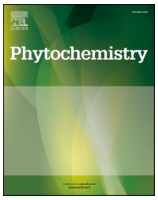
Phytochemistry 181 (2021) 112576

Contents lists available at ScienceDirect

Phytochemistry

journal homepage: www.elsevier.com/locate/phytochem



Assessing the genetic and chemical diversity of *Taraxacum* species in the Korean Peninsula

Yun Sun Lee [a,1,3], Jinkyung Kim [a,1,4], Sunmin Woo [b,1,5], Jee Young Park [a], Hyun-Seung Park [a], Hyeonah Shim [a], Hong-Il Choi [c], Jung Hwa Kang [d], Taek Joo Lee [d], Sang Hyun Sung [b,2], Tae-Jin Yang [a,**], Kyo Bin Kang [b,e,*]

[a] Department of Agriculture, Forestry and Bioresources, Plant Genomics and Breeding Institute, College of Agriculture and Life Sciences, Seoul National University, Seoul, 08826, Republic of Korea
[b] College of Pharmacy and Research Institute of Pharmaceutical Sciences, Seoul National University, Seoul, 08826, Republic of Korea
[c] Advanced Radiation Technology Institute, Korea Atomic Energy Research Institute, Jeongeup, 56212, Republic of Korea
[d] Hantaek Botanical Garden, Yongin, 17183, Republic of Korea
[e] Research Institute of Pharmaceutical Sciences, College of Pharmacy, Sookmyung Women's University, Seoul, 04310, Republic of Korea

ARTICLE INFO

Keywords:
Taraxacum
Asteraceae
Chloroplast genome
45S rDNA
Metabolomics
LC–MS
Specialized metabolites

ABSTRACT

The genetic relationship between *Taraxacum* species, also known as the dandelion, is complicated because of asexual and mixed sexual apomictic reproduction. The usage of *Taraxacum* species in traditional medicines make their specialized metabolism important, but interspecific chemical difference has rarely been reported for the genus. In this study, we assembled the chloroplast genome and 45S rDNA of six *Taraxacum* species that occur in Korea (*T. campylodes*, *T. coreanum*, *T. erythrospermum*, *T. mongolicum*, *T. platycarpum*, and *T. ussuriense*), and performed a comparative analysis, which revealed their phylogenetic relationships and possible natural hybridity. We also performed a liquid chromatography–mass spectrometry-based phytochemical analysis to reveal interspecific chemical diversity. The comparative metabolomics analysis revealed that *Taraxacum* species could be separated into three chemotypes according to their major defensive specialized metabolites, which were the sesquiterpene lactones, the phenolic inositols, and chlorogenic acid derivatives. The CP DNA- and 45S rDNA-based phylogenetic trees showed a tangled relationship, which supports the notion of ongoing hybridization of wild *Taraxacum* species. The untargeted LC–MS analysis revealed that each *Taraxacum* plant exhibits species-specific defensive specialized metabolism. Moreover, 45S rDNA-based phylogenetic tree correlated with the hierarchical cluster relied on metabolite compositions. Given the coincidence between these analyses, we represented that 45S rDNA could well reflect overall nuclear genome variation in *Taraxacum* species.

1. Introduction

The genus *Taraxacum* (Asteraceae), generally known as the dandelion, is characterized by complex genetic diversity. Hybridity, frequent polyploidy and widespread agamospermous reproduction prevail in *Taraxacum* species (Kirschner et al., 2015). This genetic complexity causes taxonomic complexity in the genus; some botanists classify the entire genus into about 60 macrospecies, while others divide it into 2000

Abbreviations: CP, chloroplast; CID, collision-induced dissociation; LC–MS, liquid chromatography–mass spectrometry; PCA, principal component analysis; OPLS-DA, orthogonal partial least squares discriminant analysis.

* Corresponding author. Research Institute of Pharmaceutical Sciences, College of Pharmacy, Sookmyung Women's University, Seoul, 04310, Republic of Korea.
** Corresponding author. Department of Agriculture, Forestry and Bioresources, Plant Genomics and Breeding Institute, College of Agriculture and Life Sciences, Seoul National University, Seoul, 08826, Republic of Korea.
E-mail addresses: tjyang@snu.ac.kr (T.-J. Yang), kbkang@sookmyung.ac.kr (K.B. Kang).

[1] These authors contributed equally to this work.
[2] Deceased on July 24th, 2018.
[3] Current address: Department of Biochemistry and Molecular Biology, Michigan State University, East Lansing, MI 48824, USA
[4] Current address: Research and Development Center for Life Science, Seegene Inc., Seoul, 05548, Republic of Korea
[5] Current address: Research Institute of Pharmaceutical Sciences, College of Pharmacy, Sookmyung Women's University, Seoul, 04310, Republic of Korea

https://doi.org/10.1016/j.phytochem.2020.112576
Received 25 September 2020; Received in revised form 19 October 2020; Accepted 24 October 2020

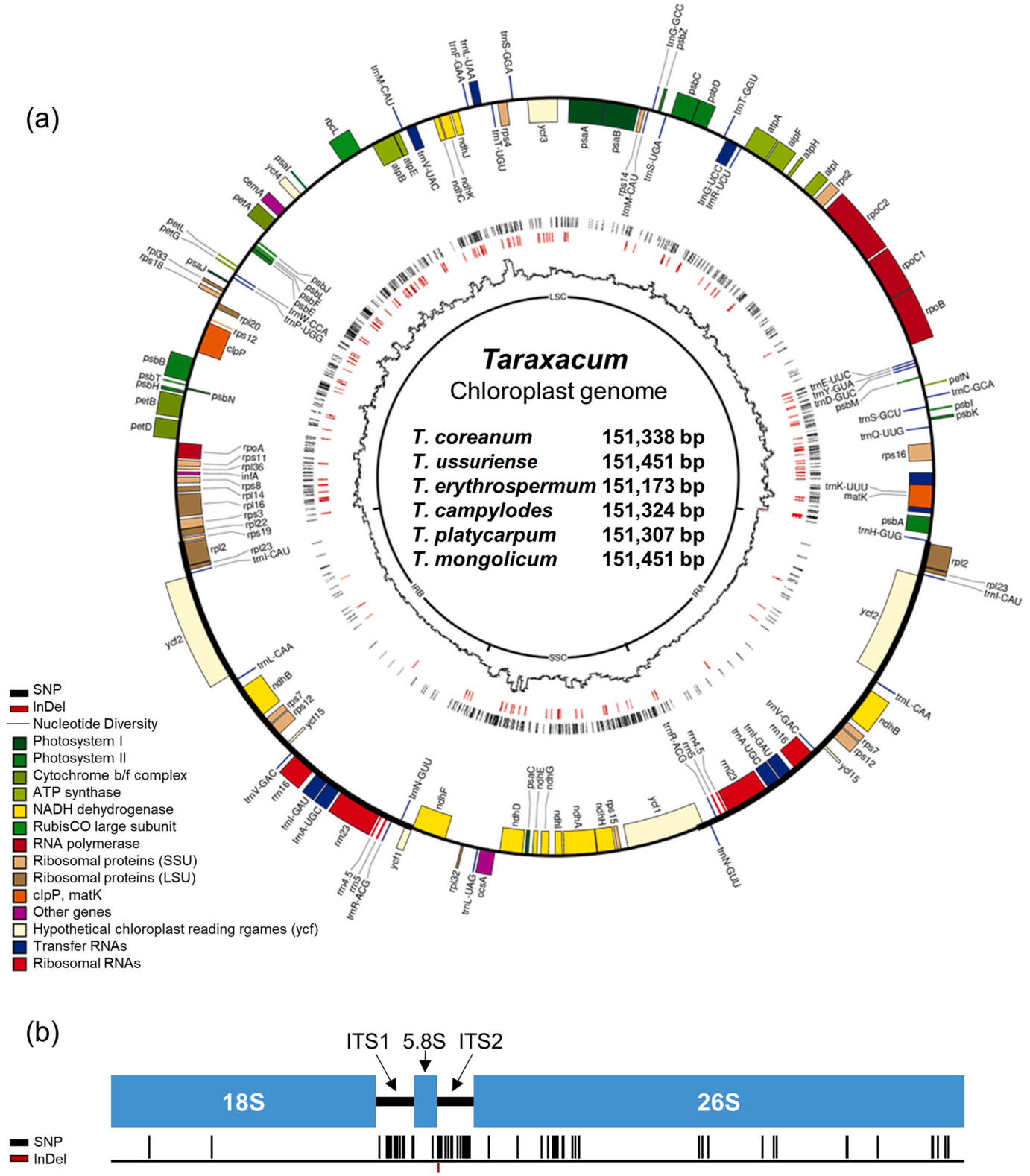


Fig. 1. Maps of the complete chloroplast genomes and 45S rDNA units of six *Taraxacum* species and the locations of polymorphic regions. The number of variable regions for each species is in Table 2 for the chloroplast genomes and Table 3 for the 45S rDNA. (A) The chloroplast genome map of the six *Taraxacum* species. Polymorphic regions among the species are indicated inside the circle in black and red bars for SNP and InDel, respectively. The innermost graph is Pi value representing nucleotide diversity. (B) The 45S rDNA coding region of the six *Taraxacum* species. SNPs and IndeLs among the six *Taraxacum* species are represented in blue and red, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

Table 1
Summary of NGS amounts and the assembly of the complete chloroplast genome and 45S rDNA of six *Taraxacum* species.

