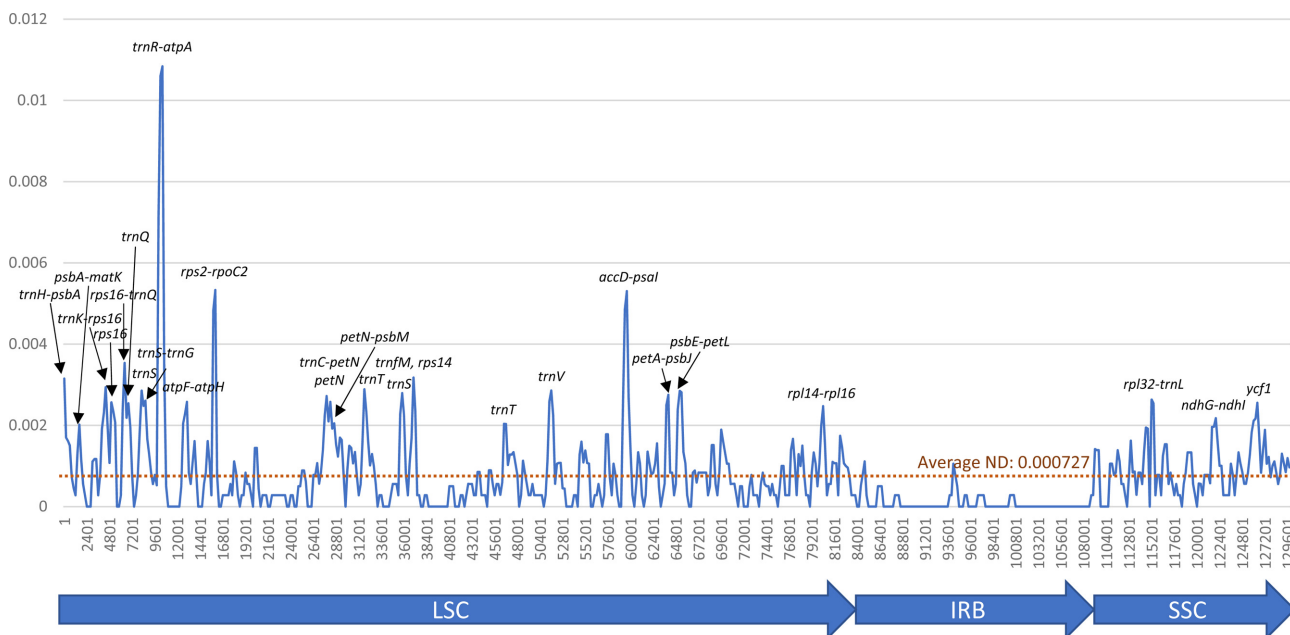


Fig. 2. Nucleotide diversity of the six *Agrimonia* chloroplast genomes. The X-axis represents the coordination of the chloroplast genome and the Y-axis represents the nucleotide diversity values from 500-bp window sliding in 200-bp steps. Highest peaks of the nucleotide diversity graph display the names of the genetic regions. The three blue arrow bars below the X-axis indicate the large single-copy (LSC), inverted repeat (IR), and small single-copy (SSC) regions, respectively.

genus (Hong et al., 2017; Liu et al., 2018; de Souza et al., 2019; Celiński et al., 2020; Loeuille et al., 2021) and even among certain species (Kim et al., 2020; Park et al., 2021d), which is also congruent with previous studies showing that *ycf1* is the most promising plastid DNA barcode in land plants (Neubig et al., 2009; Dong et al., 2015). Interestingly, *ycf1* is reportedly a target gene of PBR1 in *Arabidopsis* (Yang et al., 2016) as well as an essential part of the chloroplast protein import machinery (Kikuchi et al., 2013; de Vries et al., 2015). Moreover, *ycf1* and *ycf2* in the tobacco chloroplast genome reportedly have an essential function according to a transformation experiment (Drescher et al., 2000). Taken together, the selection pressure imposed on *ycf1* may not be high enough in comparison to other PCGs in the chloroplast genome and appeared to be distributed evenly on the gene region (Online Supplementary Material S2), also addressing the need for further investigations of *ycf1* through land plant species.

Intraspecific variations identified from the four complete chloroplast genomes of *Agrimonia pilosa*

The intraspecific variations of the four *A. pilosa* complete chloroplast genomes were investigated based on pair-wise comparisons, as was done in previous studies (Kwon et al., 2019a; Park et al., 2019f, 2019h, 2021f; Oh et al., 2021), as this method offers advantage of estimating the degree of intraspecific variation. SNPs and INDELs identified from the four *A. pilosa* chloroplast genomes ranged from 22 to 280

and 16 to 105 (86 bp to 633 bp in length) (Fig. 3A). It should be noted that the partial chloroplast genome KY419942 (Zhang et al., 2017) exhibited a high degree of intraspecific variation against the remaining three chloroplast genomes, including different varieties of *A. pilosa* and *A. pilosa* var. *nepalensis* (Fig. 3B). This phenomenon was also found in two chloroplast genomes of *Duchesnea indica* (Andrews) Th. Wolf (Kim et al., 2020), among which one is also from the same study (Zhang et al., 2017). For a comparison of the degrees of intraspecific variation in other plant species calculated from various studies which investigated intraspecific variation in plant samples isolated in Korea and China (Table 2), the proportions of SNPs and INDELs against the length of the alignment sequences were calculated. This shows that the ratios of the SNPs range from 0.000226 to 0.001801 and those of the INDELs are from 0.000522 to 0.004073. The degrees of SNPs and INDELs for *A. pilosa* are separated by two groups (blue and orange arrows in Fig. 3B) due to the numerous variations found in KY419942 chloroplast genome (Fig. 3A).

To evaluate the degrees of SNPs and INDELs identified among available *Agrimonia* chloroplast genomes, the degrees of intraspecific variation from five species in the tribe Sanguisorbeae which contains more than one chloroplast genome were also identified (Table 3). Based on the geographical locations of the species in Sanguisorbeae, including *Agrimonia* species, three major groups were defined: (1) one within Chinese isolates, (2) one between Korean and Chinese isolates, and (3) one within Korean isolates. The proportions of SNPs and INDELs identified from all three groups were distributed

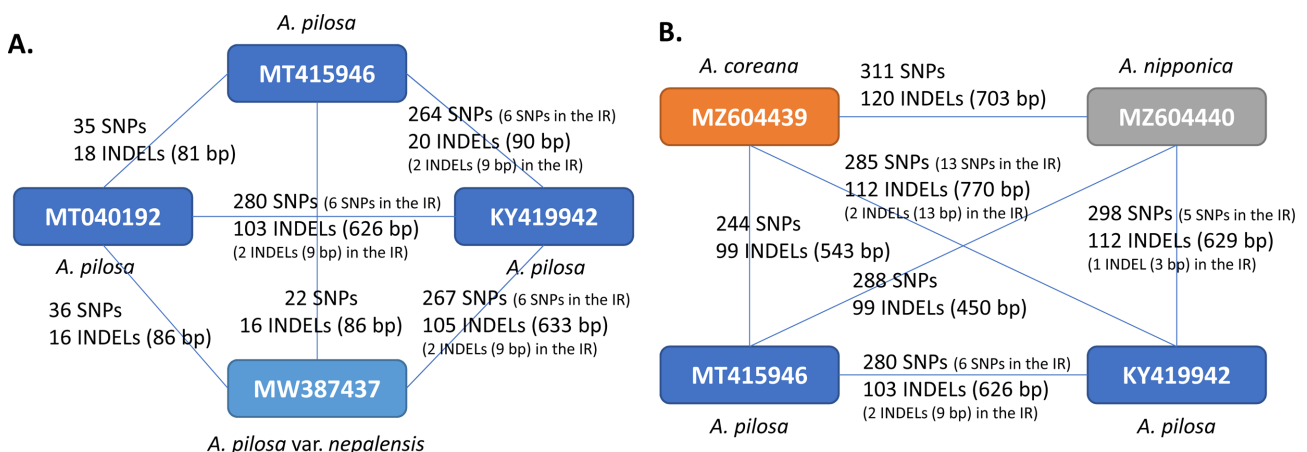


Fig. 3. Plot of intraspecific variations identified from the chloroplast genome of various plant species: (A) display of the plot on which the X-axis denotes the proportion of single nucleotide polymorphisms (SNPs) and the Y-axis is the proportion of insertions and deletions (INDELs) identified from various plant species. Three color-filled circles, shown in blue, orange, and dark blue, represent the Sanguisorbeae species, between Chinese isolates, between Korean and Chinese isolates, and between Korean isolates, respectively. Circles filled in with white represent the four groups, other species between Chinese isolates, those between Chinese isolates, those between Korean and Chinese isolates, those between Korean isolates, those between Korean and unknown isolates. B. Enlarged graph from the dotted lines in the graph in (A). Blue arrows and yellow arrows indicate the proportions of the SNPs and INDELs identified from *A. pilosa* chloroplast genomes.

Table 3. List of chloroplast genomes used for comparative analyses in this study.

