

RESEARCH ARTICLE

Genetic and chemical markers for authentication of three *Artemisia* species: *A. capillaris*, *A. gmelinii*, and *A. fukudo*

Yun Sun Lee^{1☯^{aa}}, Sunmin Woo^{2☯^{ab}}, Jin-Kyung Kim^{1^{ac}}, Jee Young Park¹, Nur Kholilatul Izzah^{1^{ad}}, Hyun-Seung Park¹, Jung Hwa Kang³, Taek Joo Lee³, Sang Hyun Sung^{2†}, Kyo Bin Kang^{2,4*}, Tae-Jin Yang^{1*}

1 Department of Plant Science, Plant Genomics and Breeding Institute, Research Institute of Agriculture and Life Sciences, College of Agriculture and Life Sciences, Seoul National University, Seoul, Republic of Korea, **2** College of Pharmacy and Research Institute of Pharmaceutical Sciences, Seoul National University, Seoul, Republic of Korea, **3** Hantaek Botanical Garden, Yongin, Republic of Korea, **4** Research Institute of Pharmaceutical Sciences, College of Pharmacy, Sookmyung Women's University, Seoul, Republic of Korea

☯ These authors contributed equally to this work.

† Deceased.

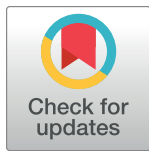
^{aa} Current address: Department of Biochemistry and Molecular Biology, Michigan State University, East Lansing, MI, United States of America

^{ab} Current address: Center for the Study of Human Health, Emory University, Atlanta, GA, United States of America

^{ac} Current address: Research and Development Center for Life Science, Seegene Inc., Seoul, Republic of Korea

^{ad} Current address: Indonesian Industrial and Beverage Crops Research Institute, Sukabumi, West Java, Indonesia

* kbkang@sookmyung.ac.kr (KBK); tjyang@snu.ac.kr (T-JY)



OPEN ACCESS

Citation: Lee YS, Woo S, Kim J-K, Park JY, Izzah NK, Park H-S, et al. (2022) Genetic and chemical markers for authentication of three *Artemisia* species: *A. capillaris*, *A. gmelinii*, and *A. fukudo*. PLoS ONE 17(3): e0264576. <https://doi.org/10.1371/journal.pone.0264576>

Editor: Branislav T. Šiler, Institute for Biological Research, University of Belgrade, SERBIA

Received: November 15, 2021

Accepted: February 13, 2022

Published: March 10, 2022

Copyright: © 2022 Lee et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All DNA sequences can be found in GenBank under the following accession numbers: KU736962 for *A. gmelinii*-A, KY073390 for *A. gmelinii*-B, KU736963 for *A. capillaris*-A, KY073391 for *A. capillaris*-B, and KU360270 for *A. fukudo*. Raw LC-MS/MS data were deposited in MassIVE (<https://massive.ucsd.edu>) under the accession number MSV000087782.

Funding: This research was supported by the Cooperative Research Program for Agriculture Science & Technology Development of the Rural

Abstract

The genus *Artemisia* is an important source of medicines in both traditional and modern pharmaceuticals, particularly in East Asia. Despite the great benefits of herbal medicine, quality assessment methods for these medicinal herbs are lacking. The young leaves from *Artemisia* species are generally used, and most of the species have similar morphology, which often leads to adulteration and misuse. This study assembled five complete chloroplast genomes of three *Artemisia* species, two accessions of *A. gmelinii* and *A. capillaris*, and one *A. fukudo*. Through comparative analysis, we revealed genomic variations and phylogenetic relationships between these species and developed seven InDel-based barcode markers which discriminated the tested species from each other. Additionally, we analyzed specialized metabolites from the species using LC-MS and suggested chemical markers for the identification and authentication of these herbs. We expect that this integrated and complementary authentication method would aid in reducing the misuse of *Artemisia* species.

Introduction

The genus *Artemisia* (Asteraceae), mainly native to Central Asia, consists of more than 500 taxa at a specific or subspecific level [1,2]. In East Asian traditional medicine, *Artemisia* species

Development Administration, Republic of Korea (PJ013238) and Ministry of Food and Drug Safety, Republic of Korea (15172MFDS246). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

have been widely used to treat jaundice, liver diseases, and renal failure [3]. Specialized metabolites biosynthesized by *Artemisia* species, especially phenolic compounds, exhibit potent bioactivities, including antibiotic, antioxidant, hepatoprotective, and antifibrotic effects [4]. The Nobel Prize awardee Youyou Tu's discovery of artemisinin from *A. annua* highlights the potential of specialized metabolites as a drug discovery source and the ethnopharmacological value of these medicinal plants [5].