	Genome information	Species[a]					
		T. campylodes	*T. coreanum*	*T. erythrospermum*	*T. mongolicum*	*T. platycarpum*	*T. ussuriense*
Filtered NGS Reads (Mbp)		875	902	2374	718	1092	1191
CP Assembly	Coverage[b] (X)	296.54	458.78	1238.61	196.51	416.26	397.80
	Total Length (bp)	151,324	151,338	151,173	151,451	151,307	151,451
	LSC (bp)	83,895	83,914	83,812	84,052	83,922	84,019
	IRA/B (bp)	24,400	24,431	24,420	24,429	24,439	24,466
	SSC (bp)	18,549	18,562	18,521	18,541	18,507	18,500
45S rDNA Assembly	Coverage[b] (X)	619	305	5832	556	472	698
	Total Length (bp)	5826	5826	5826	5.825	5826	5826
	18S	1810	1810	1810	1810	1810	1810
	5.8S	159	159	159	159	159	159
	26S	3349	3349	3349	3349	3349	3349

[a] Indicating that CPs and 45S rDNA were reported in Kim et al. (2016a); Kim et al. (2016b).
[b] Coverage indicates the total WGS read depth for the complete genomes.

Table 2
Information of SNP and InDel in chloroplast genomes of six *Taraxacum* species.

	T. campylodes	*T. coreanum*	*T. erythrospermum*	*T. mongolicum*	*T. platycarpum*	*T. ussuriense*
T. campylodes		6	48	80	43	80
T. coreanum	24		50	79	42	79
T. erythrospermum	240	248		76	46	56
T. mongolicum	414	424	429		80	48
T. platycarpum	207	196	285	458		471
T. ussuriense	446	435	463	200	84	

The upper triangle shows the number of InDels, while the lower triangle indicates the total nucleotide substitutions.

Table 3
Information of SNP and InDel in 45S rDNA genomes of six *Taraxacum* species.

	T. campylodes	*T. coreanum*-NR1	*T. coreanum*-NR2	*T. erythrospermum*	*T. mongolicum*	*T. platycarpum*	*T. ussuriense*-NR1*	*T. ussuriense*-NR2*
T. campylodes		0	0	0	1	0	0	0
T. coreanum-NR1*	0		0	0	1	0	0	0
T. coreanum-NR2*	4	4		0	1	0	0	0
T. erythrospermum	0	0	4		1	0	0	0
T. mongolicum	47	47	51	47		1	1	1
T. platycarpum	23	23	27	23	58		3	2
T. ussuriense-NR1*	26	26	30	26	61	0		0
T. ussuriense-NR2*	23	23	27	23	58	0	5	

The upper triangle shows the number of InDels, while the lower triangle indicates the total nucleotide substitutions. **T. coreanum* and *T. ussuriense* have heterotypes of nrDNA sequence, which are indicated as *T. coreanum*-NR-1 and -2, and *T. ussuriense*-NR-1 and -2, respectively.

microspecies, which are further classified into 34 sections (Richards, 1970). The leaves and buds of dandelions are edible and have been eaten for over a thousand years (Guarrera and Savo, 2016). Dandelions are also important in traditional medicines, and have been used for treating infections, spleen and liver disorders, and diarrhea (Schütz et al., 2006). One species, *Taraxacum kok-saghyz* L.E.Rodin, was recently identified as an industrially important species because of its potential for rubber production (Stolze et al., 2017). Despite their importance and value, the genetic and chemical diversity of many *Taraxacum* species remain unknown.

Chloroplast (CP) is the subcellular organelle that are maternally inherited in angiosperm (Daniell et al., 2016). The chloroplast genome

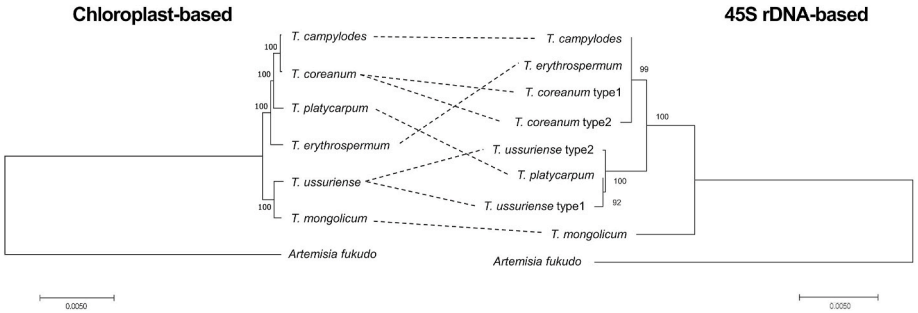


Fig. 2. The phylogenomic relationship of *Taraxacum* species based on chloroplast (CP) and 45S rDNA. The phylogenetic position of each species was analyzed using 74 protein coding sequences from the CP genomes and 45S rDNA transcription unit sequences, along with previously reported CP sequences and 45S rDNA (Genbank accession number: MT003979) of *Artemisia fukudo* (Asteraceae) (Whitehead and Bowers, 2013).

mainly consists of quadripartite structure, such as one large single copy, one small single copy, and two inverted repeat regions (Palmer, 1985). Several mutational events including inversions, oligonucleotide repeats, microstructural changes, InDel, and substitution are mainly occurred in the CP genomes (Jheng et al., 2012). Since the high conservation of CP during the evolution, the mutational events allow to develop CP genome-based markers that can be exploited for the plant barcoding (Park et al., 2020) and complex taxonomical classification (Abdullah et al., 2020; Henriquez et al., 2020a, 2020b). Recent advances in next-generation sequencing technology and newly developed *de novo* assembly of low coverage whole genome shotgun (dnaLCW) methods enable complete CP genome sequencing to be used to understand genetic diversity and evolutionary information (Kim et al., 2015). A previous study has shown that complete CP genome sequencing can discriminate *T. kok-saghyz* from its weedy relatives, *Taraxacum officinale* Weber ex F.H. Wigg and *Taraxacum brevicorniculatum* Korol. (Zhang et al., 2017). Another study revealed that genotypic diversity in the plastomes of three apomictic triploids from *T. officinale*, *Taraxacum obtusifrons* Markl, *Taraxacum stridulum* Trávníček ined, and *Taraxacum amplum* Markl. could be identified by sequencing their complete CP genomes (Salih et al., 2017). However, much of the genetic diversity in this genus remains unexplored.

Specialized metabolites, also known as secondary metabolites, are small molecules produced by higher plants for chemical defense, reproduction, and ecological interactions (Hartmann, 2007). Up to date, more than 200,000 phytochemicals have been isolated from plants and identified (Hartmann, 2007; Wink, 2003). The chemical diversity of plant metabolites is one of phenotypic results of evolution, which means the distribution of specialized metabolites could provide us with insights into the evolution and taxonomy of different plant species (Wink, 2010). Liquid chromatography–tandem mass spectrometry (LC–MS/MS) is considered the most advanced method for profiling the chemical diversity of semipolar specialized metabolites in plants (Ernst et al., 2014). LC–MS/MS analyses provide separation of metabolites followed by relative quantification based on MS1 signal intensities and putative annotation based on MS/MS fragmentation (Nakabayashi and Saito, 2013). Thus, LC–MS/MS-based metabolite profiling is an effective tool for chemotaxonomic studies, which can give insights into taxonomic relationships between genetically diverse and complex species (Martucci et al., 2014). There were a few studies that describe LC–MS/MS-based chemical profiling of *Taraxacum* species (Chen et al., 2012; Huber et al., 2016; Kenny et al., 2014; Schütz et al., 2005). However, most of these studies focused on certain classes of metabolites or a single *Taraxacum* species. Mingarro et al. performed a comparative LC–MS-based study on five South European *Taraxacum* species (*Taraxacum obovatum* (Willd.) Dc, *Taraxacum marginellum* H.Lindb, *Taraxacum hispanicum* H.Lindb, *Taraxacum lambinonii* Soest, and *Taraxacum lacistrum* Sahlin); however, this study was mainly focused on the effect of different polyphenol contents on pharmaceutical activities, and did not discuss on the interspecific differences in detail (Mingarro et al., 2015).

About a dozen *Taraxacum* species live in the wild on the Korean peninsula. Among these, *Taraxacum campylodes* G.E.Haglund (previously known as *T. officinale*) and *Taraxacum erythrospermum* Andrz. Ex Besser (also known as *T. laevigatum* (Willd. DC.)) are exotic species from Europe, and are currently considered as some of the most problematic invasive species in Korean ecosystems. In this study, we collected these two introduced species and four other *Taraxacum* species (*Taraxacum coreanum* Nakai, *Taraxacum mongolicum* Hand.-Mazz, *Taraxacum platycarpum* Dahlst, and *Taraxacum ussuriense* Kom. (also known as *T. ohwianum* Kitam.)), which are native in Korea and other Asian countries (Ryu et al., 2017). Here, we investigated and report on the phylogenetic relationship and phytochemical variation between these species to expand our knowledge on the genetics and chemistry of *Taraxacum*, by applying chloroplast and 45S rDNA genome sequences-based genomics, and LC–MS/MS-based untargeted metabolomic analyses.