Tribe	Subtribe	Species name	NCBI accession	Total length (bp)	No. of PCGs	No. of tRNAs	No. of rRNAs	GC ratio (%)	Reference
Sanguisorbeae	Agrimoniinae	<i>Agrimonia pilosa</i>	MT415496	155,125	84	37	8	34.9	Heo et al. (2020)
		<i>Agrimonia pilosa</i>	MT040192	155,188	84	37	8	34.9	Liu et al. (2020b)
		<i>Agrimonia pilosa</i> ^a	KY419942	129,106	78	30	4	34.9	Zhang et al. (2017)
		<i>Agrimonia pilosa</i> var. <i>nepalensis</i>	MW387437	155,147	84 ^b	37	8	34.8	Yang et al. (2021)
		<i>Agrimonia coreana</i>	MZ604439	155,362	84	37	8	36.9	This study
		<i>Agrimonia nipponica</i>	MZ604440	155,161	84	37	8	36.9	This study
		<i>Hagenia abyssinica</i> ^a	KY420026	129,090	78	30	4	36.0	Zhang et al. (2017)
		<i>Hagenia abyssinica</i>	KX008604	154,961	85	37	8	37.1	Gichira et al. (2017)
		<i>Leucosidea sericea</i> ^a	KY419929	128,924	78	30	4	35.9	Zhang et al. (2017)
		<i>Spenceria ramalana</i> ^a	KY419995	128,291	77	30	4	36.0	Zhang et al. (2017)
	Sanguisorbinae	<i>Sanguisorba officinalis</i>	NC_044694	155,749	84	37	8	37.2	Meng et al. (2018)
		<i>Sanguisorba officinalis</i>	MZ145058	155,412	85	37	8	37.2	Unpublished
		<i>Sanguisorba officinalis</i>	MK696193	155,393	N/A	N/A	N/A	37.2	Unpublished
		<i>Sanguisorba tenuifolia</i> var. <i>alba</i>	NC_044692	155,457	84	37	8	37.2	Meng et al. (2018)
		<i>Sanguisorba tenuifolia</i>	NC_042223	155,403	86	37	8	37.2	Park et al. (2018)
		<i>Sanguisorba tenuifolia</i>	MK696194	155,415	N/A	N/A	N/A	37.2	Unpublished
		<i>Sanguisorba stipulata</i>	MK696195	155,385	N/A	N/A	N/A	37.2	Unpublished
		<i>Sanguisorba stipulata</i>	MZ145059	155,328	85	37	8	37.2	Unpublished
		<i>Sanguisorba sitchensis</i>	NC_044691	155,127	84	37	8	37.2	Meng et al. (2018)
		<i>Sanguisorba filiformis</i>	NC_044693	154,282	84	37	8	37.3	Meng et al. (2018)
		<i>Sanguisorba filiformis</i> ^a	KY419920	128,491	78	30	4	36.3	Zhang et al. (2017)
		<i>Sanguisorba hakusanensis</i>	MZ145057	155,645	85	37	8	37.2	Unpublished
		<i>Bencomia exstipulata</i>	NC_039924	152,781	84	33	8	37.2	Unpublished
		<i>Bencomia moquiniana</i> ^a	KY420023	128,612	78	30	4	36.1	Zhang et al. (2017)
		<i>Bencomia sphaerocarpa</i> ^a	KY419986	128,439	78	30	4	36.1	Zhang et al. (2017)
		<i>Polylepis australis</i> ^a	KY419989	129,569	78	30	4	36.2	Zhang et al. (2017)
		<i>Polylepis reticulata</i> ^a	KY419921	129,512	78	30	4	36.2	Zhang et al. (2017)
		<i>Polylepis</i> sp. SCZ-2017 ^a	KY419992	129,472	78	30	4	36.2	Zhang et al. (2017)
		<i>Acaena pimatifida</i> ^a	KY419984	129,416	78	30	4	36.2	Zhang et al. (2017)
		<i>Cliffortia repens</i> ^a	KY419983	129,326	78	30	4	36.1	Zhang et al. (2017)
		<i>Margyricarpus pinnatus</i> ^a	KY419972	129,276	78	30	4	36.2	Zhang et al. (2017)
		<i>Dendriopoterium menendezii</i> ^a	KY419966	128,700	78	30	4	36.2	Zhang et al. (2017)
		<i>Poterium spinosum</i> ^a	KY419948	128,592	78	30	4	36.1	Zhang et al. (2017)
Potentilleae		<i>Potentilla stolonifera</i>	NC_044418	156,440	84	37	8	36.9	Park et al. (2019c)
Roseae		<i>Rosa rugosa</i>	NC_044126	156,989	84	37	8	37.2	Kwon et al. (2019b)

IR, inverted repeat; PCG, protein-coding gene; N/A, not available.

^aThese chloroplast genomes are partial chloroplast genome without one IR region.

^bNumber of PCGs in this chloroplast genome was updated in this study.

from low to high degrees (Fig. 3B), suggesting that these proportions are not strongly affected by the geographical distribution. This finding is also congruent with those associated with other plant species investigated in this study (Table 2); however, certain Orchidaceae species, including *Goodyera schlechtendaliana* Rehb. f. and *Gastrodia electa* Blume, displayed very high SNP and INDEL proportions (above 0.01) (Fig. 3A); these can be considered family-specific characteristics.

Two hundred and ninety-two non-redundant SNPs identified based on a multiple-sequence alignment of four *A. pilosa* chloroplast genomes indicated that most of the SNPs are common among the SNPs identified based on six pair-wise comparisons of *A. pilosa*, as the largest number of SNPs among *A. pilosa* is 280, found between MT040192 and KY419942 (Fig. 3A). In addition, 78 out of 292 SNPs and two INDELS were found in 36 of 84 PCGs (42.85%), which is a large number in comparison with other intraspecific variations of *Arabidopsis thaliana* L. (11 PCGs) (Park et al., 2020f), *C. album* (nine PCGs) (Park et al., 2021d), and *Campanula takesimana* Nakai (eight PCGs) (Park et al., 2021a). It is remarkable that only three synonymous SNPs identified in three PCGs were from MT040192; two synonymous SNPs were from MT415496 and no variations were found in *A. pilosa* var. *nepalensis* (MW387437), indicating that the only five PCGs, *matK*, *psbC*, *psaB*, *psbJ*, and *psbB*, have only synonymous SNPs apart from KY419942, lower than those identified in the three species used in this comparison.

In the same sense, *ycf1* exhibited 10 synonymous SNPs, 11 non-synonymous SNPs, and one 9-bp insertion, which did not cause a frameshift only from KY419942. It is also interesting that three *A. pilosa* chloroplast genomes but not KY419942 displayed no intraspecific variation in *ycf1*, which usually presents numerous intraspecific variations (Kim et al., 2020; Park et al., 2020f, 2021d). This provides indirect evidence of the distance between KY419942 and the remaining three *A. pilosa* chloroplast genomes.

An additional interesting factor is that certain proportions of SNPs and INDELS of *Agrimonia* species are highest among those of Sanguisorbeae species. Determination of these proportions was from pair-wise comparisons with the KY419942 chloroplast genome, which displays a high level of intraspecific variation (Fig. 3A). This difference supports the hypothesis that the KY419942 chloroplast genome evolved independently through, for instance, long-term geographical isolation, akin to island plant species (Zhang et al., 2019a; Celiński et al., 2020).

Interspecific variations identified from the chloroplast genomes of three *Agrimonia* species

Given the fewest intraspecific variations between *A. pilosa* and *A. pilosa* var. *nepalensis* among the six pair-wise comparisons of *A. pilosa* chloroplasts conducted here (Fig. 4A), the chloroplast genomes of *A. coreana*, *A. nipponica*, and *A. pilosa* were used to investigate the interspecific variations. Due to the differences from the KY419942 chloroplast genome (Figs. 4A), four *Agrimonia* chloroplast genomes were used for this investigation (Fig. 4B). The numbers of SNPs identified from the four chloroplast genomes range from 244 to 311 (Fig. 4B), higher than those of SNPs within *A. pilosa* (Fig. 4A) except for that between *A. coreana* and *A. pilosa* (GenBank accession number: MT415496). The numbers of INDEL regions and the total lengths of the INDEL regions identified from *A. pilosa* (GenBank accession number: MT415496) are slight lower than those identified from another *A. pilosa* chloroplast genome (GenBank accession number: KY419942) (Fig. 4B). Moreover, *ycf1*, which has a high number of intraspecific variations in KY419942, presented more variations than the intraspecific variations of *A. pilosa*: 22 non-synonymous SNPs, 14 synonymous SNPs, and one 9-bp insertion. These variations show the presence of four non-synonymous SNPs and four synonymous SNPs from *A. coreana*; six non-synonymous SNPs and one synonymous SNP from *A. nipponica*; seven non-synonymous SNPs, five synonymous SNPs, and a 9-bp insertion from KY419942 (*A. pilosa*); and five non-synonymous SNPs and one synonymous SNP from MT415496 (*A. pilosa*). In addition, three synonymous SNPs showed that half were from *A. coreana* and MT415496 and half were from *A. nipponica* and KY419942, suggesting that KY419942 is a different species from *A. pilosa* based on the investigation of sequence variations, a case similar to certain cryptic plant species (Gurushidze et al., 2008; Okuyama and Kato 2009; Myszczynski et al., 2017; Li et al., 2020b; Liu et al., 2020b) or hybrid-origin species such as *Arabidopsis suecica* (r.) Norrl. (O'Kane et al., 1996; Novikova et al., 2017) and *Arabis* species (Kawabe et al., 2018).

The chromosome numbers for *A. pilosa* isolated in China and on Mt. Baekdu were reported to be $2n = 4x = 28$ (Kwon et al., 2005) and $2n = 2x = 16$ (Mitrenina et al., 2020), respectively. These different chromosomal configurations of *A. pilosa* suggest at least two possible origins of *A. pilosa*. First, a tetraploid of *A. pilosa* may have been formed via allopolyploidization, as in the case of *Coffea arabica* L. which originates from *C. canephora* Pierre ex A. Froehner and *C. eugenioides* S. Moore (Clarindo and Carvalho, 2008). In this case, two types of chloroplast genomes from two parental species can be found in a hybrid species if allopolyploidization