The large genetic diversity of *Artemisia* results in a high variety of morphologies and phytochemical contents. Based on the infrageneric classification inferred from the nuclear ribosomal internal transcribed spacer (ITS), *Artemisia* species are categorized into five subgenera: *Artemisia*, *Absinthium*, *Dracunculus*, *Seriphidium*, and *Tridentata* [6]. The subgenus *Artemisia* is the most heterogeneous in morphological, chemical, ecological, and karyological characteristics [1]. *A. gmelinii* Webb. ex Stechm., belonging to the subgenera *Artemisia*, and *A. capillaris* Thunb., belonging to the subgenus *Dracunculus*, are consumed as medicinal herbs, especially the young leaves. However, the morphology of the young leaves is similar. This complicates species identification and leads to adulteration or misuse in herbal markets. Thus, a validated method to authenticate *Artemisia* species needs to be developed [3].

DNA barcoding has been suggested as a reliable tool for identifying plant species used as food and herbal medicines. The ITS and chloroplast regions, such as intergenic regions between *trnH*—*psbA*, *psbK*—*psbI*, and *atpF*—*atpH*, and genic regions including *rpoB*, *rpoC1*, *rbcL*, and *matK*, are used to classify intergenic or intragenic species and to understand the taxonomic relationship between species [7]. DNA barcoding markers for the *Artemisia* species were developed using the ITS and chloroplast intergenic regions between *trnH*—*psbA* and *trnL*—*trnF*. A sequence characterized amplified regions (SCAR) marker was developed to discriminate *A. princeps* and *A. argyi* from *A. capillaris* and *A. iwayomogi* [8], and the *trnL-F* region of the chloroplast genome was used to discriminate *A. apiacea*, *A. keiskeana*, and *A. sieversiana* from 18 other *Artemisia* species [3]. High-resolution melting analysis based on ITS2 sequences successfully discriminated *A. argyi*, *A. annua*, *A. lavandulaefolia*, *A. indica*, and *A. atrovirens* [9]. ITS2 and *psbA*—*trnH* were applied to identify *A. argyi* from 15 closely related species and counterfeits [10], and ITS, ITS2, and *psbA*—*trnH* were able to identify *A. annua*, *A. absinthium*, *A. rupestris*, *A. tonurnefortiana*, *A. austriaca*, *A. dracunculus*, *A. vulgaris*, and *A. macrocephala* [11]. However, universal DNA barcoding markers developed from plastids, widely used for the discrimination of plant species, may not be applicable for some species because of the transfer of ancestral plastid genomes into the mitochondrial genome [12]. This suggests the need to develop genetic markers specific to a given species. Limiting intraspecific variation by collecting multiple samples of a species from different geographic origins is also needed to understand the natural intraspecific diversity of wild-crafted herbal materials and prevent potential false-positive or false-negative authentication errors [12].

Another typical authentication method for medicinal plants is the chemical profiling of specialized metabolites. Chemical profiling is not sufficiently accurate for taxonomic identification because the metabolites can differ due to many non-genetic factors, such as climate, time of harvest, and symbiosis. However, chemical profiling has advantages, especially when DNA extraction is impossible. For example, authentication is required for crude extracts or manufactured goods. Previously, we showed that HPLC-UV-based chemical analysis could be used to discriminate between *A. princeps* and *A. capillaris* [13]. It was reported that HPLC-UV [14], GC-MS [15], and LC-MS [16] are powerful tools for chemical profiling of and discrimination between *Artemisia* species. This study suggests chloroplast genome-based DNA barcodes for three *Artemisia* species (i.e., *A. capillaris*, *A. gmelinii*, and *A. fukudo* Makino) for which DNA barcodes were not suggested previously. A previous study suggested that *psbA*—*trnH* is highly variable region in the *Artemisia* species [17]; however, as the region did not show high level of

variations between *A. gmelinii*, *A. capillaris*, and *A. fukudo*, we developed DNA barcoding markers specific to these species based on the other regions. Additionally, we analyzed specialized metabolites of these three species using LC-MS/MS to provide ground knowledge on the development of chemical markers for authentication.