2. Results

2.1. Assembly of chloroplast genomes and 45S rDNA, and comparative analysis in Taraxacum species

Previously we reported the complete chloroplast genomes of three *Taraxacum* species: *T. campylodes* (NC 030772.1), *T. platycarpum* (NC 031395.1), and *T. mongolicum* (NC 031396.1) (Kim et al., 2016a, 2016b). In this study, we additionally completed the *de novo* assembly of chloroplast genome sequences from three more species (*T. coreanum*, *T. ussuriense*, and *T. erythrospermum*) and of the 45S rDNA from six *Taraxacum* species (Fig. 1; Table 1; and Supplementary Table 1). The complete chloroplast sequences of three newly assembled species had lengths from 151,173 to 151,451 bp, which was similar to those of three previously reported species (151,307–151,451 bp) (Kim et al., 2016a, 2016b). The chloroplast genomes of *T. coreanum*, *T. ussuriense*, and *T. erythrospermum* showed a typical quadripartite structure with a pair of inverted repeats (24,420–24,466 bp), a large single copy region (83,812–84,019 bp), and a short single copy region (18,500–18,562 bp). The

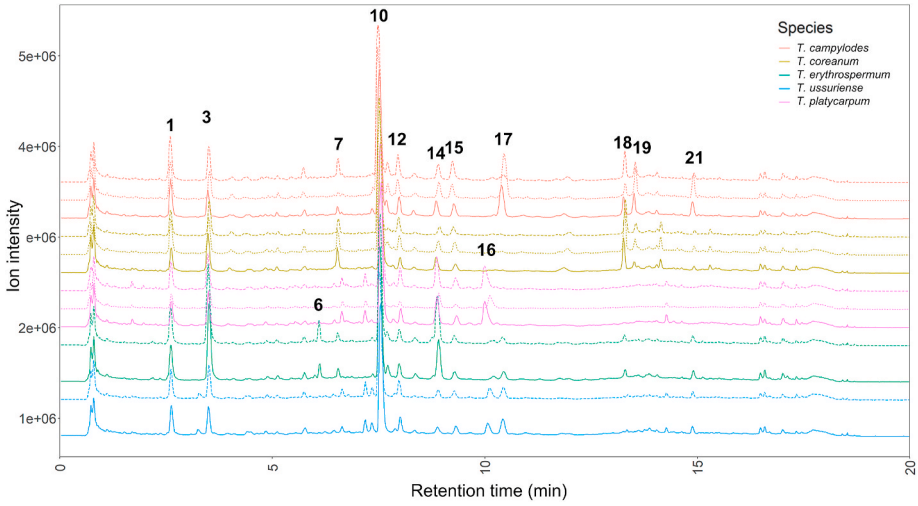


Fig. 3. LC–MS base peak ion (BPI) chromatograms of 13 *Taraxacum* extracts. The selected major chromatographic peaks are annotated with peak numbers. *T. campylodes*, *T. coreanum*, and *T. platycarpum* were triplicated, while *T. erythrospermum* and *T. ussuriense* were duplicated. Gaps between the chromatograms were added to help visualize the differences, so the y-axis values do not equal the absolute intensities.

45S rDNA consisted of 18S, ITS1, 5.8S, ITS2, and 26S transcription units in the six *Taraxacum* species, and was 5826 bp in length with 18S rRNA (1810 bp), 5.8S rRNA (159 bp), and 26S rRNA (3349 bp).

The six *Taraxacum* species had 112 genes in the same order, which consisted of 79 protein-coding genes, 29 transfer RNA (tRNA) genes, and four ribosomal RNA (rRNA) genes (Supplementary Table 2). We performed multiple alignments to compare the chloroplast genome sequences of the six *Taraxacum* species. The comparative genome analysis revealed that non-coding regions showed higher sequence variation than the coding regions, with high sequence diversity in the three intergenic regions *trnM(CAU)-aptE*, *rbcL-psaI*, and *trnL(UAG)-ccsA* (Fig. 1A; Supplementary Figure 1). A comparison of the SNP or InDel regions in the six *Taraxacum* species showed that the lowest sequence variation was between *T. campylodes* and *T. coreanum* (six InDels and 24 SNPs), while the highest sequence variation was between *T. platycarpum* and *T. ussuriense* (84 InDels and 471 SNPs) (Table 2). We also analyzed and compared the 45S rDNA from the six *Taraxacum* species (Fig. 1B). *T. coreanum* and *T. ussuriense* had heterogeneous types of 45S rDNA (Table 3). We specifically examined the heterotypes referred to as *T. coreanum*-NR-1 and *T. coreanum*-NR-2, and *T. ussuriense*-NR-1 and *T. ussuriense*-NR-2, respectively. These heterotypic 45S rDNAs had the following sequence variations: *T. coreanum* had four nucleotide variants and *T. ussuriense* had 5 nucleotide variations. The ITS1 and ITS2 regions were the most variable regions with the number as well as the density. 26S rDNA also showed comparable number of polymorphic sites with ITS regions but the density was low and 18S rDNA region was the most conserved with only 2 SNPs. Although *T. coreanum* and *T. ussuriense* had heterotypes of the 45S rDNA, the sequence comparison analysis of the six species revealed that the 45S rDNAs had high sequence similarity. A total of 71 SNPs and one InDel were identified across all *Taraxacum* species, among which *T. platycarpum* and *T. ussuriense*-NR-1 had the lowest sequence variation with two substitutions, while *T. mongolicum* and *T. ussuriense*-NR-2 had the highest sequence variation with 61 substitutions.

A phylogenetic analysis was conducted using 74 protein coding sequences from CP genomes and 45S rDNA, and the taxonomic relationships between the six *Taraxacum* species were compared based on the phylogenetic trees produced (Fig. 2). A phylogenetic analysis based on the whole CP genome was also performed, and it gave a tree of which topology was same to one of the 74 protein coding sequences-based tree (Supplementary Figure 2).

2.2. The tentative identification of specialized metabolites

The LC–MS analyses on the whole parts of five *Taraxacum* species (excluding *T. mongolicum* due to insufficient sample material) were performed in the negative ion mode using electrospray ionization (ESI−). Compared to the positive ion mode (ESI+), the ESI− provided better sensitivity and lower detection limits, as reported previously (Liigand et al., 2017). *Taraxacum* species are known to contain various phenolics and sesquiterpene lactones (Martinez et al., 2015), so the formation of anions was preferred because of the carboxyl and phenyl groups in these compounds. The stacked chromatogram showed that the specialized metabolites of the five tested species were qualitatively similar (Fig. 3).

The MS^E method (Plumb et al., 2006) allowed us to acquire high-energy collision-induced dissociation (CID) MS (pseudo tandem MS) data, which supported the tentative identification of the major chromatographic peaks. Many peaks were cautiously identified by comparing their relative retention time, UV absorption spectra, and high-energy CID MS spectra to the literature, as shown in Table 4 (the MS^E pseudo-fragmentation spectra and chemical structures of selected compounds are shown in Supplementary Figure 3). The investigated metabolites belonged to various classes, including mono- and di-isomers of caffeoylquinic acids, flavonoid glycosides, sesquiterpene lactones, and phenolic inositols, as described in the literature (Chen et al., 2012; Huber et al., 2016; Kenny et al., 2014; Mingarro et al., 2015; Schütz et al., 2005).

2.3. Specialized metabolite diversity between the five Taraxacum species

A total of 5283 MS features were acquired from the entire LC–MS

Table 4
The putative identification of metabolites in *Taraxacum* extracts analyzed by LC–MS^E.

Peak no.	Identity	t_R (min)	Observed *m/z* (Da)	Molecular formula	Calculated *m/z* (Da)[b]	UV λ_{max} (nm)	Key MS^E fragment ions (Da)
1	*cis*-caftaric acid[a]	2.63	311.0400	$C_{13}H_{12}O_9$	311.0403		179, 149, 135, 133
2		3.33	359.0808	$C_{11}H_{20}O_3$	359.0826		
3	chlorogenic acid[a]	3.51	353.0810	$C_{16}H_{18}O_9$	353.0873	243, 327	191, 135, 85
4		4.85	305.0668	$C_{15}H_{14}O_7$	305.0661		
5	5-*O*-feruloylquinic acid	6.13	367.1032	$C_{17}H_{20}O_9$	387.1029	246, 282, 325	
6	luteolin dihexoside	6.64	609.1445	$C_{27}H_{30}O_{16}$	609.1456	246, 328	469, 285
7	austricin 8-O-*β*-D-glucopyranoside[a]	7.18	469.1700	$C_{21}H_{28}O_9$	469.1710[c]	248, 324	427, 343, 161
8	luteolin 7-*O*-rutinoside[a]	7.33	593.1507	$C_{27}H_{30}O_{15}$	593.1506		447, 285, 179, 151, 135
9	sesquiterpene lactone glycoside	7.35	467.1562	$C_{22}H_{28}O_{11}$	467.1533		
10	chicoric acid[a]	7.54	473.0709	$C_{22}H_{18}O_{12}$	473.0720	223, 244, 328	293, 179, 153, 133, 112
11	luteolin 7-*O*-glucoside[a]	7.71	447.0927	$C_{21}H_{20}O_{11}$	447.0927		285
12		8.00	473.0732				
13		8.33	505.0988	$C_{23}H_{22}O_{13}$	505.0982		
14	3,5-di-*O*-caffeoylquinic acid[a]	8.91	515.1183	$C_{25}H_{24}O_{12}$	515.1190	249, 325	353, 191, 179, 135
15		9.32	457.0779	$C_{22}H_{18}O_{11}$	457.0771		
16		10.07	343.0848				235, 122
17	taraxinic acid	10.43	261.1125	$C_{15}H_{18}O_4$	261.1127	248, 325	
18	1D-4,5,6-tri-*O*-*p*-hydroxyphenylacetyl-chiro-inositol[a]	13.26	581.1651	$C_{30}H_{30}O_{12}$	581.1659		429, 295, 161, 151, 133
19	1D-1,2,3-tri-*O*-*p*-hydroxyphenylacetyl-chiro-inositol[a]	13.51	581.1664	$C_{30}H_{30}O_{12}$	581.1659		429, 295, 161, 151, 133
20		14.32	453.2123	$C_{23}H_{34}O_9$	453.2125		
21		14.89	545.2154	$C_{32}H_{34}O_8$	545.2175		261, 217, 173

[a] Identifications were confirmed by comparing t_R and MS spectra to standard compounds.
[b] $[M–H]^-$ except peak **7**.
[c] $[M + HCOOH–H]^-$.

profiles of the five *Taraxacum* species. A principal component analysis (PCA) was applied to analyze the chemical diversity between the tested *Taraxacum* species. At first, the PCA model was built using every tested and quality control (QC) sample (a mixture of every analyzed samples). In the PC (principal component)1-PC2 score plot of this PCA model, every QC sample was placed at the center of the score plot, which validated our experimental method (Supplementary Figure 4). Further multivariate analyses were performed without the QC samples to maximize the variance between samples. Another PCA model was built using the 13 tested samples; the PC1-PC2 score plot from this analysis is shown in Fig. 4A. In this model, PC1 and PC2 accounted for 32.1% and 26.3% of the total variance, respectively (Supplementary Figure 5). Two clusters were formed in the score plot, the first comprising *T. campylodes* and *T. coreanum*, and the second containing *T. ussuriense* and *T. platycarpum* (Fig. 4A). *T. erythrospermum* was relatively more different from the other species. The PC1-PC2 loading plot revealed that the separation of *T. erythrospermum* from the other species was caused by the relatively high concentration of phenolic acid derivatives, such as chlorogenic acid (**3**) and 3,5-di-*O*-caffeoylquinic acid (**14**), and a relatively low concentration of chicoric acid (**10**). On the other hand, *T. campylodes* and *T. coreanum* showed relatively higher concentrations of 1D-4,5,6-tri-*O*-*p*-hydroxyphenylacetyl-*chiro*-inositol (**18**) (Fig. 4B). An unknown metabolite (**16**) was abundant in *T. ussuriense* and *T. platycarpum*; however, it could not be annotated because its possible molecular formula did not match any compounds previously reported in *Taraxacum* species.