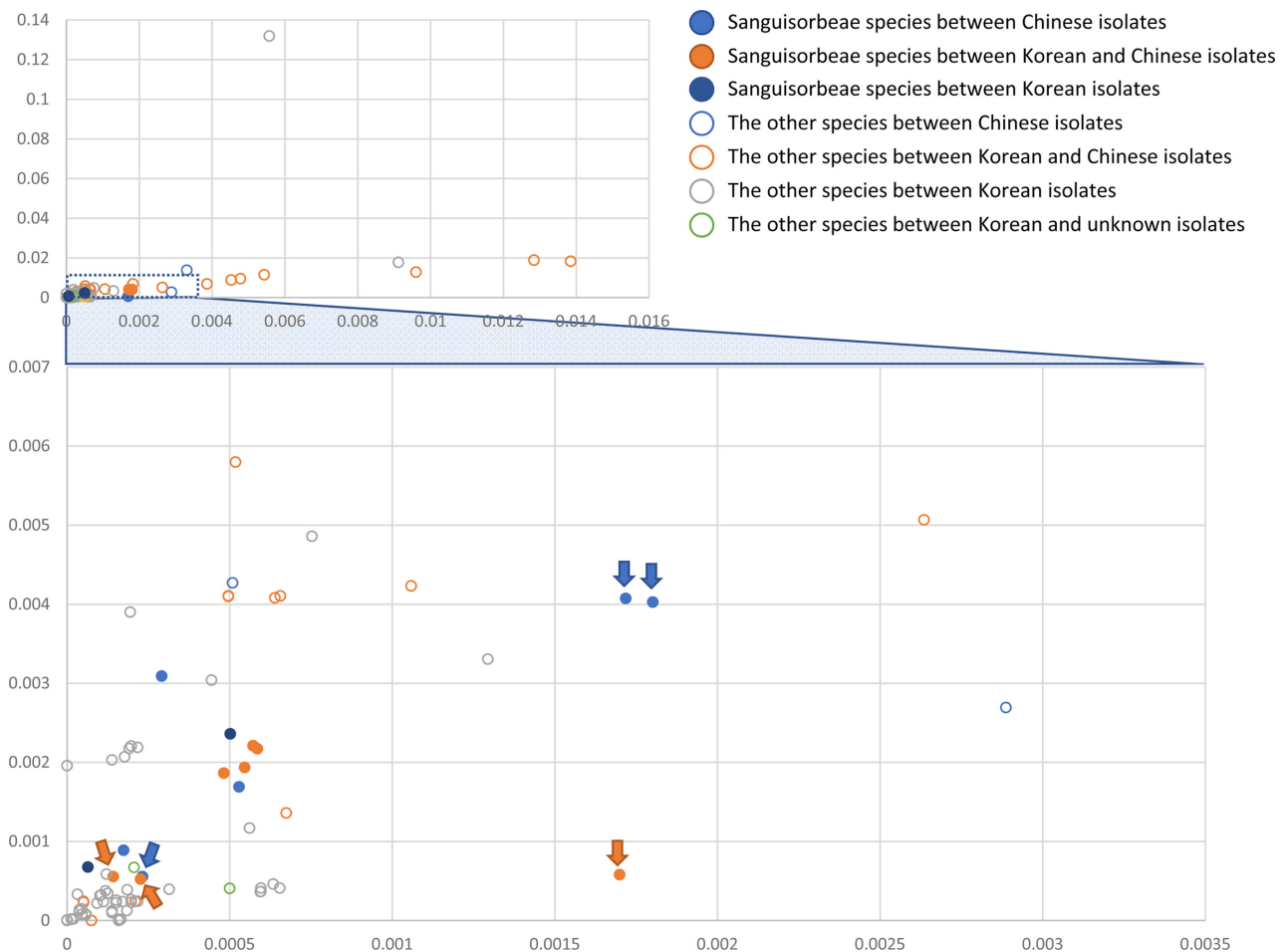


Fig. 4. Numbers of single nucleotide polymorphisms (SNPs) and insertions and deletions (INDELs) identified among the four *Agrimonia* chloroplast genomes. Yellow, grey, and blue boxes indicate the *A. coreana*, *A. nipponica*, and *A. pilosa* chloroplast genomes, respectively. Numbers of SNPs and INDEL regions with the total lengths of INDEL regions are displayed on the lines connecting two chloroplast genomes. The comparisons with the KY419942 chloroplast genome exhibit additional information about the numbers of SNPs and INDELs in the IR region due to a lack of one of the IR regions in the KY419942 chloroplast genome.

occurred in both directions, in agreement with the two distinct types of chloroplast genomes found in this study. The second scenario involves two different speciation events, resulting in two different basic chromosome numbers, $n = 7$ (Kwon et al., 2005) and $n = 8$ (Mitrenina et al., 2020). From the two speciation routes, two different types of chloroplast genomes could be passed down to *A. pilosa*, which can explain the current distinct between the two chloroplast genomes.

Phylogenetic analysis of the *Agrimonia* chloroplast genome sequence

ML and BI phylogenetic trees were constructed based on the 33 Sanguisorbeae species, including the two *Agrimonia* chloroplast genomes assembled in this study. *Agrimonia* exhibited monophyletic characteristics in that it formed one clade clustered with the *Hagenia* and *Leucosidea* genera (Fig. 5), congruent with the taxonomy of the subtribe Agrimoniinae

(Chung, 2008) and previous phylogenetic analyses (Helfgott et al., 2000; Zhang et al., 2017). In addition, *A. pilosa* var. *nepalensis* was clustered with two *A. pilosa* complete chloroplast genomes isolated in Korea and China (Fig. 5), also in agreement with the low number of intraspecific variations among the four *A. pilosa* chloroplast genomes (Fig. 4A).

In contrast, the *A. pilosa* chloroplast (GenBank accession number: KY419942), which is different from the remaining three *A. pilosa* chloroplast genomes, is a sister to *A. nipponica* (Fig. 5), in contrast to the expected results based on the numbers of SNPs and INDELs (Fig. 3B). This indicates a need for an additional investigation of *A. pilosa* to test the suggested hypotheses that *A. pilosa* was formed from hybridization or independent speciation events. The chromosome numbers of the samples of *A. pilosa* from which chloroplast genomes were determined are unknown. The chromosome number heterogeneity of *A. pilosa*, as discussed above,

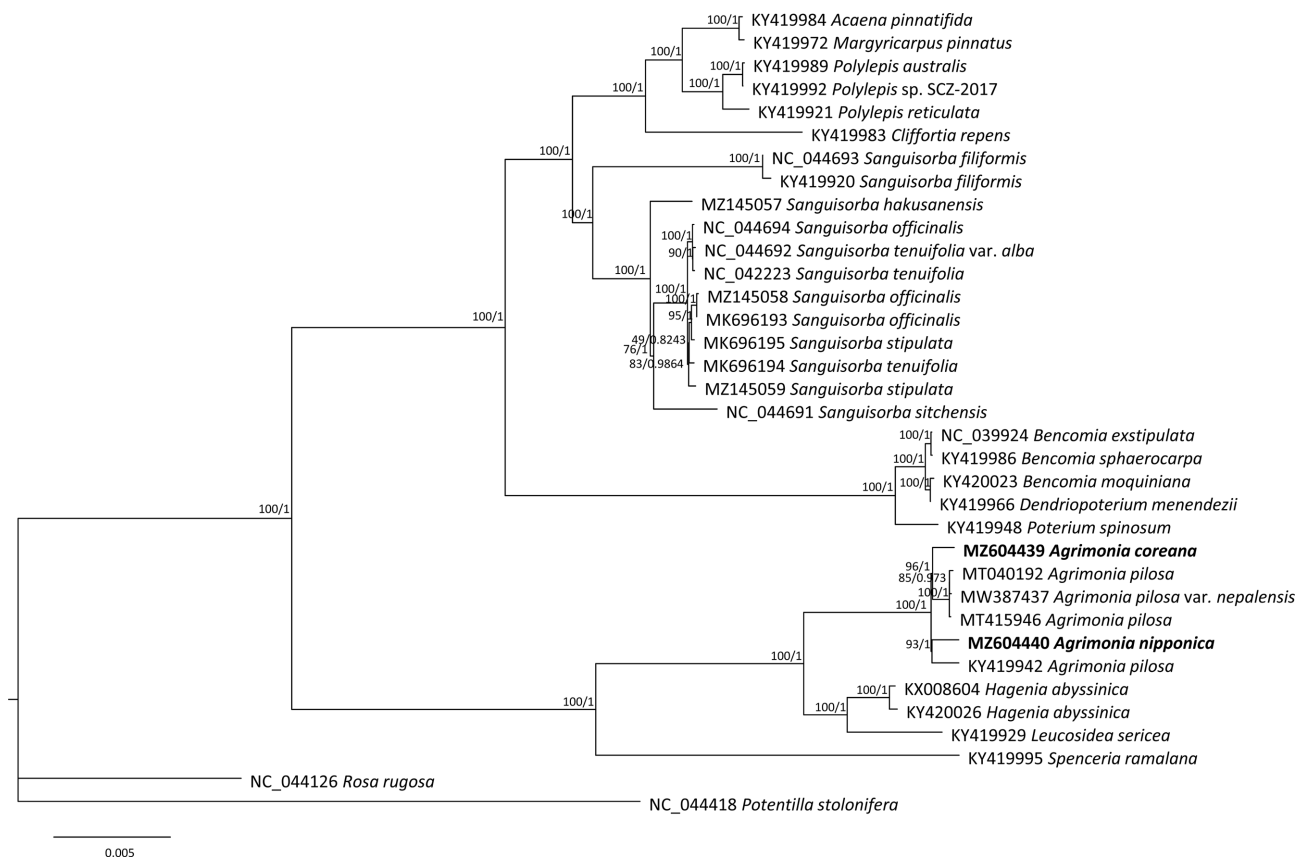


Fig. 5. Maximum likelihood tree of 33 *Sanguisorbeae* chloroplast genomes with two outgroup species. Numbers on branches in the phylogenetic tree indicate the bootstrap values of ML and the prior possibility of Bayesian inference. Chloroplast genomes sequenced in this study are presented in bold text.

associated with hybridization and polyploidization, may be responsible for the close relationship between the *A. pilosa* chloroplast (GenBank accession number: KY419942) and *A. nipponica*. The chromosome numbers of *A. nipponica* and *A. coreana* are $2n = 4x = 28$ (Iwatsubo et al., 1993), reflecting the complex speciation history of *Agrimonia* species. Based on leaf morphological characteristics, *A. coreana* and *A. nipponica* are closely related by having 1–2 pairs of lateral leaflets in the mid-caule region and an abaxial surface of the leaflets with densely glandular and tomentose hairs (Chung, 2017), in contrast to the phylogenetic tree (Fig. 5). This indicates that additional research is needed to investigate the corresponding morphological features possibly to support the current phylogenetic tree or to consider the possibility of hybridized individuals of *A. pilosa*. Furthermore, *A. pilosa* may contain cryptic species, or there may even have been a misidentification of *A. pilosa* chloroplast genome (KY419942), as in the *C. album* (Park et al., 2021d) and *Magnolia insignis* (Wall.) Blume and *Magnolia alba* Figlar & Noot cases. Park (2020) can be referenced for further analyses.

Comparative analysis of SSR polymorphisms on six *Agrimonia* chloroplast genomes

SSRs were identified from the six *Agrimonia* chloroplast genomes, denoting the total numbers of SSRs, including normal SSRs, potential SSRs, and the extended SSRs ranging from 546 to 558 apart from KY419942 due to the lack of one IR region (Table 4). PentaSSR and hexaSSR identified in the chloroplast genome are usually classified into the potential SSR of which the length is 10 or 12 bp with two repeats. Only the *A. coreana* chloroplast genome contained two normal pentaSSRs (P0000056 and P0000057) with the three repeats (Fig. 6A). The numbers of normal SSRs for the six chloroplast genomes ranged from 73 to 83 (Fig. 6A). The four *A. pilosa* chloroplast genomes exhibited a range of 73 to 74, in good agreement with other plant species, such as *C. album* (53 to 55) (Park et al., 2021d). In addition, KY419942 without one IR region displayed the same level of normal SSRs (Fig. 6A), reflecting that the IR region of *A. pilosa* does not contain normal SSRs. The *A. coreana* and *A. nipponica* chloroplast genomes contained more normal SSRs than those of *A. pilosa*: *A. coreana* covers more monoSSRs and pentaSSRs, while *A.*

Table 4. List of normal SSRs, potential SSRs, and extended SSRs identified from the six *Agrimonia* chloroplast genomes.

SSR type		<i>A. coreana</i>	<i>A. nipponica</i>	<i>A. pilosa</i> MT415946	<i>A. pilosa</i> MT040192	<i>A. pilosa</i> KY419942	<i>A. pilosa</i> var. <i>nepalensis</i>
Normal SSRs	MonoSSR	62	57	55	55	56	54
	DiSSR	10	12	9	9	9	9
	TriSSR	3	3	3	3	2	3
	TetraSSR	6	7	7	7	6	7
	PentaSSR	2	0	0	0	0	0
Potential SSRs	PentaSSR	307	310	312	312	248	311
	HexaSSR	106	103	104	105	86	104
Extended SSRs	HeptaSSR	32	31	33	32	28	32
	OctaSSR	16	14	13	13	10	13
	NonaSSR	9	9	9	9	10	9
	DecaSR	5	4	4	4	5	4
Total		558	550	549	549	460	546

SSR, simple sequence repeat.