Materials and methods

Plant materials

Samples of *A. gmelinii*, *A. capillaris*, and *A. fukudo* were collected from different regions in Korea and maintained at the Hantaek Botanical Garden (Yongin, Korea) or Department of Herbal Crop Research, Rural Development Administration (Eumseong, Korea). *A. gmelinii* and *A. capillaris* were biologically duplicated by collecting individuals from different geographical locations. The biological duplicates were named *A. gmelinii*-A, *A. gmelinii*-B, *A. capillaris*-A, and *A. capillaris*-B. Detailed information on the collection site and the location of each sample is described in [S1 Table](#). The plants were grown in a greenhouse, with the light, temperature, and humidity maintained at similar levels across the greenhouse.

Chemicals and reagents

HPLC-grade water, MeOH, and MeCN were purchased from Avantor Performance Materials Inc. (Center Valley, PA, USA). Formic acid and leucine enkephalin were purchased from Sigma-Aldrich (St. Louis, MO, USA). Reference standards for the identification of LC-MS/MS peaks were purchased as follows: neochlorogenic acid (**2**) from Interpharm (Koyang, Korea), scopoletin (**12**) from Sigma Aldrich, chlorogenic acid (**6**), hyperoside (**16**), luteolin-7-*O*-glucoside (**18**), and isorhamnetin-3-*O*- β -D-glucoside (**24**) from Chromadex (Irvine, CA, USA).

Chloroplast genome sequencing and annotation

Total genomic DNA was extracted from the leaf tissue of each *Artemisia* species using a modified cetyltrimethylammonium bromide method [18]. Our previous studies described the methods for chloroplast genome assembly of *A. gmelinii*, *A. capillaris*, and *A. fukudo* [19,20]. The three *Artemisia* species were sequenced using the Illumina MiSeq platform (Illumina, San Diego, CA, USA). High-quality paired-end reads of 1.2–2.1 Gb were generated and separately assembled using CLC genome assembler 4.6 (CLC Inc., Aarhus, Denmark). All contigs representing chloroplast sequences were retrieved, ordered, and combined into a single sequence based on the chloroplast genome of *A. frigida* (JX293720) [21]. Sequences can be found in GenBank under the following accession numbers: KU736962 for *A. gmelinii*-A, KY073390 for *A. gmelinii*-B, KU736963 for *A. capillaris*-A, KY073391 for *A. capillaris*-B, and KU360270 for *A. fukudo*.

Complete chloroplast genomes of the three *Artemisia* species were annotated using DOGMA [22] and manually corrected by comparison with chloroplast genomes deposited in GenBank. Chloroplast genome maps of the three *Artemisia* species were drawn using OGDRAW [23]. The genomic differences among the three *Artemisia* species were analyzed using Mvista [24]. The repeat sequences, including forward, palindrome, reverse, and complement, were identified using the REPuter software [25] with the following criteria: cutoff $n \geq 30$ bp and a sequence identity of $\geq 90\%$.

Phylogenetic analysis

A phylogenetic tree was constructed using 64 protein coding sequences extracted from 18 species belonging to the Asteraceae family including *A. fukudo* and 2 sequences each from *A.*

gmelinii and *A. capillaris*. The accession numbers of the chloroplast genomes used for phylogenetic analysis are listed in [S2 Table](#). All coding sequences were aligned using MAFFT [26]. A phylogenetic tree was constructed using the neighbor-joining method with 1000 bootstrap values, as implemented in MEGA 6.0 [27].

Development of InDel barcode markers and blind test analysis

InDel barcode markers were developed based on polymorphic regions identified in the aligned chloroplast genome sequences of *A. gmelinii*, *A. capillaris*, and *A. fukudo*. Interspecific polymorphic targets without intraspecific diversity among three species were selected as candidate regions for marker development. Primer sets were designed using Primer 3 version 0.4.0 [28], and the primer set information can be found in [S3 Table](#). PCR was performed in a final volume of 25 μ L, consisting of 20 ng DNA, 2 units of Taq polymerase (Vivagen, Seongnam, Korea), 2.5 mM dNTPs, and 20 pmol of each primer. The PCR conditions were as follows: 5 min at 94°C; 35 cycles of 30 s at 94°C, 30 s at 54°C, and 20 s at 72°C; and 7 min at 72°C as the final extension. PCR amplicons were analyzed on a 3% agarose gel and visualized under UV trans-illuminator using a gel documentation system.