The metabolomic diversity between *T. campylodes* and *T. coreanum*, and between *T. ussuriense* and *T. platycarpum* was analyzed further using a stepwise orthogonal partial least squares discriminant analysis (OPLS-DA) (Fig. 4C–F). The R2X values of the OPLS-DA models for

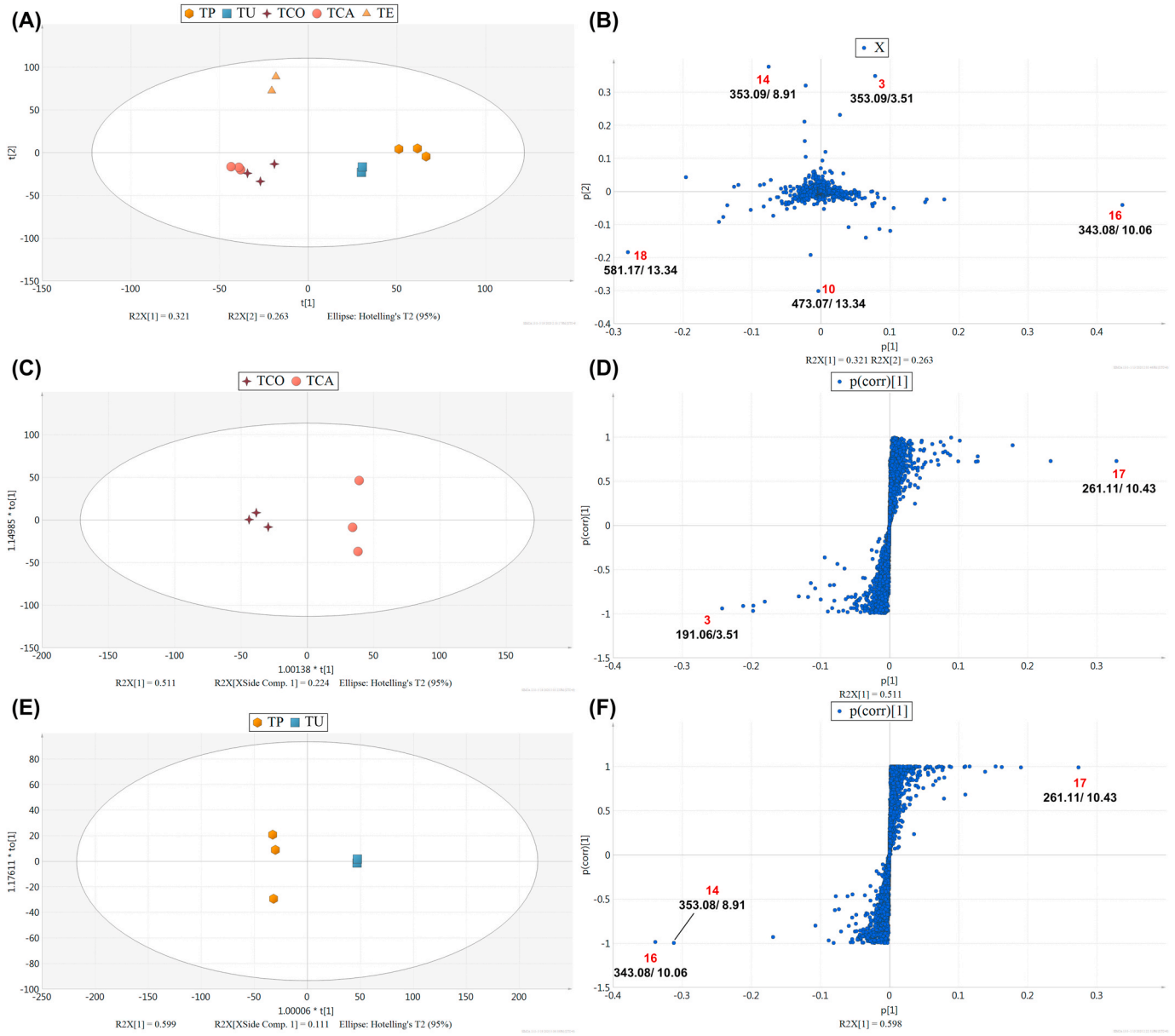


Fig. 4. The multivariate analysis of the metabolite data derived from 13 *Taraxacum* extracts. (A) The score plot and (B) loading plot of the PCA model showing PC1-PC2. (C) The score plot and (D) *S*-plot of the OPLS-DA model for *T. campylodes* (TCA) and *T. coreanum* (TCO), showing PC1-PC2. (E) The score plot and (F) *S*-plot of the OPLS-DA model for *T. platycarpum* (TP) and *T. ussuriense* (TU), showing PC1-PC2. In the score plots, the orange hexagons, blue squares, brown plus signs, pink circles, and yellow triangles correspond to *T. platycarpum*, *T. ussuriense*, *T. coreanum*, *T. campylodes*, and *T. erythrospermum* (TE), respectively. In the loading plots, important markers are labelled with xxx/yy, where xxx means *m/z* values and yy means retention time. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

T. campylodes/T. coreanum and *T. ussuriense/T. platycarpum* were 0.511 and 0.598, respectively, which indicated that 51.1% and 59.8% of the variation between analyzed species were characterized by the OPLS-DA models. The OPLS-DA *S*-plot for *T. campylodes* and *T. coreanum* showed relatively high concentrations of chlorogenic acid (**3**) in *T. coreanum*, and taraxinic acid (**17**) in *T. campylodes* (Fig. 4D). For *T. ussuriense* and *T. platycarpum*, the OPLS-DA *S*-plot revealed high concentrations of taraxinic acid (**17**) in *T. ussuriense*, and 3,5-di-*O*-caffeoylquinic acid (**14**) and the unknown metabolite (**16**) in *T. platycarpum* (Fig. 4F).

To investigate the relative quantitative differences in the specialized metabolites of the tested species, we compared the ion abundance of the MS features, which were selected based on the multivariate analysis. As shown in Fig. 5, the relative abundance of selected ions agreed with the pattern observed in the PCA and OPLS-DA. Concentrations of chlorogenic acid (**3**) and 3,5-di-*O*-caffeoylquinic acid (**14**) showed were significantly higher in *T. erythrospermum*, while phenolic inositols (**18** and **19**) were abundant in *T. campylodes* and *T. coreanum*. Taraxinic acid (**17**) concentrations were relatively higher in *T. ussuriense* and *T. campylodes* than in other species. Austricin 8-*O*-*β*-D-glucopyranoside (**7**) was not found as a marker compound in the multivariate analyses, but it also showed a species-specific distribution. In contrast, caftaric acid (**1**) and chicoric acid (**10**) concentrations were similar in every analyzed extract (Supplementary Figure 6).