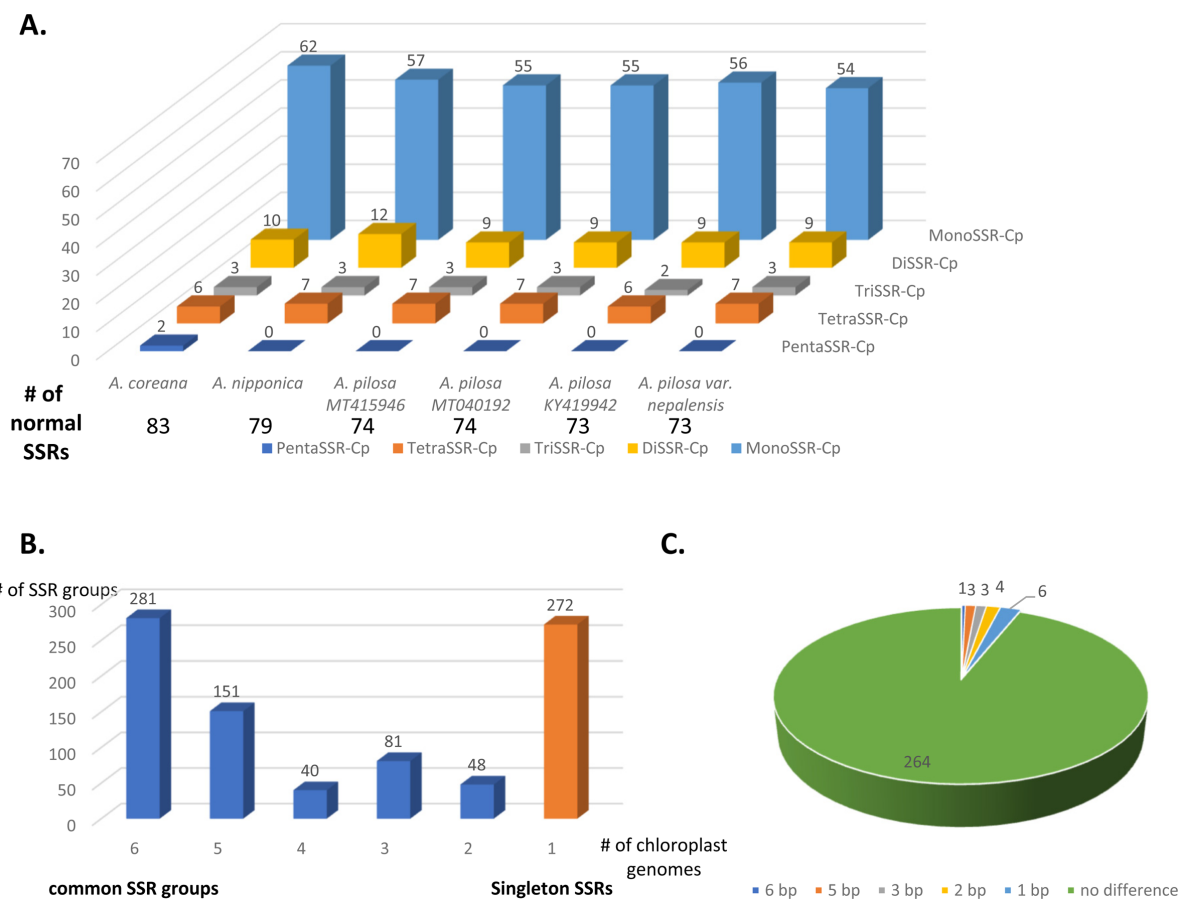


Fig. 6. Number of normal simple sequence repeats (SSRs) and distribution of SSR groups identified on six *Agrimonia* chloroplast genomes. **A.** Number of SSR types in each of the six *Agrimonia* chloroplast genomes. The X-axis represents the six *Agrimonia* chloroplasts and the Y-axis indicates the number of normal SSRs. The five different colors indicate the five types of SSRs, i.e., monoSSRs, diSSRs, triSSRs, tetraSSRs, and pentaSSRs. The table below the X-axis represents the numbers of normal SSRs along with the six *Agrimonia* samples. **B.** Number of SSR groups along with the number of samples in each SSR group. The X-axis is the number of species in the SSR groups and the Y-axis is the number of SSR groups. The orange-colored bar indicates singleton SSRs. **C.** Pie graph presenting SSR groups, showing the length differences in the SSR group.

Table 5. List of SSRs from common SSR groups displaying length difference of SSR regions.

Name	Type	<i>A. coreana</i>	<i>A. nipponica</i>	<i>A. pilosa</i> MT415946	<i>A. pilosa</i> MT040192	<i>A. pilosa</i> KY419942	<i>A. pilosa</i> var. <i>nepalensis</i>
SSRGroup 281	diSSR	16	10	10	10	8	10
SSRGroup 280	pentaSSR	15	10	10	10	10	10
SSRGroup 223	monoSSR	10	13	11	11	15	11
SSRGroup 58	pentaSSR	15	10	10	10	10	10
SSRGroup 567	monoSSR	10	11	10	10	13	10
SSRGroup 371	monoSSR	11	12	11	11	11	14
SSRGroup 120	monoSSR	11	11	10	11	13	11
SSRGroup 455	monoSSR	11	12	11	11	10	11
SSRGroup 373	monoSSR	11	10	11	10	12	11
SSRGroup 359	diSSR	14	12	12	12	12	12
SSRGroup 189	HexaSSR PentaSSR	12	10	12	12	12	12
SSRGroup 502	monoSSR	11	12	11	12	12	11
SSRGroup 338	monoSSR	10	10	11	11	10	11
SSRGroup 195	monoSSR	12	11	12	12	12	12
SSRGroup 92	monoSSR	13	13	12	12	13	12
SSRGroup 57	monoSSR	10	11	11	11	11	11
SSRGroup 36	monoSSR	10	11	11	11	11	11

SSR, simple sequence repeat.

nipponica had more diSSRs and tetraSSRs (Fig. 6A).

Based on a comparison of the flanking sequences of SSRs identified from the six *Agrimonia* chloroplast genomes, 601 SSR groups and 272 singleton SSRs were identified (Fig. 6B). In addition, 281 of 601 SSR groups (46.76%), covering six SSRs from six different chloroplast genomes, were identified (Fig. 6B). This indicates that fewer than half of the *Agrimonia* SSRs are shared among the six chloroplast genomes, referred to as common SSR groups. One hundred and fifty-one SSR groups (25.12%) contained five SSRs that originated from five chloroplast genomes (Fig. 6B). Seventeen out of 281 common SSR groups (6.05%) exhibited length variations in the SSR regions (Fig. 6C), representing a good molecular marker candidate with which to distinguish species or populations. The length differences in the SSR region ranged from 1 bp to 6 bp (Fig. 6C). Twelve out of 17 common SSR groups (70.59%) were monoSSR and the remaining five common SSR groups consisted of two diSSRs, two pentaSSRs, and one pentaSSR/hexaSSR (Table 5). The greatest difference (6 bp found in SSR Group 281) was caused by three additional time repeats of diSSRs, indicating that the SSRs in KY419942 and the *A. coreana* chloroplast genomes were eight-fold TA diSSR while the remaining SSRs are five-fold TA diSSR (Table 5). Interestingly,

SSR Group 189 displayed one of six SSRs that was the pentaSSR type (*A. nipponica*), while the remaining SSRs were hexaSSRs (Fig. 6B, C). These variations in the 17 common SSR groups can serve as a good example for developing molecular markers based on SSRs in chloroplast genomes. In detail, the 151 SSR groups covering five *Agrimonia* chloroplast genomes can be considered as another type of molecular marker candidate because one of the six chloroplast genomes lacks the SSR, which is clear evidence allowing the detection of a specific *Agrimonia* sample. Moreover, because the number of SSRs can be affected by the evolution of chloroplast genomes (Sawicki et al., 2020), SSRs identified from the six *Agrimonia* chloroplast genomes can be used to understand the corresponding evolutionary features with the additional chloroplast genome of *Agrimonia* and a neighbor genus in the near future. Taken together, the variations identified in the SSRs from six *Agrimonia* chloroplast genomes exhibit not only the dynamic characteristics of *Agrimonia* SSRs but also suggest potential uses of *Agrimonia* chloroplast genomes.

CONCLUSION

We completed the chloroplast genomes of *A. coreana* and

A. nipponica isolated in Korea and conducted comparative analyses of six *Agrimonia* chloroplast genomes originating from three species. The *Agrimonia coreana* and *A. nipponica* chloroplast genomes are 151,362 bp and 155,161 bp long, respectively. High nucleotide diversity was detected in the *ycfI* gene, congruent with earlier work (Hong et al., 2017; Jiang et al., 2017; Liu et al., 2018; de Souza et al., 2019; Kim et al., 2019d; Li et al., 2019; Celiński et al., 2020; Park et al., 2020e; Loeuille et al., 2021). The intraspecific variations of four *A. pilosa* chloroplast genomes revealed that KY419942 is distinct from the remaining three *A. pilosa* chloroplast genomes. In addition, the interspecific variations among *A. coreana*, *A. nipponica*, and *A. pilosa* indicated that the distance between two *A. pilosa* chloroplast genomes is similar to that between *A. coreana* and *A. nipponica*. A phylogenetic analysis found that KY419942 was clustered with *A. nipponica*, suggesting that KY419942 (*A. pilosa*) might have been misidentified or a cryptic species caused by complex evolutionary histories. A comparison of SSRs identified from the six *Agrimonia* chloroplast genomes suggests potential molecular markers with which to distinguish between species or among populations of the same species. Considering these aspects together with the results of this study, our *A. coreana* and *A. nipponica* chloroplast genomes provide insight into the phylogenetic relationships among *Agrimonia* species in Korea and about the morphological as well as interspecific and intraspecific features of chloroplast genomes, including variations of chloroplast genomes. Moreover, information about the two *Agrimonia* chloroplast genomes serves to unravel the complex evolutionary history of *A. pilosa*, *A. coreana*, and *A. nipponica*, which require further analyses that can include deciphering additional chloroplast genomes of *A. pilosa* and neighboring species.

ONLINE SUPPLEMENTARY MATERIALS

S1 and S2 are available at <https://doi.org/10.11110/kjpt.2024.54.1.47>

- S1. Nucleotide diversity of *ycfI* in the six *Agrimonia* chloroplast genomes
- S2. Numbers and proportions of SNPs and INDELS identified among the four *Agrimonia pilosa* chloroplast genomes

ORCID: Jongsun PARK <http://orcid.org/0000-0003-0786-4701>; Hwa-Jung SUH <http://orcid.org/0000-0003-2528-425X>; Hong XI <https://orcid.org/0000-0003-0731-2210>

ACKNOWLEDGMENTS

We would like to express our appreciation of Dr. Sang-Hun

Oh for the useful advice about this manuscript. This work was supported by an InfoBoss Research Grant (IBG-0035).