Five InDel-based barcode markers specific to *A. gmelinii* and *A. capillaris* were used in a blind test on 20 dried commercial *Artemisia* species provided by the Ministry of Food and Drug Safety of Korea. The samples were labeled as follows: seven as *A. gmelinii*, six as *A. capillaris*, six as *A. apiacea*, and one as *A. annua*. The amplified PCR products were genotyped by capillary electrophoresis using a fragment analyzer (Advanced Analytical Technologies Inc., USA), according to the manufacturer's protocol. The genotyping results were obtained by analysis using the PROSize 2.0 software (Agilent Technologies Inc., Santa Clara, CA, USA).

LC-MS/MS-based specialized metabolite profiling

Accurately weighed powdered plant samples (100.0 mg) were extracted with 1.0 mL MeOH/H₂O (5:5, v/v) and sonicated at room temperature for 15 min. The extracts were centrifuged at 15,000 \times g for 10 min, and the supernatant (0.5 mL) was collected. Chicoric acid (Interpharm) was added as an internal standard (IS) at a 0.1 mg/mL final concentration. The samples were filtered through 0.20 μ m Minisart RC15 filters (Sartorius Stedim Biotech, Göttingen, Germany). Each sample was analyzed in triplicate.

The samples were analyzed using a Waters Acquity UPLC system (Waters Co., Milford, MA, USA) coupled with a Waters Xevo G2 Q/TOF mass spectrometer (Waters MS Technologies, Manchester, UK) equipped with an electrospray ionization interface (ESI). Analytes were separated on a Waters Acquity UPLC HSS C₁₈ column (1.8 μ m, 2.1 \times 100 mm) eluted with a mixture of 0.1% formic acid in water (solvent A) and acetonitrile containing 0.1% formic acid (solvent B) with the following gradient: 5% B, 0 min; 10% B, 5 min; 10% B, 6 min; 16% B, 7 min; 20% B, 13 min; 50% B, 18 min; 100% B, 20 min; 100% B, 23 min, 5% B, 23.1 min; 5% B, 26 min. The flow rate was 0.3 mL/min, and the injection volume was 1.0 μ L. MS/MS data were acquired in the data-independent acquisition (MS^E) mode. The detailed parameters of the MS setting are described in the [S1 File](#). The order of data acquisition was randomized. A quality control (QC) sample (a mixture of every analyzed sample) was prepared and used to precondition the UHPLC column with five continuous injections. The QC sample was also injected after every five samples to monitor the precision of the analysis. Raw LC-MS/MS data were deposited in MassIVE (<https://massive.ucsd.edu>) under the accession number MSV000087782.

LC-MS data were processed to extract MS features and peak area-based intensities using MZmine 2.53 software [29]. The detailed parameters are described in [S4 Table](#). The intensity

of each ion was normalized to the IS intensity. Principal component analysis (PCA) was performed using the Python package *pca* (<https://github.com/erdogant/pca>). One-way ANOVA was performed using GraphPad Prism (version 7.0; GraphPad Software, Inc., La Jolla, CA, USA).

Results and discussion

Development of chloroplast genome in the three *Artemisia* species

Previous studies assembled the chloroplast genomes of three *Artemisia* species: *A. fukudo* [30], *A. gmelinii*, and *A. capillaris* [31]. In this study, we completed chloroplast genome sequences from different individuals of *A. gmelinii* (*A. gmelinii*-B) and *A. capillaris* (*A. capillaris*-B) and compared them with previously characterized sequences (*A. gmelinii*-A and *A. capillaris*-A) (Fig 1). The completed chloroplast genomes had a typical quadripartite structure consisting of a long single-copy section (LSC), inverted repeat (IR), and short single-copy section (SSC) (Table 1). The genome sizes were similar among the three species, ranging from 151,011 bp to 151,318 bp. Variation was detected within the same species: 151,318 bp for *A. gmelinii*-A and 151,050 bp for *A. gmelinii*-B, and 151,056 bp for *A. capillaris*-A and 151,020 bp for *A. capillaris*-B. The genome size of *A. fukudo* was 151,011 bp. The variations in genome size between intra- and inter-species were mainly due to the LSC region, as the lengths varied from 82,751 to 83,061 bp. Significant length differences of 271bp in the LSC region were observed between *A. gmelinii*-A and *A. gmelinii*-B. This difference came from 43 intraspecific InDels in the LSC region; most were in intergenic regions, except for three exonic ones. The average GC content of the chloroplast genomes was almost identical (37.4–37.5%) in the three species. We constructed a phylogenetic tree with 64 protein-coding sequences from the chloroplast genomes of 18 species, including seven tribes and three subfamilies of the Asteraceae family (Fig 2). For *Artemisia*, three additional species (i.e., *A. frigida*, *A. montana*, and *A. argyi*), were incorporated, as these chloroplast genome sequences are publicly available in NCBI. Overall, the phylogenetic tree was well-clustered by tribes and subfamilies. The six *Artemisia* species were grouped in the same clade along with *Chrysanthemum* × *morifolium* within the tribe Anthemideae, which is consistent with the phylogenetic analysis based on the protein-coding regions in chloroplasts [32]. Two individuals of *A. gmelinii* and *A. capillaris* collected from different locations were placed in the same clade, indicating that the individuals were most likely the same species.