We performed a hierarchical cluster analysis (HCA) to find out if the

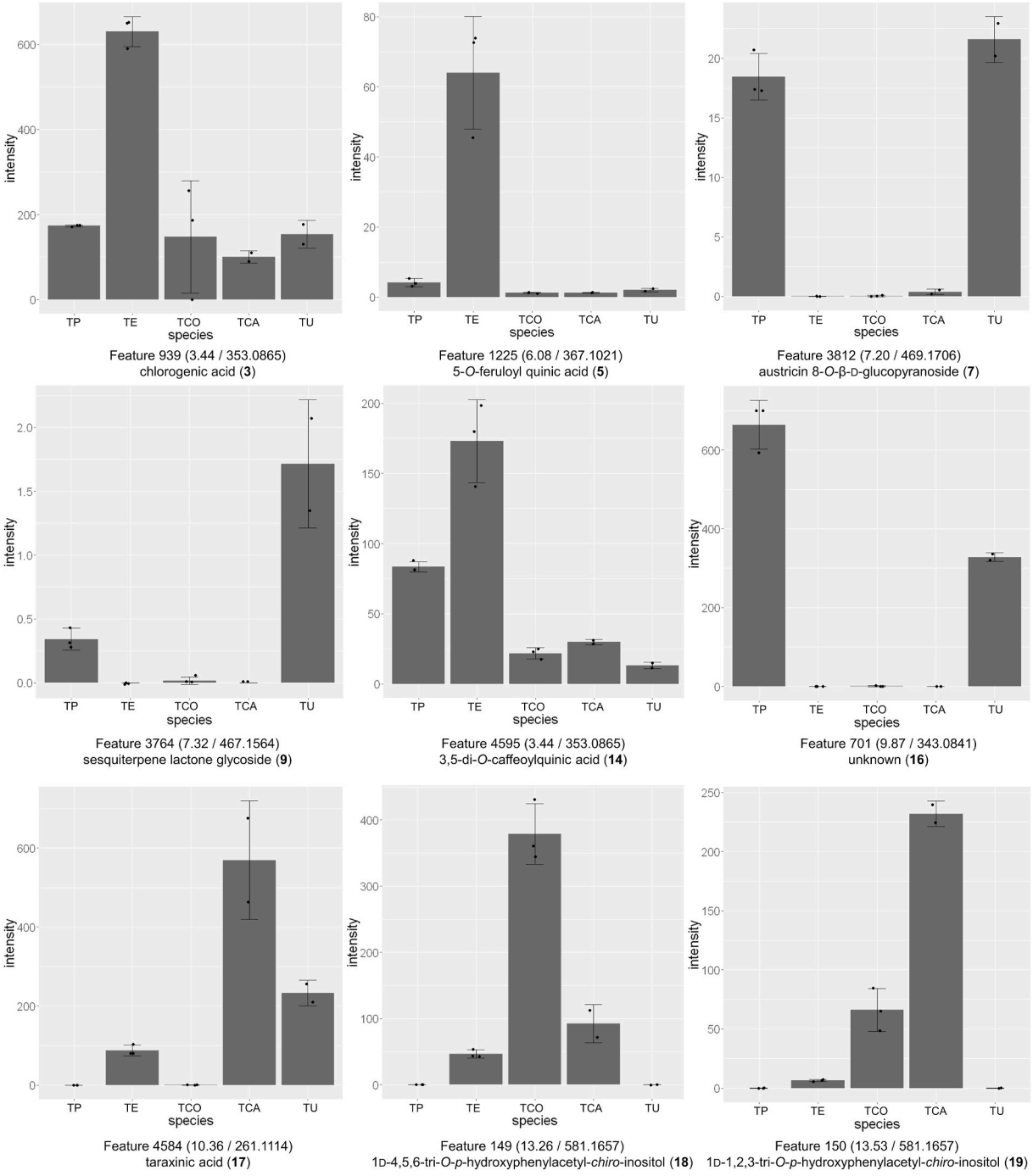


Fig. 5. Bar plots showing the ion intensities of selected defensive metabolites (**3**, **5**, **7**, **9**, **14**, **16**, **17**, **18**, and **19**) in analyzed *Taraxacum* samples. TCA: *T. campylodes*; TCO: *T. coreanum*; TE: *T. erythrospermum*; TU: *T. ussuriense*; TP: *T. platycarpum*.

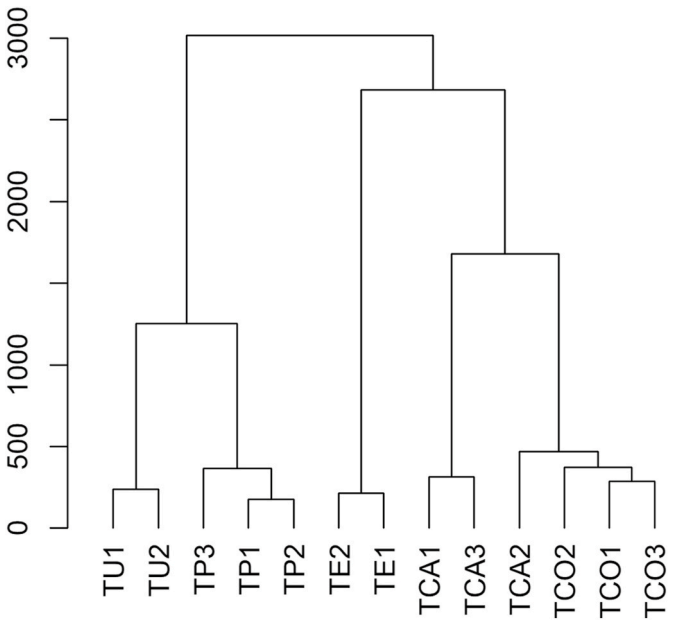


Fig. 6. The chemical dendrogram obtained by HCA (Euclidean distance and Ward's linkage method) on the LC–MS feature table. TCA: *T. campylodes*; TCO: *T. coreanum*; TE: *T. erythrospermum*; TU: *T. ussuriense*; TP: *T. platycarpum*.

tested samples could be grouped into different chemotypes (Fig. 6). The chemical dendrogram acquired from the HCA resembled the phylogram drawn by the nrDNA (Fig. 2); *T. ussuriense* and *T. platycarpum* were clustered into one subgroup, while *T. campylodes*, *T. coreanum*, and *T. erythrospermum* formed another subgroup. When this relationship was compared to the relative ion abundances (Fig. 5), it was found that species in the same subgroup shared defensive metabolites in similar scaffold. *T. ussuriense* and *T. platycarpum* showed high concentrations of austricin 8-*O*-*β*-D-glucopyranoside (**7**) and an unidentified sesquiterpene lactone glycoside (**9**), which made us assume that the unknown metabolite **16** might be a sesquiterpene derivative; however, the presence of another sesquiterpene, taraxinic acid (**17**), did not follow the subgrouping pattern. Phenolic inositols (**18** and **19**) were abundant in *T. campylodes* and *T. coreanum*, while phenyl propanoidal quinic acids (**3**, **5**, and **14**) were dominant in *T. erythrospermum* (Fig. 5).

3. Discussion

In this study, we assembled completed chloroplast genomes from three *Taraxacum* species, *T. coreanum, T. erythrospermum, T. ussuriense* and conducted comparative genome analysis with previously reported three other *Taraxacum* (Kim et al., 2016a, 2016b). The six *Taraxacum* CP genomes have a highly similarity in genome size, structure, and gene numbers (Zhang et al., 2017), indicating that genome feature is conserved between the species. However, three intergenic regions such as *trnM(CAU)-aptE*, *rbcL-psaI*, and *trnL(UAG)-ccsA* show sequence variation between the six *Taraxacum* species. Especially high frequency sequence variations were found in two intergenic regions *rbcL-psaI* and *trnL(UAG)-ccsA* in *Stipa* species (Krawczyk et al., 2018) and *Bupleurum* species (Li et al., 2020), suggesting that those two regions could be further consider barcoding markers to discriminate *Taraxacum* species. Additionally, we also assembled 45S rDNA simultaneously from them and found that heterogeneous form of 45S rDNA occurred in *T. coreanum* and *T. ussuriense*. Comparison of whole 45S rDNA showed 26S rDNA had enough amount of interspecific variations comparable with both ITS1 and ITS2. Together with ITS polymorphisms, the variations in 26S rDNA can help to increase molecular phylogenetic resolution.

In genome-based comparative analyses, the CP genome- and 45S rDNA-based trees showed an incongruence, which is common because of the maternal inheritance of CP genomes (Liu et al., 2017; Pelser et al., 2010). In the chloroplast-based phylogenetic tree, the Korean native species (*T. coreanum* and *T. platycarpum*) were clustered with *T. campylodes* and *T. erythrospermum*, which mainly grow in Europe or certain parts of North America (Chon and Park, 2012; Kim et al., 2009). This indicated that these four species possibly have close maternal ancestors, despite their different geographical origins. The two heterotypes of *T. coreanum*-NR-1 and *T. coreanum*-NR-2 were clustered in the same clade with *T. campylodes* and *T. erythrospermum* in the 45S rDNA-based phylogenetic tree. One of the possible explanations for this genetic similarity between the Korean native species and the introduced species was natural hybridization between *T. campylodes*, *T. coreanum*, and *T. erythrospermum*. The incongruence between the CP DNA and the nuclear genome analyses have been reported in many other *Taraxacum* studies, and were regarded as evidence of an ancient persistent hybridity in the genus (King, 1993; Kirschner et al., 2015; Van Der Hulst et al., 2003). Previous studies on wild *Taraxacum* in Japan reported that wild *Taraxacum* species were under ongoing hybridization at a fast rate (Matsuyama et al., 2018; Shibaike et al., 2002), which supports the possibility of natural hybridization between *T. campylodes*, *T. coreanum*, and *T. erythrospermum*. *T. platycarpum* was different to *T. coreanum* and clustered with *T. ussuriense* in the 45S rDNA-based tree, while *T. ussuriense* formed a clustered clade with *T. mongolicum* in the CP genome-based tree. *T. mongolicum* showed close phylogenetic distances to the other species in the CP tree, while it did not cluster with any other species in the 45S rDNA-based tree. This suggests that the maternal ancestor of *T. mongolicum* was close to other species, especially *T. ussuriense*, while the paternal ancestor may be genetically distant from the other species analyzed in our study.

Most of the specialized metabolites analyzed in our study are known as defensive metabolites against root herbivores. Caffeoylquinic acids (**3** and **14**) were implicated in plant resistance to thrips, a pest that *Taraxacum* species can be hosts of (Leiss et al., 2009; Smith et al., 2011). A glucoside of taraxinic acid (**17**) and phenolic inositol derivatives (**18** and **19**) were also reported as defensive metabolites of *T. campylodes* against insect herbivores (Agrawal et al., 2018; Huber et al., 2016). Chicoric acid (**10**) was discovered from multiple taxa of the plant kingdom, and was reported to help protect against insects, viruses, bacteria, fungi, and nematodes (Lee and Scagel, 2013). Specialized metabolism is usually regulated by two critical factors: genotype and environmental stimuli. As all of the individual plants we tested were maintained in a same botanical garden, we can assume that any observed differences in defensive chemicals we measured were driven by genetic differences between species, not by environmental stimuli. This assumption is also supported by the congruence between the 45S rDNA-based phylogenetic tree and the chemical dendrogram. The consistency between the two was interesting, because in most other studies, chemical dendrograms generally do not correlate with phylogenetic trees (Ernst et al., 2019; Kang et al., 2018). From the matched pattern of the two trees, we assumed that nucleus was involved in species-specific specialized metabolism in *Taraxacum* rather than the chloroplast. However, this hypothesis is only tentative owing to the low number of samples tested and a lack of whole genome data. Considering the high possibility of ongoing hybridization between *Taraxacum* species in wild, we can suppose that the chemotypes of wild dandelions will be further diversified in the future. Interspecific hybridization is considered as one of the most important events driving diversification of plant specialized metabolism (Orians, 2000; Whitehead and Bowers, 2013), because hybridization can give a chemical trait introgression. From this point of view, tracking alterations in defensive metabolism in *Taraxacum* plants would provide further insights about the effects of hybridization on chemical diversity.