CONFLICTS OF INTEREST

The authors declare that there are no conflicts of interest.

LITERATURE CITED

- Alzahrani, D. A., S. S. Yaradua, E. J. Albokhari and A. Abba. 2020. Complete chloroplast genome sequence of *Barleria prionitis*, comparative chloroplast genomics and phylogenetic relationships among Acanthoideae. *BMC Genomics* 21: 393.
- Baek, J., S. Park, J. Lee, J. Min, J. Park and G. W. Lee. 2021. The complete chloroplast genome of *Chrysanthemum zawadskii* Herbich (Asteraceae) isolated in Korea. *Mitochondrial DNA Part B Resources* 6: 1956–1958.
- Bolger, A. M., M. Lohse and B. Usadel. 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* 30: 2114–2120.
- Bum, H. M., H. K. Kim, N. S. Kim, W. Kwon and J. Park. 2020. The complete chloroplast genome of *Douinia plicata* (Lindb.) Konstant. & Vilnet (Scapaniaceae, Jungermanniales). *Mitochondrial DNA Part B Resources* 5: 3680–3682.
- Cai, Z., C. Penafior, J. V. Kuehl, J. Leebens-Mack, J. E. Carlson, C. W. dePamphilis, J. L. Boore and R. K. Jansen. 2006. Complete plastid genome sequences of *Drimys*, *Liriodendron*, and *Piper*: Implications for the phylogenetic relationships of magnoliids. *BMC Evolutionary Biology* 6: 77.
- Celiński, K., H. Kijak and J. Wiland-Szymańska. 2020. Complete chloroplast genome sequence and phylogenetic inference of the Canary Islands Dragon Tree (*Dracaena draco* L.). *Forests* 11: 309.
- Chang, H., L. Zhang, H. Xie, J. Liu, Z. Xi and X. Xu. 2021. The conservation of chloroplast genome structure and improved resolution of infrafamilial relationships of Crassulaceae. *Frontiers in Plant Science* 12: 631884.
- Chen, L. and Y.-H. Kang. 2014. Antioxidant activities of *Agrimonia pilosa* ledeb: *In vitro* comparative activities of its different fractions. *Korean Journal of Plant Resources* 27: 642–649.
- Cheon, K.-S., K.-A. Kim, S.-K. Jang and K.-O. Yoo. 2016. Complete chloroplast genome sequence of *Campanula takesimana* (Campanulaceae), an endemic to Korea. *Mitochondrial DNA Part A, DNA Mapping, Sequencing, and Analysis* 27: 2169–2171.
- Cho, M.-S., Y. Kim, S.-C. Kim and J. Park. 2019. The complete chloroplast genome of Korean *Pyrus ussuriensis* Maxim. (Rosaceae): Providing genetic background of two types of *P. ussuriensis*. *Mitochondrial DNA Part B Resources* 4: 2424–

- 2425.
- Choi, N. J., H. Xi and J. Park. 2021. A comparative analyses of the complete mitochondrial genomes of fungal endosymbionts in *Sogatella furcifera*, white-backed planthoppers. *International Journal of Genomics* 2021: 6652508.
- Choi, Y. G., N. Yun, J. Park, H. Xi, J. Min, Y. Kim and S.-H. Oh. 2020. The second complete chloroplast genome sequence of the *Viburnum erosum* (Adoxaceae) showed a low level of intra-species variations. *Mitochondrial DNA Part B Resources* 5: 271–272.
- Chung, K.-S. 2008. A systematic study of genus *Agrimonia* (Rosaceae). PhD dissertation, The University of Oklahoma, Norman, OK, USA, 183 pp.
- Chung, K.-S. 2017. The first record of *Agrimonia gorovoi* Rumjantsev in Korea (Rosaceae). *Korean Journal of Plant Taxonomy* 47: 132–136.
- Chung, K. S. and Y.-S. Kim. 2000. A taxonomic study on the genus *Agrimonia* (Rosaceae) in Korea. *Korean Journal of Plant Taxonomy* 30: 315–337.
- Clarindo, W. R. and C. R. Carvalho. 2008. First *Coffea arabica* karyogram showing that this species is a true allotetraploid. *Plant Systematics and Evolution* 274: 237–241.
- Darlington, C. D. and A. P. Wylie. 1955. *Chromosome Atlas of Flower Plants*. George Allen and Union Ltd., London, 519 pp.
- de Souza, U. J. B., R. Nunes, C. P. Targueta, J. A. F. Diniz-Filho and M. P. de Campos Telles. 2019. The complete chloroplast genome of *Stryphnodendron adstringens* (Leguminosae -Caesalpinioideae): Comparative analysis with related Mimosoid species. *Scientific Reports* 9: 14206.
- de Vries, J., F. L. Sousa, B. Bölter, J. Soll and S. B. Gould. 2015. YCF1: A green TIC? *The Plant Cell* 27: 1827–1833.
- Deamer, D., M. Akeson and D. Branton. 2016. Three decades of nanopore sequencing. *Nature Biotechnology* 34: 518–524.
- Dong, W., C. Xu, C. Li, J. Sun, Y. Zuo, S. Shi, T. Cheng, J. Guo and S. Zhou. 2015. *ycf1*, the most promising plastid DNA barcode of land plants. *Scientific Reports* 5: 8348.
- Drescher, A., S. Ruf, T. Calsa Jr., H. Carrer and R. Bock. 2000. The two largest chloroplast genome-encoded open reading frames of higher plants are essential genes. *The Plant Journal* 22: 97–104.
- Du, Y.-P., Y. Bi, F.-P. Yang, M.-F. Zhang, X.-Q. Chen, J. Xue and X.-H. Zhang. 2017. Complete chloroplast genome sequences of *Lilium*: Insights into evolutionary dynamics and phylogenetic analyses. *Scientific Reports* 7: 5751.
- Gichira, A. W., Z. Li, J. K. Saina, Z. Long, G. Hu, R. W. Gituru, Q. Wang and J. Chen. 2017. The complete chloroplast genome sequence of an endemic monotypic genus *Hagenia* (Rosaceae): Structural comparative analysis, gene content and microsatellite detection. *PeerJ* 5: e2846.
- Gil, H.-Y., Y. Kim, S.-H. Kim, J.-H. Jeon, Y. Kwon, S.-C. Kim and J. Park. 2019. The complete chloroplast genome of *Pyrus ussuriensis* Maxim. (Rosaceae). *Mitochondrial DNA Part B Resources* 4: 1000–1001.
- Goodwin, S., J. D. McPherson and W. R. McCombie. 2016. Coming of age: Ten years of next-generation sequencing technologies. *Nature Reviews Genetics* 17: 333–351.
- Greiner, S., P. Lehwark and R. Bock. 2019. OrganellarGenomeDRAW (OGDRAW) version 1.3. 1: Expanded toolkit for the graphical visualization of organellar genomes. *Nucleic Acids Research* 47: W59–W64.
- Gurushidze, M., R. M. Fritsch and F. R. Blattner. 2008. Phylogenetic analysis of *Allium* subg. *Melanocrommyum* infers cryptic species and demands a new sectional classification. *Molecular Phylogenetics and Evolution* 49: 997–1007.
- Hassemer, G., S. Bruun-Lund, A. B. Shipunov, B. G. Briggs, H. M. Meudt and N. Rønsted. 2019. The application of high-throughput sequencing for taxonomy: The case of *Plantago* subg. *Plantago* (Plantaginaceae). *Molecular Phylogenetics and Evolution* 138: 156–173.
- Helfgott, D. M., J. Francisco-Ortega, A. Santos-Guerra, R. K. Jansen and B. B. Simpson. 2000. Biogeography and breeding system evolution of the woody *Bencomia* alliance (Rosaceae) in Macaronesia based on ITS sequence data. *Systematic Botany* 25: 82–97.
- Heo, K.-I., J. Park, H. Xi and J. Min. 2020. The complete chloroplast genome of *Agrimonia pilosa* Ledeb. isolated in Korea (Rosaceae): Investigation of intraspecific variations on its chloroplast genomes. *Mitochondrial DNA Part B Resources* 5: 2264–2266.
- Hong, K.-J., W. Ki, I.-J. Lee, H. Lee, J. Park and W. Lee. 2022. The complete mitochondrial genome of *Aclees taiwanensis* Kôno, 1933 (Coleoptera: Curculionidae). *Mitochondrial DNA Part B Resources* 7: 1460–1462.
- Hong, S.-Y., K.-S. Cheon, K.-O. Yoo, H.-O. Lee, K.-S. Cho, J.-T. Suh, S.-J. Kim, J.-H. Nam, H.-B. Sohn and Y.-H. Kim. 2017. Complete chloroplast genome sequences and comparative analysis of *Chenopodium quinoa* and *C. album*. *Frontiers in Plant Science* 8: 1696.
- Huelsenbeck, J. P. and F. Ronquist. 2001. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* 17: 754–755.
- Hutchinson, J. 1964. *Agrimonia*. In *The Genera of Flowering Plants*. Vol. 1. Angiospermae Dicotyledones. Clarendon Press, Oxford. P. 206.
- Iwatsubo, Y., M. Mishima and N. Naruhashi. 1993. Chromosome studies of Japanese *Agrimonia* (Rosaceae). *Cytologia* 58: 453–461.
- Jiang, D., Z. Zhao, T. Zhang, W. Zhong, C. Liu, Q. Yuan and L. Huang. 2017. The chloroplast genome sequence of *Scutel-*