Comparative analysis on chloroplast genomes for barcode marker development

For a deeper investigation of variations between the *Artemisia* chloroplast genomes, three *Artemisia* species were compared to previously reported *A. frigida* [21], with annotation of *A. gmelinii*-A as a reference. While more sequence variations were found in non-coding regions than in coding regions in all species, the 47 genes had at least one genic SNP. Seven genes, i.e., *matK*, *rpoB*, *rpoC2*, *accD*, *ndhF*, *ycf2*, and *ycf1*, were highly divergent among the three *Artemisia* species (Fig 3). Intraspecific variations were found in fifteen genes of *A. gmelinii* and twelve of *A. capillaris* (Figs 1 and 3). The coding regions of *A. fukudo* were less conserved than those of the other *Artemisia* species. As in other plant species [16,21,33], *ycf1* and *accD* were highly varied regions, suggesting these particular genes are prone to natural variations in plant chloroplasts. Regarding SNPs and InDels, more intraspecific differences were observed in *A. gmelinii* than in *A. capillaris*, with 107 and 30 SNPs, respectively. The number of InDels was 43 and 10 for *A. gmelinii* and *A. capillaris*, respectively (Table 2). Individuals of the same species might have undergone independent mutation events in different growing regions.

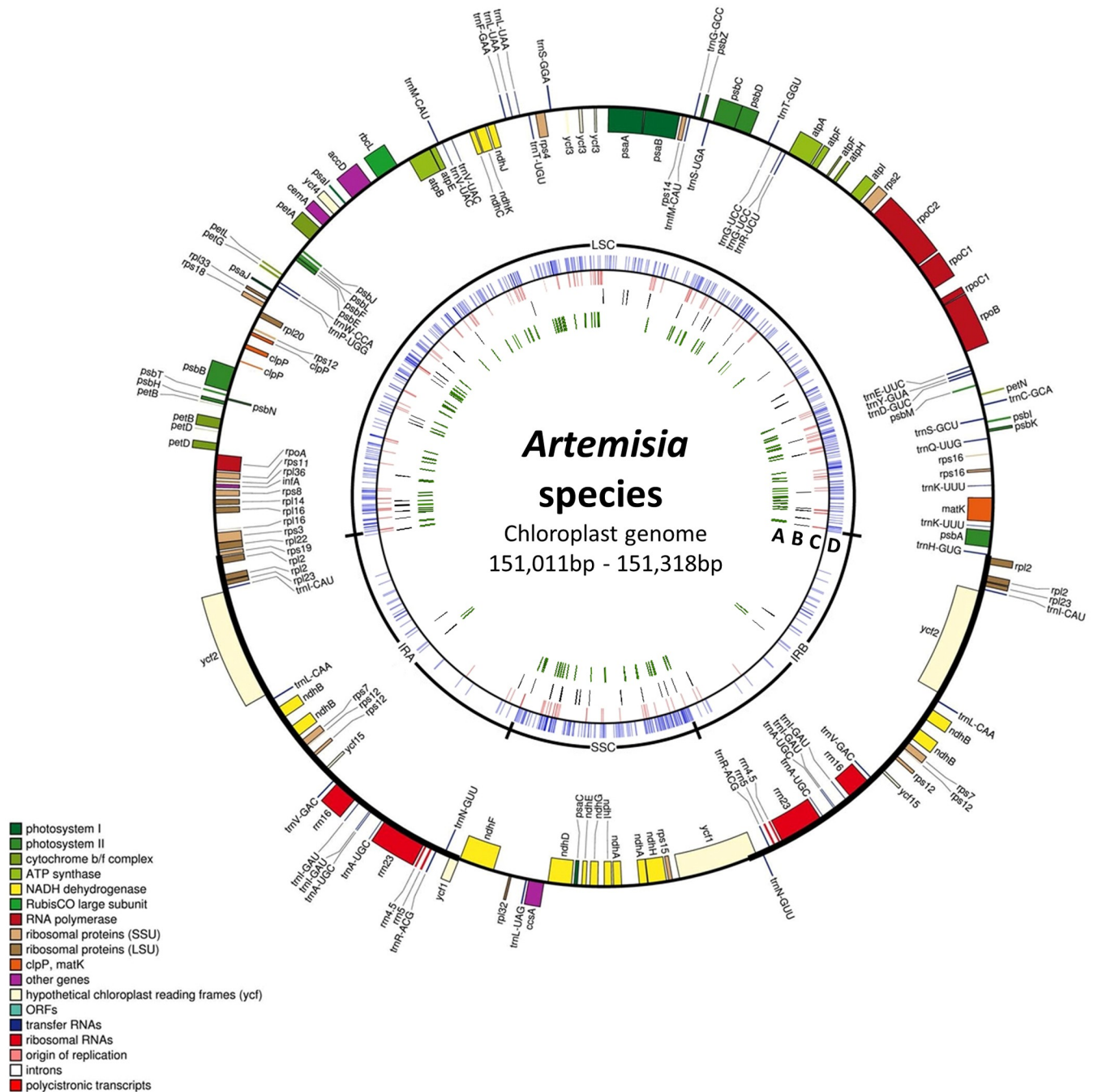


Fig 1. Chloroplast genome maps of *A. gmelinii*, *A. capillaris*, and *A. fukudo*. Genes represented on the outside of the outer circle are transcribed clockwise, and those shown on the inside of the outer circle are transcribed counterclockwise. Genes belonging to the same functional group were indicated in the same color. The inner tracks represent inter- and intraspecific variations. Track A and B are intraspecific variations of *A. gmelinii* and *A. capillaris*, respectively. Track C and D indicate total SNPs and InDels among five *Artemisia* chloroplast genomes, respectively.