Despite their universal presence, *Taraxacum* species have rarely been investigated for their genomic and chemical diversity. In the present study, we reported on the complete CP DNA and 45S rDNA sequences of six *Taraxacum* species that inhabit Korea. Through the CP DNA- and 45S

rDNA-based phylogenetic study, we showed that *Taraxacum* species did not cluster by their geographic origin. Instead, there was a tangled relationship between the CP DNA- and 45S rDNA-based phylogenetic trees, which supports the notion of ongoing hybridization of *Taraxacum* species in the wild. Additionally, we investigated the chemical diversity of five *Taraxacum* species using untargeted LC–MS analysis, which revealed that each *Taraxacum* plant exhibits species-specific defensive specialized metabolism. More intensive investigations are needed to discover the associations between the genetic and chemical relationships of *Taraxacum* species; however, we expect that our comparative study will be a small cornerstone for further studies on this cosmopolitan genus.

4. Experimental

4.1. Chemicals and reagents

High performance liquid chromatography (HPLC) grade water, MeOH and acetonitrile were purchased from Avantor Performance Materials, Inc. (Center Valley, PA, USA). Formic acid, leucine-enkephalin and sodium hydroxide were bought from Sigma-Aldrich (St. Louis, MO, USA). Ultrapure water was triple deionized using the Super-Q water purification system (Millipore, Bedford, MA, USA). Reference standards of *cis*-caftaric acid (**1**), chlorogenic acid (**3**), luteolin 7-*O*-rutinoside (**8**), chicoric acid (**10**), luteolin 7-*O*-glucoside (**11**) and 3,5-di-*O*-caffeoylquinic acid (**14**) were purchased from InterPharm (Koyang, Korea). 1D-4,5,6-tri-*O*-*p*-hydroxyphenylacetyl-*chiro*-inositol (**18**) and 1D-1,2,3-tri-*O*-*p*-hydroxyphenylacetyl-*chiro*-inositol (**19**) were isolated from *T. campylodes* (Choi, 2018). Austricin 8-*O*-β-D-glucopyranoside (**7**) was isolated from *T. ussuriense* through successive preparative chromatography and then identified by NMR spectroscopy.

4.2. Plant materials

Taraxacum campylodes G.E.Haglund, *Taraxacum coreanum* Nakai, *Taraxacum erythrospermum* Willd. DC, *Taraxacum mongolicum* Hand.-Mazz, *Taraxacum ussuriense* Kom, and *Taraxacum platycarpum* Dahlst. were collected from the Hantaek Botanical Garden, Yongin, Korea (GPS N37°05′, E127°24′), at the same time (April 2015). These plants had been maintained in the botanical garden for years, which might minimize the effects of the surrounding environment on the specialized metabolite profiles. The original collection sites for the plants are as follows: Yongin (GPS N37°13′, E127°13′; *T. campylodes* and *T. ussuriense*), Seosan (GPS N36°47′, E126°27′; *T. coreanum*), Cheongwon (GPS N36°44′, E127°25′; *T. erythrospermum*), Namwon (GPS N35°24′, E127°23′; *T. mongolicum*), and Chuncheon (GPS N37°53′, E127°44′; *T. platycarpum*). Collection of the wild-grown *Taraxacum* plants were performed under the permission of land owners. The samples were authenticated by Mr. T. J. Lee (Hantaek Botanical Garden), and voucher specimens (SUPH-1504-01, *T. coreanum*; SUPH-1504-02, *T. erythrospermum*; SUPH-1504-03, *T. mongolicum*; SUPH-1504-04, *T. campylodes*; SUPH-1504-05, *T. ussuriense*, SUPH-1504-06, *T. platycarpum*) are deposited in the Herbarium of the Medicinal Plant Garden, Seoul National University (Koyang, Korea).

4.3. Sample preparation

4.3.1. Whole genome sequencing

Total genomic DNAs were extracted from the fresh leaves of each species using the cetyltrimethylammonium bromide method (Allen et al., 2006). After checking the quality and quantity of DNAs on agarose gel, a library was constructed using the DNAs, followed by genomic sequencing with Illumina Miseq and Nextseq according to manufacturer's protocols (Illumina, San Diego, CA, USA).

4.3.2. Assembly of the CP genome and the 45S rDNA sequence, and phylogenetic analysis

The paired-end whole genome sequences from each species were assembled by a CLC *de novo* genome assembler (v. beta 4.6, CLC Inc, Aarhus, Denmark), as described in our previous studies (Kim et al, 2015, 2017). Draft CP genomes and 45S rDNA were edited by manual curation using mapping of the raw reads. After assembly, CP gene annotation was performed using the GeSeq program (Tillich et al., 2017) by comparisons with homologous genes from other chloroplast sequences in Asteraceae. Among the assembled genomes, CP genome sequences of three species, *T. campylodes*, *T. platycarpum*, and *T. mongolicum* were previously reported by our group (Kim et al., 2016a, 2016b). 45S rDNA was annotated based on comparisons with other plant 45S rDNA by BLAST in NCBI (http://www.ncbi.nlm.nih.gov/BLAST/) (Camacho et al., 2009). The CP and 45S rDNA sequences were aligned by the MAFFT program (https://mafft.cbrc.jp/alignment/server/) (Katoh et al., 2002), while the phylogeny was drawn using the neighbor-joining (NJ) method with 1000 bootstrap values in MEGA 6.0 (Tamura et al., 2013). Previously reported CP sequences and 45S rDNA (Genbank accession number: MT003979) of *Artemisia fukudo* (Asteraceae) was applied as an outgroup (Lee et al., 2016). The pi value for nucleotide divergence was calculated by using DnaSP v6 (Rozas et al., 2017) with parameter of 600bp of windows size, 200bp of sliding size and exclusion of gap.

4.4. The LC–MS analysis of specialized metabolites

The powdered whole plant bodies were accurately weighed for 100.0 mg and then were extracted with 1.0 ml of MeOH/H_2O (5:5, v/v), sonicated at room temperature for 15 min, centrifuged at 10,000 *g* for 5 min, and then the 0.5 ml supernatant was collected. The supernatant was filtered with a polyvinylidene difluoride (PVDF) filter for the LC–MS analysis. Samples were biologically multiplicated using different individuals from the same species; *T. campylodes*, *T. coreanum*, and *T. platycarpum* were triplicated, while *T. erythrospermum* and *T. ussuriense* were duplicated due to the lack of maintained individuals. A Branson 8510 ultrasonic bath (Branson Ultrasonics Corporation, Danbury, CT) was used for extractions. Centrifugation was performed using a HANIL micro centrifuge (Micro17TR, Hanil scientific industrial, Seoul, Korea).

The LC–MS analysis was performed on a Waters Acquity UPLC system (Waters Co, Milford, MA, USA) coupled to a Waters Xevo G2 QTOF mass spectrometer (Waters MS Technologies, Manchester, UK), which was equipped with an electrospray ionization interface (ESI). Chromatographic separations were performed on a Waters Acquity HSS T3 (100 × 2.1 mm, 1.8 μm) column. The mobile phase was comprised of H_2O (A) and acetonitrile (B), both of which were acidified with 0.1% formic acid. The column temperature and sample organizer were maintained at 40 °C and 15 °C, respectively. A stepwise gradient method at a constant flow rate of 0.3 ml/min was used to elute the column with the following conditions: 10–11% B (0–3 min); 11–20% B (3–6 min); 20% B (6–11 min); 20–40% B (11–14 min); and 40–100% B (14–17 min), followed by 3 min of washing and 3 min of reconditioning. The photo diode array (PDA) detector scanning range was from 200 to 500 nm with a resolution of 1.2 nm and a sampling rate of 20 times/s. Analyses of the samples (1.0 μL injected into the partial loop in the needle overfill mode) were performed in the negative ion mode in the *m/z* 50–1200 Da range, with acquisition times of 0.2 s in the centroid mode. The ESI conditions were set as follows: the capillary voltage was 3.0 kV, the cone voltage was 25 V, the source temperature was 120 °C, the desolvation temperature was 450 °C, the cone gas flow was 50 L/h, and the desolvation gas flow was 800 L/h. High-purity nitrogen was used as the nebulizer and auxiliary gas, and argon was used as the collision gas. The $[M-H]^-$ ion of leucine enkephalin at *m/z* 554.2615 was used as the lock mass to ensure mass accuracy and reproducibility. The MS^E methodology was applied to acquire high-energy CID MS spectra for tentative

identification of the metabolites (Plumb et al., 2006). The low collision energy for the detection of the precursor ions was set to 3 eV, while the high collision energy for fragmentation was set to 20–40 eV. For data acquisition, pooled samples were used for quality control checks. Sample acquisition was randomized and the QC sample was analyzed every 5 injections to monitor and correct for changes in the instrument response.

4.5. Data preprocessing and multivariate analyses for metabolomic data

LC–MS raw data were preprocessed into an ion marker table using MarkerLynx XS software (version 4.1, Waters Co.). The data matrix was created with a method using the following parameters: retention time (t_R) 2.0–17.0 min, mass range *m/z* 100–1200 Da, mass tolerance of 0.02 Da, and an intensity threshold of 100 counts. The alignment of peaks across the samples was performed within the range of ±0.02 Da mass and ±0.20 min t_R windows. For the parameters in the ApexTrack algorithm, the control peak detection by peak width (peak width at 5% height) and baseline threshold (peak-to-peak baseline ratio) were automatically calculated by MarkerLynx. The noise elimination level was set to 100. As a result, 5283 ion features were extracted from the LC–MS profiles and arranged into a peak table. Peak areas were normalized to the total peak area of each sample, then PCA and OPLS-DA were performed using SIMCA 13.0 (Umetrics, Umeå, Sweden). Hierarchical clustering analysis (HCA) was performed using the R functions *dist* and *hclust*, with Euclidean distance and Ward's linkage method.