- laria baicalensis* provides insight into intraspecific and interspecific chloroplast genome diversity in *Scutellaria*. *Genes* 8: 227.
- Jung, C. H., J.-H. Kim, S. Park, D.-H. Kweon, S.-H. Kim and S.-G. Ko. 2010. Inhibitory effect of *Agrimonia pilosa* Ledeb. on inflammation by suppression of iNOS and ROS production. *Immunological Investigations* 39: 159–170.
- Kalkman, C. 2004. Rosaceae. In *Flowering Plants · Dicotyledons*. Kubitzki, K. (ed.), Springer, Berlin. Pp. 343–386.
- Kang, M.-J., S.-C. Kim, H.-R. Lee, S.-A. Lee, J.-W. Lee, T.-D. Kim and E.-J. Park. 2020. The complete chloroplast genome of Korean *Gastrodia elata* Blume. *Mitochondrial DNA Part B Resources* 5: 1015–1016.
- Katoh, K. and D. M. Standley. 2013. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution* 30: 772–780.
- Kawabe, A., H. Nukii and H. Y. Furihata. 2018. Exploring the history of chloroplast capture in *Arabis* using whole chloroplast genome sequencing. *International Journal of Molecular Sciences* 19: 602.
- Kikuchi, S., J. Bédard, M. Hirano, Y. Hirabayashi, M. Oishi, M. Imai, M. Takase, T. Ide and M. Nakai. 2013. Uncovering the protein translocon at the chloroplast inner envelope membrane. *Science* 339: 571–574.
- Kim, D.-H., B.-J. An and J.-Y. Lee. 2011a. Whitening activities of the *Agrimonia pilosa* L. extracts. *Journal of Applied Biological Chemistry* 54: 284–289.
- Kim, D. E., P. Kim, H. Lee, N.-H. Kim, D. Kim, M.-J. Lee, Y.-G. Ban, B.-J. Jang and J. Park. 2023. Comprehensive analysis of the complete mitochondrial genome of *Melanoplus differentialis* (Acrididae: Melanoplinae) captured in Korea. *Entomological Research* 53: 66–81.
- Kim, J., Y. Kim and J. Park. 2019a. Complete chloroplast genome sequence of the *Salix koriyanagi* Kimura ex Goerz (Salicaceae). *Mitochondrial DNA Part B Resources* 4: 549–550.
- Kim, M.-H., S. Park, J. Lee, J. Baek, J. Park and G. W. Lee. 2021a. The complete chloroplast genome of *Glycyrrhiza uralensis* Fisch. isolated in Korea (Fabaceae). *Korean Journal of Plant Taxonomy* 51: 353–362.
- Kim, M., H. Xi and J. Park. 2021b. Genome-wide comparative analyses of GATA transcription factors among 19 *Arabidopsis* ecotype genomes: Intraspecific characteristics of GATA transcription factors. *PLoS ONE* 16: e0252181.
- Kim, M., H. Xi, S. Park, Y. Yun and J. Park. 2021c. Genome-wide comparative analyses of GATA transcription factors among seven *Populus* genomes. *Scientific Reports* 11: 16578.
- Kim, S.-H., C. H. Cho, M. Yang and S.-C. Kim. 2017. The complete chloroplast genome sequence of the Japanese *Camellia* (*Camellia japonica* L.). *Mitochondrial DNA Part B Resources* 2: 583–584.
- Kim, S.-T., S.-H. Oh and J. Park. 2021d. The complete chloroplast genome of *Diarthron linifolium* (Thymelaeaceae), a species found on a limestone outcrop in eastern Asia. *Korean Journal of Plant Taxonomy* 51: 345–352.
- Kim, T. H., J. M. Kim, J. M. Baek, T. W. Kim, D. J. Kim, J. H. Park and M. Choe. 2011b. Antioxidant and whitening effects of *Agrimonia pilosa* Ledeb. water extract. *Korean Journal of Medicinal Crop Science* 19: 177–184.
- Kim, Y., Y. Chung and J. Park. 2019b. The complete chloroplast genome sequence of *Dysphania pumilio* (R. Br.) Mosyakin & Clemants (Amaranthaceae). *Mitochondrial DNA Part B Resources* 4: 403–404.
- Kim, Y., K.-I. Heo, S. Lee and J. Park. 2019c. The complete chloroplast genome sequence of *Pseudostellaria palibiniana* (Takeda) Ohwi (Caryophyllaceae). *Mitochondrial DNA Part B Resources* 4: 973–974.
- Kim, Y., J. Park and Y. Chung. 2019d. Comparative analysis of chloroplast genome of *Dysphania ambrosioides* (L.) Mosyakin & Clemants understanding phylogenetic relationship in genus *Dysphania* R. Br. *Korean Journal of Plant Resources* 32: 644–668.
- Kim, Y., J. Park and Y. Chung. 2019e. The complete chloroplast genome of *Suaeda japonica* Makino (Amaranthaceae). *Mitochondrial DNA Part B Resources* 4: 1505–1507.
- Kim, Y., J. Park and Y. Chung. 2020. The comparison of the complete chloroplast genome of *Suaeda japonica* Makino presenting different external morphology (Amaranthaceae). *Mitochondrial DNA Part B Resources* 5: 1616–1618.
- Kim, Y., H. Xi and J. Park. 2021e. The complete chloroplast genome of *Limonium tetragonum* (Plumbaginaceae) isolated in Korea. *Korean Journal of Plant Taxonomy* 51: 337–344.
- Kim, Y., J.-S. Yi, J. Min, H. Xi, D. Y. Kim, J. Son, J. Park and J.-I. Jeon. 2019f. The complete chloroplast genome of *Aconitum coreanum* (H. Lév.) Rapaics (Ranunculaceae). *Mitochondrial DNA Part B Resources* 4: 3404–3406.
- Kwon, W., Y. Kim and J. Park. 2019a. The complete chloroplast genome of Korean *Marchantia polymorpha* subsp. *ruderalis* Bischl. & Boisselier: Low genetic diversity between Korea and Japan. *Mitochondrial DNA Part B Resources* 4: 959–960.
- Kwon, W., Y. Kim and J. Park. 2019b. The complete mitochondrial genome of *Dumortiera hirsuta* (Sw.) Nees (Dumortieraceae, Marchantiophyta). *Mitochondrial DNA Part B Resources* 4: 1586–1587.
- Kwon, W., Y. Kim and J. Park. 2019c. The complete mitochondrial genome of Korean *Marchantia polymorpha* subsp. *ruderalis* Bischl. & Boisselier: Inverted repeats on mitochondrial genome between Korean and Japanese isolates. *Mitochondrial DNA Part B Resources* 4: 769–770.

- Kwon, W., J. Min, H. Xi and J. Park. 2019d. The complete chloroplast genome of *Fissidens nobilis* Griff. (Fissidentaceae, Bryophyta). *Mitochondrial DNA Part B Resources* 4: 2225–2226.
- Kwon, Y. J., M. R. Sul, J. K. An, C. H. Kim and B.-Y. Sun. 2005. Chromosome numbers of vascular plants of Mt. Baekdu and adjacent area in China. *Korean Journal of Plant Taxonomy* 35: 47–55.
- Lee, B. and J. Park. 2021a. The complete chloroplast genome of *Zoysia japonica* Steud. isolated in Korea (Poaceae): Investigation of potential molecular markers on *Z. japonica* chloroplast genomes. *Plant Biotechnology Reports* 15: 707–715.
- Lee, B. and J. Park. 2021b. The complete chloroplast genome of *Zoysia matrella* (L.) Merr. isolated in Korea (Poaceae): Investigation of intraspecific variations on chloroplast genomes. *Mitochondrial DNA Part B Resources* 6: 572–574.
- Lee, J., J. Park, H. Xi and J. Park. 2020. Comprehensive analyses of the complete mitochondrial genome of *Figulus binodulus* (Coleoptera: Lucanidae). *Journal of Insect Science* 20: 10.
- Lee, Y. S., J. Y. Park, J.-K. Kim, H. O. Lee, H.-S. Park, S.-C. Lee, J. H. Kang, T. J. Lee, S. H. Sung and T.-J. Yang. 2016. Complete chloroplast genome sequence of *Artemisia fukudo* Makino (Asteraceae). *Mitochondrial DNA Part B Resources* 1: 376–377.
- Li, C., Y. Zheng and P. Huang. 2020a. Molecular markers from the chloroplast genome of rose provide a complementary tool for variety discrimination and profiling. *Scientific Reports* 10: 12188.
- Li, H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv preprint at: <https://arxiv.org/abs/1303.3997>.
- Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin and 1000 Genome Project Data Processing Subgroup. 2009. The sequence alignment/map format and SAMtools. *Bioinformatics* 25: 2078–2079.
- Li, W., Y. Liu, Y. Yang, X. Xie, Y. Lu, Z. Yang, X. Jin, W. Dong and Z. Suo. 2018. Interspecific chloroplast genome sequence diversity and genomic resources in *Diospyros*. *BMC Plant Biology* 18: 210.
- Li, W., C. Zhang, X. Guo, Q. Liu and K. Wang. 2019. Complete chloroplast genome of *Camellia japonica* genome structures, comparative and phylogenetic analysis. *PLoS ONE* 14: e0216645.
- Li, Z.-Z., B. K. Ngarega, S. Lehtonen, A. W. Gichira, M. J. Kari-chu, Q.-F. Wang and J.-M. Chen. 2020b. Cryptic diversity within the African aquatic plant *Ottelia ulvifolia* (Hydrocharitaceae) revealed by population genetic and phylogenetic analyses. *Journal of Plant Research* 133: 373–381.
- Liang, H., Y. Zhang, J. Deng, G. Gao, C. Ding, L. Zhang and R. Yang. 2020. The complete chloroplast genome sequences of 14 *Curcuma* species: Insights into genome evolution and phylogenetic relationships within zingiberales. *Frontiers in Genetics* 11: 802.
- Liu, Q., X. Li, M. Li, W. Xu, T. Schwarzacher and J. S. Heslop-Harrison. 2020a. Comparative chloroplast genome analyses of *Avena*: insights into evolutionary dynamics and phylogeny. *BMC Plant Biology* 20: 406.
- Liu, T., J. Qian, Z. Du, J. Wang and B. Duan. 2020b. Complete genome and phylogenetic analysis of *Agrimonia pilosa* Ldb. *Mitochondrial DNA Part B Resources* 5: 1435–1436.
- Liu, W., H. Kong, J. Zhou, P. W. Fritsch, G. Hao and W. Gong. 2018. Complete chloroplast genome of *Cercis chuniana* (Fabaceae) with structural and genetic comparison to six species in Caesalpinioideae. *International Journal of Molecular Sciences* 19: 1286.
- Loeuille, B., V. Thode, C. Siniscalchi, S. Andrade, M. Rossi and J. R. Pirani. 2021. Extremely low nucleotide diversity among thirty-six new chloroplast genome sequences from *Aldama* (Heliantheae, Asteraceae) and comparative chloroplast genomics analyses with closely related genera. *PeerJ* 9: e10886.
- Lowe, T. M. and P. P. Chan. 2016. tRNAscan-SE On-line: Integrating search and context for analysis of transfer RNA genes. *Nucleic Acids Research* 44: W54–W57.
- Meng, X.-X., Y.-F. Xian, L. Xiang, D. Zhang, Y.-H. Shi, M.-L. Wu, G.-Q. Dong, S.-P. Ip, Z.-X. Lin, L. Wu and W. Sun. 2018. Complete chloroplast genomes from *Sanguisorba*: Identity and variation among four species. *Molecules* 23: 2137.
- Min, J., Y. Kim, H. Xi, K.-I. Heo and J. Park. 2019a. The complete chloroplast genome of coffee tree, *Coffea arabica* L. ‘Typica’ (Rubiaceae). *Mitochondrial DNA Part B Resources* 4: 2240–2241.
- Min, J., Y. Kim, H. Xi, T. Jang, G. Kim, J. Park and J.-H. Park. 2019b. The complete chloroplast genome of a new candidate cultivar, Sang Jae, of *Abeliophyllum distichum* Nakai (Oleaceae): Initial step of *A. distichum* intraspecies variations atlas. *Mitochondrial DNA Part B Resources* 4: 3716–3718.
- Min, J., W. Kwon, H. Xi and J. Park. 2019c. The complete chloroplast genome of *Leucobryum juniperoides* (brid.) C. Müll. (Leucobryaceae, Bryophyta). *Mitochondrial DNA Part B Resources* 4: 2962–2963.
- Min, J., J. Park, Y. Kim and W. Kwon. 2019d. The complete chloroplast genome of *Artemisia fukudo* Makino (Asteraceae): Providing insight of intraspecies variations. *Mitochondrial DNA Part B Resources* 4: 1510–1512.
- Mitrenina, E., M. Skaptsov, M. Kutsev, A. Kuznetsov, H. Ikeda and A. Erst. 2020. A new diploid cytotype of *Agrimonia pilosa* (Rosaceae). *Caryologia* 73: 67–73.