<https://doi.org/10.1371/journal.pone.0264576.g001>

Comparative analyses of the chloroplast genomes revealed the possibility of developing InDel-based barcode markers, providing useful guidance for identifying *Artemisia* species. We designed seven markers ar09, ar16, ar20, ar32, ar42, ar44, and ar46 based on the polymorphic regions of *petN-psbM*, *psaA-ycf3*, *ycf3-trnS(GGA)*, *rps11-rpl36*, *ycf1*, *psbZ-trnG(GCC)*, and

Table 1. Summary of the NGS data and chloroplast genomes of *A. gmelinii*, *A. capillaris*, and *A. fukudo*.

sample	number of raw reads	total read length (bp)	mapped read number	chloroplast genome coverage (x)	chloroplast genome length (bp)	long single copy (LSC)	inverted repeat (IR)	short single copy (SSC)
<i>A. gmelinii</i> -A	7,295,868	2,191,126,001	187,330	300.24	151,318	83,061	24,961	18,335
<i>A. gmelinii</i> -B	8,762,916	2,626,280,574	52,077	78.01	151,050	82,790	24,961	18,338
<i>A. capillaris</i> -A	5,360,722	1,610,083,995	121,704	193.74	151,056	82,821	24,963	18,309
<i>A. capillaris</i> -B	9,532,322	2,859,007,916	143,982	217.70	151,020	82,786	24,963	18,302
<i>A. fukudo</i>	4,314,036	1,289,617,713	61,476	93.82	151,011	82,751	24,956	18,348

<https://doi.org/10.1371/journal.pone.0264576.t001>

rpl32-trn(UAG) (Figs 4 and S1). Five markers, ar09, ar16, ar20, ar32, and ar42, were used to distinguish between *A. gmelinii* and *A. capillaris*. Two markers, ar44 and ar46, were specifically identified in *A. fukudo*. PCR analysis showed that the developed markers produced PCR amplicons of different sizes, regardless of intraspecific variation in *A. capillaris* and *A. gmelinii*, with successful discrimination among the three species (Fig 4).