Data availability

Assembled genomes were deposited in NCBI and Genbank with the following accession. CP data: *T. coreanum*, MN689809; *T. erythrospermum*, MN689810; *T. ussuriense*, MN689808.45S rDNA data: *T. campylodes*, MT077849; *T. coreanum*-NR-1, MT077854; *T. coreanum*-NR-2, MT077855; *T. erythrospermum*, MT077856; *T. mongolicum*, MT077853; *T. ussuriense*-NR-1, MT077851; *T. ussuriense*-NR-1, MT077852; *T. platycarpum*, MT077850. All LC–MS raw data files and the extracted ion table are deposited in MassIVE (https://massive.ucsd.edu) with the accession number MSV000084931.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This research was supported by the Cooperative Research Program for Agriculture Science & Technology Development of the Rural Development Administration, Republic of Korea (PJ013238) and Ministry of Food and Drug Safety, Republic of Korea (15172MFDS246). KBK was supported by Sookmyung Women's University Research Grants (1-1903-2004). The funders had no roles in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

References

Abdullah, Mehmood, F., Shahzadi, I., Waseem, S., Mirza, B., Ahmed, I., Waheed, M.T., 2020. Chloroplast genome of *Hibiscus rosa-sinensis* (Malvaceae): comparative analyses and identification of mutational hotspots. Genomics 112, 581–591. https://doi.org/10.1016/j.ygeno.2019.04.010.

Agrawal, A.A., Hastings, A.P., Fines, D.M., Bogdanowicz, S., Huber, M., 2018. Insect herbivory and plant adaptation in an early successional community. Evolution 72 (5), 1020–1033. https://doi.org/10.1111/evo.13451.

Allen, G.C., Flores-Vergara, M.A., Krasynanski, S., Kumar, S., Thompson, W.F., 2006. A modified protocol for rapid DNA isolation from plant tissues using cetyltrimethylammonium bromide. Nat. Protoc. 1 (5), 2320–2325. https://doi.org/10.1038/nprot.2006.384.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., Madden, T.L., 2009. BLAST+: architecture and applications. BMC Bioinform. 10, 421. https://doi.org/10.1186/1471-2105-10-421.

Chen, H.J., Inbaraj, B.S., Chen, B.H., 2012. Determination of phenolic acids and flavonoids in *Taraxacum formosanum* Kitam by liquid chromatography-tandem mass spectrometry coupled with a post-column derivatization technique. Int. J. Mol. Sci. 13 (1), 260–285. https://doi.org/10.3390/ijms13010260.

Choi, J., 2018. Chemical Constituents from Taraxacum officinale and Their α-glucosidase Inhibitory Activities. Seoul National University. Doctoral dissertation.

Chon, S.-U., Park, J.-S., 2012. Change in plant growth and physiologically-active compounds content of *Taraxacum officinale* under plastic house condition. Korean J. Crop Sci. 57 (4), 449–455. https://doi.org/10.7740/kjcs.2012.57.4.449.

Daniell, H., Lin, C.S., Yu, M., Chang, W.J., 2016. Chloroplast genomes: diversity, evolution, and applications in genetic engineering. Genome Biol. 17, 134. https://doi.org/10.1186/s13059-016-1004-2.

Ernst, M., Nothias, L.F., van der Hooft, J.J.J., Silva, R.R., Saslis-Lagoudakis, C.H., Grace, O.M., Martinez-Swatson, K., Hassemer, G., Funez, L.A., Simonsen, H.T., Medema, M.H., Staerk, D., Nilsson, N., Lovato, P., Dorrestein, P.C., Rønsted, N., 2019. Assessing specialized metabolite diversity in the cosmopolitan plant genus *Euphorbia*. Front. Plant Sci. 10, 846. https://doi.org/10.3389/fpls.2019.00846.

Ernst, M., Silva, D.B., Silva, R.R., Vêncio, R.Z.N., Lopes, N.P., 2014. Mass spectrometry in plant metabolomics strategies: from analytical platforms to data acquisition and processing. Nat. Prod. Rep. 31 (6), 784–806. https://doi.org/10.1039/c3np70086k.

Guarrera, P.M., Savo, V., 2016. Wild food plants used in traditional vegetable mixtures in Italy. J. Ethnopharmacol. 185, 202–234. https://doi.org/10.1016/j.jep.2016.02.050.

Hartmann, T., 2007. From waste products to ecochemicals: fifty years research of plant secondary metabolism. Phytochemistry 68 (22–24), 2831–2846. https://doi.org/10.1016/j.phytochem.2007.09.017.

Henriquez, C.L., Abdullah Ahmed, I., Carlsen, M.M., Zuluaga, A., Croat, T.B., McKain, M. R., 2020a. Molecular evolution of chloroplast genomes in Monsteroideae (Araceae). Planta 251, 72. https://doi.org/10.1007/s00425-020-03365-7.

Henriquez, C.L., Abdullah Ahmed, I., Carlsen, M.M., Zuluaga, A., Croat, T.B., McKain, M. R., 2020b. Evolutionary dynamics of chloroplast genomes in subfamily Aroideae (Araceae). Genomics 112 (3), 2349–2360. https://doi.org/10.1016/j.ygeno.2020.01.006.

Huber, M., Epping, J., Schulze Gronover, C., Fricke, J., Aziz, Z., Brillatz, T., Swyers, M., Köllner, T.G., Vogel, H., Hammerbacher, A., Triebwasser-Freese, D., Robert, C.A.M., Verhoeven, K., Preite, V., Gershenzon, J., Erb, M., 2016. A latex metabolite benefits plant fitness under root herbivore attack. PLoS Biol. 14 (1), e1002332 https://doi.org/10.1371/journal.pbio.1002332.

Jheng, C.F., Chen, Tien Chih, Lin, J.Y., Chen, Ting Chieh, Wu, W.L., Chang, C.C., 2012. The comparative chloroplast genomic analysis of photosynthetic orchids and developing DNA markers to distinguish *Phalaenopsis* orchids. Plant Sci. 190, 62–73. https://doi.org/10.1016/j.plantsci.2012.04.001.

Kang, K. Bin, Kang, S.J., Kim, M.S., Lee, D.Y., Han, S. Il, Kim, T.B., Park, J.Y., Kim, J., Yang, T.J., Sung, S.H., 2018. Chemical and genomic diversity of six *Lonicera* species occurring in Korea. Phytochemistry 115, 126–135. https://doi.org/10.1016/j.phytochem.2018.07.012.

Katoh, K., Misawa, K., Kuma, K.I., Miyata, T., 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 30 (14), 3059–3066. https://doi.org/10.1093/nar/gkf436.

Kenny, O., Smyth, T.J., Hewage, C.M., Brunton, N.P., McLoughlin, P., 2014. 4-Hydroxyphenylacetic acid derivatives of inositol from dandelion (*Taraxacum officinale*) root characterised using LC-SPE-NMR and LC-MS techniques. Phytochemistry 98, 197–203. https://doi.org/10.1016/j.phytochem.2013.11.022.

Kim, J.K., Park, J.Y., Lee, Y.S., Lee, H.O., Park, H.S., Lee, S.C., Kang, J.H., Lee, T.J., Sung, S.H., Yang, T.J., 2016a. The complete chloroplast genome sequence of the *Taraxacum officinale* F.H.Wigg (Asteraceae). Mitochondrial DNA Part B Resour 1 (1), 228–229. https://doi.org/10.1080/23802359.2016.1155425.

Kim, J.K., Park, J.Y., Lee, Y.S., Woo, S.M., Park, H.S., Lee, S.C., Kang, J.H., Lee, T.J., Sung, S.H., Yang, T.J., 2016b. The complete chloroplast genomes of two *Taraxacum* species, *T. platycarpum* Dahlst. and *T. mongolicum* Hand.-Mazz. (Asteraceae). Mitochondrial DNA Part B Resour 1 (1), 412–413. https://doi.org/10.1080/23802359.2016.1176881.

Kim, K., Lee, S.C., Lee, Junki, Yu, Y., Yang, K., Choi, B.S., Koh, H.J., Waminal, N.E., Choi, H. Il, Kim, N.H., Jang, W., Park, H.S., Lee, Jonghoon, Lee, H.O., Joh, H.J., Lee, H.J., Park, J.Y., Perumal, S., Jayakodi, M., Lee, Y.S., Kim, B., Copetti, D., Kim, Soonok, Kim, Sunggil, Lim, K.B., Kim, Y.D., Lee, Jungho, Cho, K.S., Park, B.S., Wing, R.A., Yang, T.J., 2015. Complete chloroplast and ribosomal sequences for 30 accessions elucidate evolution of *Oryza* AA genome species. Sci. Rep. 5, 15655. https://doi.org/10.1038/srep15655.

Kim, K., Nguyen, V.B., Dong, J., Wang, Y., Park, J.Y., Lee, S.C., Yang, T.J., 2017. Evolution of the Araliaceae family inferred from complete chloroplast genomes and 45S nrDNAs of 10 *Panax*-related species. Sci. Rep. 7, 4917. https://doi.org/10.1038/s41598-017-05218-y.

Kim, Y.H., Hamayun, M., Khan, A.L., Na, C.I., Kang, S.M., Han, H.H., Lee, I.J., 2009. Exogenous application of plant growth regulators increased the total flavonoid content in *Taraxacum officinale* Wigg. Afr. J. Biotechnol. 8 (21), 5727–5732. https://doi.org/10.5897/ajb09.927.

King, L.M., 1993. Origins of genotypic variation in North American dandelions inferred from ribosomal DNA and chloroplast DNA restriction enzyme analysis. Evolution 47 (1), 136–151. https://doi.org/10.2307/2410124.

Kirschner, J., Záveská Drábková, L., Štěpánek, J., Uhlemann, I., 2015. Towards a better understanding of the *Taraxacum* evolution (Compositae–Cichorieae) on the basis of nrDNA of sexually reproducing species. Plant Systemat. Evol. 301, 1135–1156. https://doi.org/10.1007/s00606-014-1139-0.

Krawczyk, K., Nobis, M., Myszczyński, K., Klichowska, E., Sawicki, J., 2018. Plastid super-barcodes as a tool for species discrimination in feather grasses (Poaceae: *Stipa*). Sci. Rep. 8, 1924. https://doi.org/10.1038/s41598-018-20399-w.