- Murata, G. and K. Umemoto. 1983. Critical notes on the Asiatic *Agrimonia* with special reference to the distribution patterns of calcium oxalate crystals. *Acta Phytotaxonomica et Geobotanica* 34: 166–177.
- Myszczyński, K., A. Bączkiewicz, K. Buczkowska, M. Ślipiko, M. Szczecińska and J. Sawicki. 2017. The extraordinary variation of the organellar genomes of the *Aneura pinguis* revealed advanced cryptic speciation of the early land plants. *Scientific Reports* 7: 9804.
- Nei, M. and W.-H. Li. 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proceedings of the National Academy of Sciences of the United States of America* 76: 5269–5273.
- Neubig, K. M., W. M. Whitten, B. S. Carlswald, M. A. Blanco, L. Endara, N. H. Williams and M. Moore. 2009. Phylogenetic utility of *ycf1* in orchids: A plastid gene more variable than *matK*. *Plant Systematics and Evolution* 277: 75–84.
- Nguyen, L.-T., H. A. Schmidt, A. von Haeseler and B. Q. Minh. 2015. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution* 32: 268–274.
- Niu, Z., Q. Xue, S. Zhu, J. Sun, W. Liu and X. Ding. 2017. The complete plastome sequences of four orchid species: Insights into the evolution of the Orchidaceae and the utility of plastomic mutational hotspots. *Frontiers in Plant Science* 8: 715.
- Novikova, P. Y., T. Tsuchimatsu, S. Simon, V. Nizhynska, V. Voronin, R. Burns, O. M. Fedorenko, S. Holm, T. Säll, E. Prat., W. Marande, V. Castric and M. Nordborg. 2017. Genome sequencing reveals the origin of the allotetraploid *Arabidopsis suecica*. *Molecular Biology and Evolution* 34: 957–968.
- Oh, S.-D., S.-K. Lee, D.-W. Yun, H.-J. Sun, H.-G. Kang, H.-Y. Lee, H. Xi, J. Park and B. Lee. 2021. The complete chloroplast genome of *Zoysia macrostachya* (Poaceae): Insights into intraspecific variations and species delimitation of the *Zoysia* species. *Korean Journal of Plant Taxonomy* 51: 326–331.
- Oh, S.-H. and J. Park. 2020. The complete chloroplast genome of *Euscaphis japonica* (Thunb.) Kanitz (Staphyleaceae) isolated in Korea. *Mitochondrial DNA Part B Resources* 5: 3751–3753.
- Oh, S.-H., H. J. Suh, J. Park, Y. Kim and S. Kim. 2019a. The complete chloroplast genome sequence of a morphotype of *Goodyera schlechtendaliana* (Orchidaceae) with the column appendages. *Mitochondrial DNA Part B Resources* 4: 626–627.
- Oh, S.-H., H. J. Suh, J. Park, Y. Kim and S. Kim. 2019b. The complete chloroplast genome sequence of *Goodyera schlechtendaliana* in Korea (Orchidaceae). *Mitochondrial DNA Part B Resources* 4: 2692–2693.
- O'Kane, S. L. Jr., B. A. Schaal and I. A. Al-Shehbaz. 1996. The origins of *Arabidopsis suecica* (Brassicaceae) as indicated by nuclear rDNA sequences. *Systematic Botany* 21:559–566.
- Okuyama, Y. and M. Kato. 2009. Unveiling cryptic species diversity of flowering plants: Successful biological species identification of Asian *Mitella* using nuclear ribosomal DNA sequences. *BMC Evolutionary Biology* 9: 105.
- Park, I., W.-J. Kim, S. Yang, S.-M. Yeo, H. Li and B. C. Moon. 2017. The complete chloroplast genome sequence of *Aconitum coreanum* and *Aconitum carmichaelii* and comparative analysis with other *Aconitum* species. *PLoS ONE* 12: e0184257.
- Park, I., S. Yang, W. J. Kim, P. Noh, H. O. Lee and B. C. Moon. 2018. Complete chloroplast genome of *Sanguisorba × tenuifolia* Fisch. ex Link. *Mitochondrial DNA Part B Resources* 3: 909–910.
- Park, J.-S., D.-P. Jin, J.-W. Park and B.-H. Choi. 2019a. Complete chloroplast genome of *Fagus multinervis*, a beech species endemic to Ulleung Island in South Korea. *Mitochondrial DNA Part B Resources* 4: 1698–1699.
- Park, J., J.-H. An, Y. Kim, D. Kim, B.-G. Yang and T. Kim. 2020a. Database of National Species List of Korea: The taxonomical systematics platform for managing scientific names of Korean native species. *Journal of Species Research* 9: 233–246.
- Park, J., Y. Bae, B.-Y. Kim, G.-H. Nam, J.-M. Park, B. Y. Lee, H.-J. Suh and S.-H. Oh. 2021a. The complete chloroplast genome of *Campanula takesimana* Nakai from Dokdo Island in Korea (Campanulaceae). *Mitochondrial DNA Part B Resources* 6: 135–137.
- Park, J., Y. G. Choi, N. Yun, H. Xi, J. Min, Y. Kim and S.-H. Oh. 2019b. The complete chloroplast genome sequence of *Viburnum erosum* (Adoxaceae). *Mitochondrial DNA Part B Resources* 4: 3278–3279.
- Park, J., K.-I. Heo, Y. Kim and W. Kwon. 2019c. The complete chloroplast genome of *Potentilla fragarioides* var. *major* Maxim. *Mitochondrial DNA Part B Resources* 4: 1265–1266.
- Park, J., Y. Kim, W. Kwon, H. Xi and M. Kwon. 2019d. The complete chloroplast genome of tulip tree, *Liriodendron tulifipera* L. (Magnoliaceae): Investigation of intra-species chloroplast variations. *Mitochondrial DNA Part B Resources* 4: 2523–2524.
- Park, J., Y. Kim, W. Kwon, H. Xi and C.-H. Park. 2021b. The complete chloroplast genome sequence of new species candidate of *Plantago depressa* Willd. in Korea (Plantaginaceae). *Mitochondrial DNA Part B Resources* 6: 1961–1963.
- Park, J., Y. Kim, J. Min and S.-T. Kim. 2022a. The complete chloroplast genome of *Atriplex gmelinii* C. A. Mey. ex Bong. (Amaranthaceae). *Mitochondrial DNA Part B Resources* 7: 541–543.
- Park, J., Y. Kim and H. Xi. 2019e. The complete chloroplast

- genome sequence of male individual of Korean endemic willow, *Salix koriyanagi* Kimura ex Goerz (Salicaceae). Mitochondrial DNA Part B Resources 4: 1619–1621.
- Park, J., Y. Kim, H. Xi, M. Nho, J. Woo and Y. Seo. 2019f. The complete chloroplast genome of high production individual tree of *Coffea arabica* L. (Rubiaceae). Mitochondrial DNA Part B Resources 4: 1541–1542.
- Park, J., J. Lee and J. Park. 2022b. The investigation of intraspecific characteristics and comparative analyses of the complete mitochondrial genome of *Stegobium paniceum* (Linnaeus, 1758) (Coleoptera: Ptinidae) assembled from public NGS raw reads of the black truffle, *Tuber melanosporum*. Science Progress 105: 368504211072355.
- Park, J., S. H. Lee and J. H. Kim. 2021c. Complete genome sequence of the endosymbiotic bacterium “*Candidatus* Riesia pediculicola”. Microbiology Resource Announcements 10: e01181-20.
- Park, J., J. Min, Y. Kim and Y. Chung. 2021d. The comparative analyses of six complete chloroplast genomes of morphologically diverse *Chenopodium album* L. (Amaranthaceae) collected in Korea. International Journal of Genomics 2021: 6643444.
- Park, J., J. Min, Y. Kim, H. Xi, W. Kwon, T. Jang, G. Kim and J.-H. Park. 2019g. The complete chloroplast genome of a new candidate cultivar, Dae Ryun, of *Abeliophyllum distichum* Nakai (Oleaceae). Mitochondrial DNA Part B Resources 4: 3713–3715.
- Park, J. and S.-H. Oh. 2020. A second complete chloroplast genome sequence of *Fagus multinervis* Nakai (Fagaceae): Intraspecific variations on chloroplast genome. Mitochondrial DNA Part B Resources 5: 1868–1869.
- Park, J., S. Park, T. Jang, G. Kim and J.-H. Park. 2021e. The complete chloroplast genome of *Abeliophyllum distichum* f. *lilacinum* Nakai (Oleaceae) from the Chungbuk Province, Korea. Mitochondrial DNA Part B Resources 6: 1754–1756.
- Park, J., Y. Suh and S. Kim. 2020b. A complete chloroplast genome sequence of *Gastrodia elata* (Orchidaceae) represents high sequence variation in the species. Mitochondrial DNA Part B Resources 5: 517–519.
- Park, J. and H. Xi. 2021. Investigation of nucleotide diversity based on 17 sea cucumber mitochondrial genomes and assessment of sea cucumber mitochondrial gene markers. Advances in Oceanography & Marine Biology 2: 2021.
- Park, J. and H. Xi. 2022. The complete chloroplast genome of *Campsis grandiflora* (Bignoniaceae). Korean Journal of Plant Taxonomy 52: 156–172.
- Park, J., H. Xi and Y. Kim. 2020c. The complete chloroplast genome of *Arabidopsis thaliana* isolated in Korea (Brassicaceae): An investigation of intraspecific variations of the chloroplast genome of Korean *A. thaliana*. International Journal of Genomics 2020: 3236461.
- Park, J., H. Xi and Y. Kim. 2021f. The complete mitochondrial genome of *Arabidopsis thaliana* (Brassicaceae) isolated in Korea. Korean Journal of Plant Taxonomy. 51: 176–180.
- Park, J., H. Xi, Y. Kim, K.-I. Heo, M. Nho, J. Woo, Y. Seo and J. H. Yang. 2019h. The complete chloroplast genome of cold hardiness individual of *Coffea arabica* L. (Rubiaceae). Mitochondrial DNA Part B Resources 4: 1083–1084.
- Park, J., H. Xi, Y. Kim and M. Kim. 2021g. Complete genome sequence of *Lentilactobacillus parabuchneri* strain KEM. Microbiology Resource Announcements 10: e01208-20.
- Park, J., H. Xi, Y. Kim and S.-T. Kim. 2022c. The complete chloroplast genome of *Utricularia tenuicaulis* Miki (Lentibulariaceae) isolated in Korea. Mitochondrial DNA Part B Resources 7: 1143–1145.
- Park, J., H. Xi, Y. Kim, S. Nam and K.-I. Heo. 2020d. The complete mitochondrial genome of new species candidate of *Rosa rugosa* (Rosaceae). Mitochondrial DNA Part B Resources 5: 3435–3437.
- Park, J., H. Xi and S.-H. Oh. 2020e. Comparative chloroplast genomics and phylogenetic analysis of the *Viburnum dilatatum* complex (Adoxaceae) in Korea. Korean Journal of Plant Taxonomy 50: 8–16.
- Park, J., H. Xi and J. Park. 2021h. Complete genome sequence of a *Blochmannia* endosymbiont of *Colobopsis nipponica*. Microbiology Resource Announcements 10: e01195-20.
- Park, J., H. Xi, J. Park, S. J. Nam and Y.-D. Lee. 2020f. Complete genome sequence of the *Blochmannia* endosymbiont of *Campotonotus nipponensis*. Microbiology Resource Announcements 9: e00703-20.
- Park, J., H. Xi, J. Park and B. Y. Seo. 2021i. A new mitochondrial genome of *Sogatella furcifera* (Horváth) (Hemiptera: Delphacidae) and mitogenome-wide investigation on polymorphisms. Insects 12: 1066.
- Park, J., N. Yun and S.-H. Oh. 2019i. The complete chloroplast genome of an endangered species in Korea, *Halenia corniculata* (L.) Cornaz (Gentianaceae). Mitochondrial DNA Part B Resources 4: 1539–1540.
- Park, S.-H., Y.-B. Sim, Y.-J. Kang, J.-K. Lee, S.-S. Lim and H.-W. Suh. 2012. Effect of *Agrimonia pilosa* Ledeb extract on the antinociception and mechanisms in mouse. The Korean Journal of Physiology & Pharmacology 16: 119–123.
- Park, S. 2020. A phylogenomic study of Magnoliaceae and its evolutionary implications. PhD dissertation, Sungshin Womens University, Seoul, Korea, 283 pp.
- Potter, D., T. Eriksson, R. C. Evans, S. Oh, J. E. E. Smedmark, D. R. Morgan, M. Kerr, K. R. Robertson, M. Arsenault, T. A. Dickinson and C. S. Cambell. 2007. Phylogeny and classifi-

- cation of Rosaceae. *Plant Systematics and Evolution* 266: 5–43.
- Roberts, R. J., M. O. Carneiro and M. C. Schatz. 2013. The advantages of SMRT sequencing. *Genome Biology* 14: 405.
- Robertson, K. R. 1974. The genera of Rosaceae in the southeastern United States. *Journal of the Arnold Arboretum* 55: 344–401.
- Sawicki, J., A. Bączkiewicz, K. Buczkowska, P. Górski, K. Krawczyk, P. Mizia, K. Myszczyński, M. Ślipiko and M. Szczecińska. 2020. The increase of simple sequence repeats during diversification of Marchantiidae, an early land plant lineage, leads to the first known expansion of inverted repeats in the evolutionarily-stable structure of liverwort plastomes. *Genes* 11: 299.
- Seo, H. S., B. H. Chung and Y. G. Cho. 2008. Antioxidant and anticancer effects of agrimony (*Agrimonia pilosa* L.) and Chinese Lizardtail (*Saururus chinensis* Baill). *Korean Journal of Medicinal Crop Science* 16: 139–143.
- Shin, W.-J., K.-H. Lee, M.-H. Park and B.-L. Seong. 2010. Broad-spectrum antiviral effect of *Agrimonia pilosa* extract on influenza viruses. *Microbiology and Immunology* 54: 11–19.
- Skalický, V. 1962. Ein Beitrag zur Erkenntnis der europäischen Arten der Gattung *Agrimonia*. *Acta Horti Botanici Pragensis* 1962: 87–108.
- Skalický, V. 1973. Amerikanische Arten der Gattung *Agrimonia* L. ser. Tuberosae ser. nova. *Folia Geobotanica and Phytotaxonomica* 8: 95–104.
- Suh, H.-J., J. Min, J. Park and S.-H. Oh. 2021. The complete chloroplast genome of *Aruncus dioicus* var. *kamtschaticus* (Rosaceae). *Mitochondrial DNA Part B Resources* 6: 1256–1258.
- Van Dijk, E. L., H. Auger, Y. Jaszczyszyn and C. Thermes. 2014. Ten years of next-generation sequencing technology. *Trends in Genetics* 30: 418–426.
- Wallroth, C. F. W. 1842. Monographischer Versuch über die Gewächs-Gattung *Agrimonia* Celsi. *Beiträge zur Botanik* (Wallroth) 1: 1–66.
- Wang, C.-L., M.-Q. Ding, C.-Y. Zou, X.-M. Zhu, Y. Tang, M.-L. Zhou and J.-R. Shao. 2017. Comparative analysis of four buckwheat species based on morphology and complete chloroplast genome sequences. *Scientific Reports* 7: 6514.
- Wang, H., S.-Y. Park, S.-H. Song, M.-L. San, Y.-C. Kim, T.-H. Ham, D.-Y. Kim, T.-S. Kim, J. Lee and S.-W. Kwon. 2020. Analysis of complete chloroplast genome sequence of Korean landrace *Cymbidium goeringii*. *3 Biotech* 10: 29.
- World Health Organization. 1998. *Medicinal Plants in the Republic of Korea: Information on 150 Commonly Used Medicinal Plants*. WHO Regional Office for the Western Pacific, Manila, 316 pp.
- Xiang, S., X.-D. Liu, W.-H. Sun, S.-R. Lan, Z.-J. Liu and S.-Q. Zou. 2019. The complete chloroplast genome sequence of *Euscaphis japonica* (Staphyleaceae). *Mitochondrial DNA Part B Resources* 4: 3484–3485.
- Xie, D.-F., H.-X. Yu, M. Price, C. Xie, Y.-Q. Deng, J.-P. Chen, Y. Yu, S.-D. Zhou and X.-J. He. 2019. Phylogeny of Chinese *Allium* species in section *Daghestanica* and adaptive evolution of *Allium* (Amaryllidaceae, Allioidae) species revealed by the chloroplast complete genome. *Frontiers in Plant Science* 10: 460.
- Xu, Z., T. Xin, D. Bartels, Y. Li, W. Gu, H. Yao, S. Liu, H. Yu, X. Pu, J. Zhou, J. Xu, C. Xi, H. Lei, J. Song and S. Chen. 2018. Genome analysis of the ancient tracheophyte *Selaginella tamariiscina* reveals evolutionary features relevant to the acquisition of desiccation tolerance. *Molecular Plant* 11: 983–994.
- Yang, J., K. Takayama, J.-S. Youn, J.-H. Pak and S.-C. Kim. 2020. Plastome characterization and phylogenomics of East Asian beeches with a special emphasis on *Fagus multinervis* on Ulleung Island, Korea. *Genes* 11: 1338.
- Yang, L., Y. Teng, S.-F. Gong, Z.-M. Feng and Y. Zhang. 2021. The complete chloroplast genome of *Agrimonia pilosa* var. *nepalensis* (D. Don) Nakai. *Mitochondrial DNA Part B Resources* 6: 938–940.
- Yang, X.-F., Y.-T. Wang, S.-T. Chen, J.-K. Li, H.-T. Shen and F.-Q. Guo. 2016. PBR1 selectively controls biogenesis of photosynthetic complexes by modulating translation of the large chloroplast gene *Ycf1* in *Arabidopsis*. *Cell Discovery* 2: 16003.
- Yoo, S.-C., S.-H. Oh and J. Park. 2021. Phylogenetic position of *Daphne genkwa* (Thymelaeaceae) inferred from complete chloroplast data. *Korean Journal of Plant Taxonomy* 51: 171–175.
- Yoo, S.-C., J. Park and S.-H. Oh. 2023. Complete chloroplast genome of *Daphne pseudomezereum* var. *koreana* (Thymelaeaceae). *Mitochondrial DNA Part B Resources* 8: 305–309.
- Yoon, J.-Y., S.-Y. Lee, H.-J. Jun and J.-Y. Lee. 2012. Anti-aging effects of solvent fraction from *Agrimonia pilosa* L. extracts. *Journal of Applied Biological Chemistry* 55: 35–39. (in Korean)
- Yuan, Y., X. Jin, J. Liu, X. Zhao, J. Zhou, X. Wang, D. Wang, C. Lai, W. Xu, J. Huang, L. Zha, D. Liu, X. Ma, L. Wang, M. Zhou, Z. Jiang, H. Meng, H. Peng, Y. Liang, R. Li, C. Jiang, Y. Zhao, T. Nan, Y. Jin, Z. Zhan, J. Yang, W. Jiang and L. Huang. 2018. The *Gastrodia elata* genome provides insights into plant adaptation to heterotrophy. *Nature Communications* 9: 1615.
- Zerbino, D. R. and E. Birney. 2008. Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. *Genome Research* 18: 821–829.
- Zhang, S.-D., J.-J. Jin, S.-Y. Chen, M. W. Chase, D. E. Soltis, H.-T. Li, J.-B. Yang, D.-Z. Li and T.-S. Yi. 2017. Diversification of Rosaceae since the Late Cretaceous based on plastid phylog-

- enomics. *New Phytologist* 214: 1355–1367.
- Zhang, W., Y. Zhao, G. Yang, J. Peng, S. Chen and Z. Xu. 2019a. Determination of the evolutionary pressure on *Camellia oleifera* on Hainan Island using the complete chloroplast genome sequence. *PeerJ* 7: e7210.
- Zhang, Z., X. Jiang, Y. Chen, P. Zhu, L. Li, Y. Zeng and T. Tang. 2019b. Characterization of the complete chloroplast genome sequence of *Conyza canadensis* and its phylogenetic implications. *Mitochondrial DNA Part B Resources* 4: 2028–2030.
- Zhao, Q.-Y., Wang, Y., Kong, Y.-M., Luo, D., Li, X. and Hao, P. 2011. Optimizing *de novo* transcriptome assembly from short-read RNA-Seq data: A comparative study. *BMC Bioinformatics* 12: S2.
- Zhou, X., L. Ren, Q. Meng, Y. Li, Y. Yu and J. Yu. 2010. The next-generation sequencing technology and application. *Protein & Cell* 1: 520–536.