Lee, J., Scagel, C.F., 2013. Chicoric acid: chemistry, distribution, and production. Front. Chem. 1, 40. https://doi.org/10.3389/fchem.2013.00040.

Lee, Y.S., Park, J.Y., Kim, J.K., Lee, H.O., Park, H.S., Lee, S.C., Kang, J.H., Lee, T.J., Sung, S.H., Yang, T.J., 2016. Complete chloroplast genome sequence of *Artemisia fukudo* Makino (Asteraceae). Mitochondrial DNA Part B Resour 1 (1), 376–377. https://doi.org/10.1080/23802359.2016.1155426.

Leiss, K.A., Maltese, F., Choi, Y.H., Verpoorte, R., Klinkhamer, P.G.L., 2009. Identification of chlorogenic acid as a resistance factor for thrips in chrysanthemum. Plant Physiol. 150 (3), 1567–1575. https://doi.org/10.1104/pp.109.138131.

Li, J., Xie, D.F., Guo, X.L., Zheng, Z.Y., He, X.J., Zhou, S.D., 2020. Comparative analysis of the complete plastid genome of five *Bupleurum* species and new insights into DNA barcoding and phylogenetic relationship. Plants 9, 543. https://doi.org/10.3390/plants9040543.

Liigand, P., Kaupmees, K., Haav, K., Liigand, J., Leito, I., Girod, M., Antoine, R., Kruve, A., 2017. Think negative: finding the best electrospray ionization/MS mode for your analyte. Anal. Chem. 89 (11), 5665–5668. https://doi.org/10.1021/acs.analchem.7b00096.

Liu, X., Wang, Z., Shao, W., Ye, Z., Zhang, J., 2017. Phylogenetic and taxonomic status analyses of the abaso section from multiple nuclear genes and plastid fragments reveal new insights int the North America origin of *Populu* (Salicaceae). Front. Plant Sci. 7, 2022. https://doi.org/10.3389/fpls.2016.02022.

Martinez, M., Poirrier, P., Chamy, R., Prüfer, D., Schulze-Gronover, C., Jorquera, L., Ruiz, G., 2015. *Taraxacum officinale* and related species - an ethnopharmacological review and its potential as a commercial medicinal plant. J. Ethnopharmacol. 169, 244–262. https://doi.org/10.1016/j.jep.2015.03.067.

Martucci, M.E.P., De Vos, R.C.H., Carollo, C.A., Gobbo-Neto, L., 2014. Metabolomics as a potential chemotaxonomical tool: application in the genus *Vernonia* Schreb. PloS One 9 (4), e93149. https://doi.org/10.1371/journal.pone.0093149.

Matsuyama, S., Morimoto, M., Harata, T., Nanami, S., Itoh, A., 2018. Hybridization rate and genotypic diversity of apomictic hybrids between native (*Taraxacum japonicum*) and introduced (*T. officinale*) dandelions in western Japan. Conserv. Genet. 19, 181–191. https://doi.org/10.1007/s10592-017-1014-y.

Mingarro, D.M., Plaza, A., Galán, A., Vicente, J.A., Martínez, M.P., Acero, N., 2015. The effect of five *Taraxacum* species on in vitro and in vivo antioxidant and antiproliferative activity. Food Funct 6 (8), 2787–2793. https://doi.org/10.1039/c5fo00645g.

Nakabayashi, R., Saito, K., 2013. Metabolomics for unknown plant metabolites. Anal. Bioanal. Chem. 405 (15), 5005–5011. https://doi.org/10.1007/s00216-013-6869-2.

Orians, C.M., 2000. The effects of hybridization in plants on secondary chemistry: implications for the ecology and evolution of plant-herbivore interactions. Am. J. Bot. 87 (12), 1749–1756. https://doi.org/10.2307/2656824.

Palmer, J.D., 1985. Comparative organization of chloroplast genomes. Annu. Rev. Genet. 19, 325–354. https://doi.org/10.1146/annurev.ge.19.120185.001545.

Park, H.S., Jayakodi, M., Lee, S.H., Jeon, J.H., Lee, H.O., Park, J.Y., Moon, B.C., Kim, C. K., Wing, R.A., Newmaster, S.G., Kim, J.Y., Yang, T.J., 2020. Mitochondrial plastid DNA can cause DNA barcoding paradox in plants. Sci. Rep. 10, 6112. https://doi.org/10.1038/s41598-020-63233-y.

Pelser, P.B., Kennedy, A.H., Tepe, E.J., Shidler, J.B., Nordenstam, B., Kadereit, J.W., Watson, L.E., 2010. Patterns and causes of incongruence between plastid and nuclear Senecioneae (Asteraceae) phylogenies. Am. J. Bot. 97 (5), 856–873. https://doi.org/10.3732/ajb.0900287.

Plumb, R.S., Johnson, K.A., Rainville, P., Smith, B.W., Wilson, I.D., Castro-Perez, J.M., Nicholson, J.K., 2006. UPLC/MS^{E}; a new approach for generating molecular fragment information for biomarker structure elucidation. Rapid Commun. Mass Spectrom. 20 (13), 1989–1994. https://doi.org/10.1002/rcm.2550.

Richards, A.J., 1970. Eutriploid facultative agamospermy in *Taraxacum*. New Phytol. 69 (3), 761–774. https://doi.org/10.1111/j.1469-8137.1970.tb02461.x.

Rozas, J., Ferrer-Mata, A., Sanchez-DelBarrio, J.C., Guirao-Rico, S., Librado, P., Ramos-Onsins, S.E., Sanchez-Gracia, A., 2017. DnaSP 6: DNA sequence polymorphism analysis of large data sets. Mol. Biol. Evol. 34 (12), 3299–3302. https://doi.org/10.1093/molbev/msx248.

Ryu, J., Jae-il, L., Chang-Hyu, B., 2017. Genetic variation and phylogenetic relationship of *Taraxacum* based on chloroplast DNA (trnL-trnF and rps16-trnK) sequences. Korean J. Polar Res. 30 (5), 522–534. https://doi.org/10.7732/kjpr.2017.30.5.522.

Salih, R.H.M., Majeský, L., Schwarzacher, T., Gornall, R., Heslop-Harrison, P., 2017. Complete chloroplast genomes from apomictic *Taraxacum* (Asteraceae): identity and variation between three microspecies. PloS One 12 (2), e0168008. https://doi.org/10.1371/journal.pone.0168008.

Schütz, K., Carle, R., Schieber, A., 2006. *Taraxacum*-A review on its phytochemical and pharmacological profile. J. Ethnopharmacol. 107 (3), 313–323. https://doi.org/10.1016/j.jep.2006.07.021.

Schütz, K., Kammerer, D.R., Carle, R., Schieber, A., 2005. Characterization of phenolic acids and flavonoids in dandelion (*Taraxacum officinale* WEB. ex WIGG.) root and herb by high-performance liquid chromatography/electrospray ionization mass spectrometry. Rapid Commun. Mass Spectrom. 19 (2), 179–186. https://doi.org/10.1002/rcm.1767.

Shibaike, H., Akiyama, H., Uchiyama, S., Kasai, K., Morita, T., 2002. Hybridization between European and asian dandelions (*Taraxacum* section *ruderalia* and section *mongolica*). J. Plant Res. 115 (5), 321–328. https://doi.org/10.1007/s10265-002-0045-7.

Smith, E.A., Ditommaso, A., Fuchs, M., Shelton, A.M., Nault, B.A., 2011. Weed hosts for onion thrips (Thysanoptera: thripidae) and their potential role in the epidemiology of *Iris yellow spot virus* in an onion ecosystem. Environ. Entomol. 40 (2), 194–203. https://doi.org/10.1603/EN10246.

Stolze, A., Wanke, A., van Deenen, N., Geyer, R., Prüfer, D., Schulze Gronover, C., 2017. Development of rubber-enriched dandelion varieties by metabolic engineering of the inulin pathway. Plant Biotechnol. 15 (6), 740–753. https://doi.org/10.1111/pbi.12672. J.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., Kumar, S., 2013. MEGA6: molecular evolutionary genetics analysis version 6.0. Mol. Biol. Evol. 30 (12), 2725–2729. https://doi.org/10.1093/molbev/mst197.

Tillich, M., Lehwark, P., Pellizzer, T., Ulbricht-Jones, E.S., Fischer, A., Bock, R., Greiner, S., 2017. GeSeq - versatile and accurate annotation of organelle genomes. Nucleic Acids Res. 45 (W1) https://doi.org/10.1093/nar/gkx391. W6–11.

Van Der Hulst, R.G.M., Mes, T.H.M., Falque, M., Stam, P., Den Nijs, J.C.M., Bachmann, K., 2003. Genetic structure of a population sample of apomictic dandelions. Heredity 90 (4), 326–335. https://doi.org/10.1038/sj.hdy.6800248.

Whitehead, S.R., Bowers, M.D., 2013. Iridoid and secoiridoid glycosides in a hybrid complex of bush honeysuckles (*Lonicera* spp., Caprifolicaceae): implications for evolutionary ecology and invasion biology. Phytochemistry 86, 57–63. https://doi.org/10.1016/j.phytochem.2012.10.012.

Wink, M., 2010. Introduction: biochemistry, physiology and ecological functions of secondary metabolites. In: Wink, M. (Ed.), Annual Plant Reviews: Biochemistry of Plant Secondary Metabolism, vol. 40. Wiley-Blackwell, Hoboken, NJ, pp. 1–19. https://doi.org/10.1002/9781444320503.ch1.

Wink, M., 2003. Evolution of secondary metabolites from an ecological and molecular phylogenetic perspective. Phytochemistry 64 (1), 3–19. https://doi.org/10.1016/S0031-9422(03)00300-5.

Zhang, Y., Iaffaldano, B.J., Zhuang, X., Cardina, J., Cornish, K., 2017. Chloroplast genome resources and molecular markers differentiate rubber dandelion species from weedy relatives. BMC Plant Biol. 17 (1), 34. https://doi.org/10.1186/s12870-016-0967-1.