We then tested whether our barcode markers could be applied to real-world *Artemisia* products. Twenty dried herbs labeled *A. gmelinii*, *A. capillaris*, *A. annua*, and *A. apiacea*, were collected from the market and tested. Five markers, ar09, ar16, ar20, ar32, and ar42,

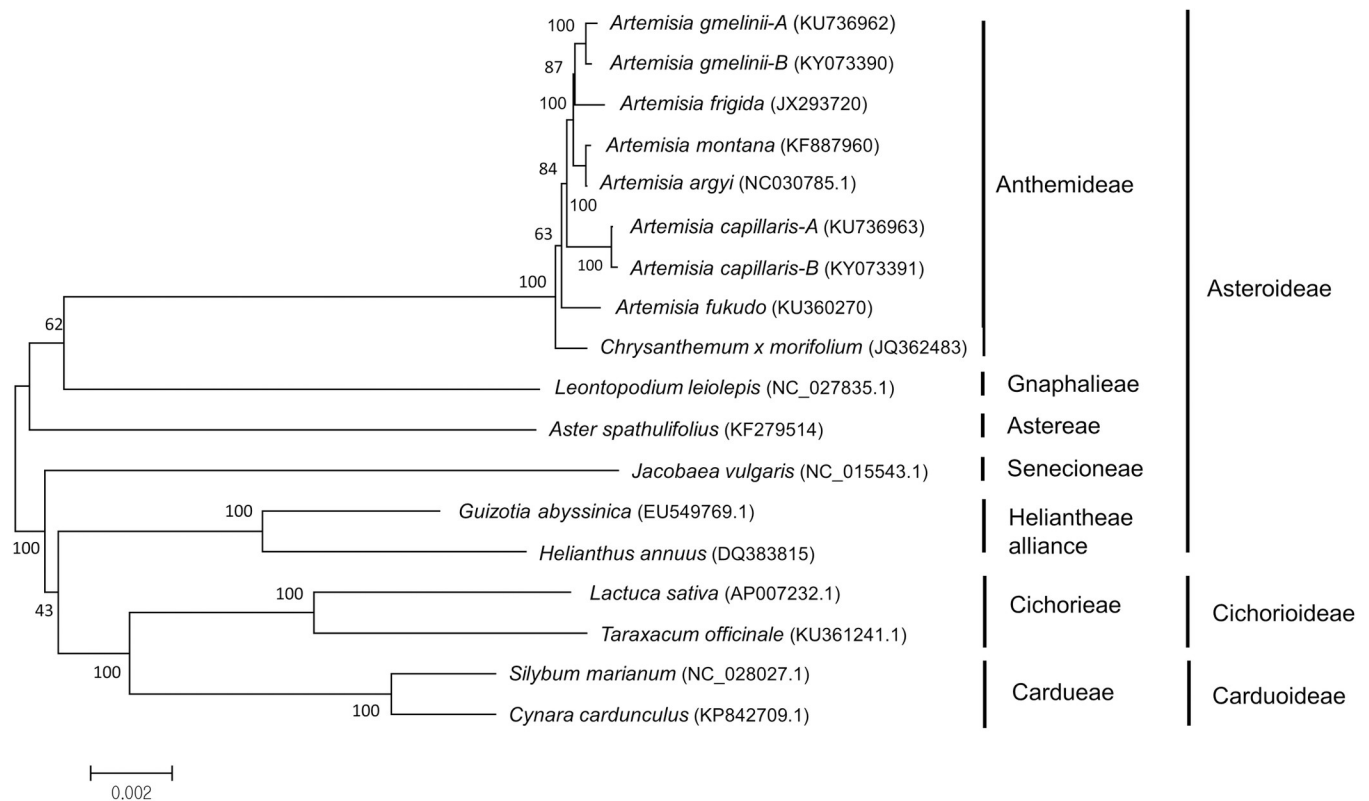


Fig 2. A phylogenetic tree of 18 species belonging to the Asteraceae family. A total of 64 coding regions in the chloroplast genome were used to analyze phylogenetic relationships among the species. The phylogenetic tree was constructed by the neighbor-joining method with 1000 bootstrap values. Numbers on each node were bootstrap values. The accession number of the chloroplast genomes used for the phylogenetic tree are in parenthesis and S2 Table.

<https://doi.org/10.1371/journal.pone.0264576.g002>

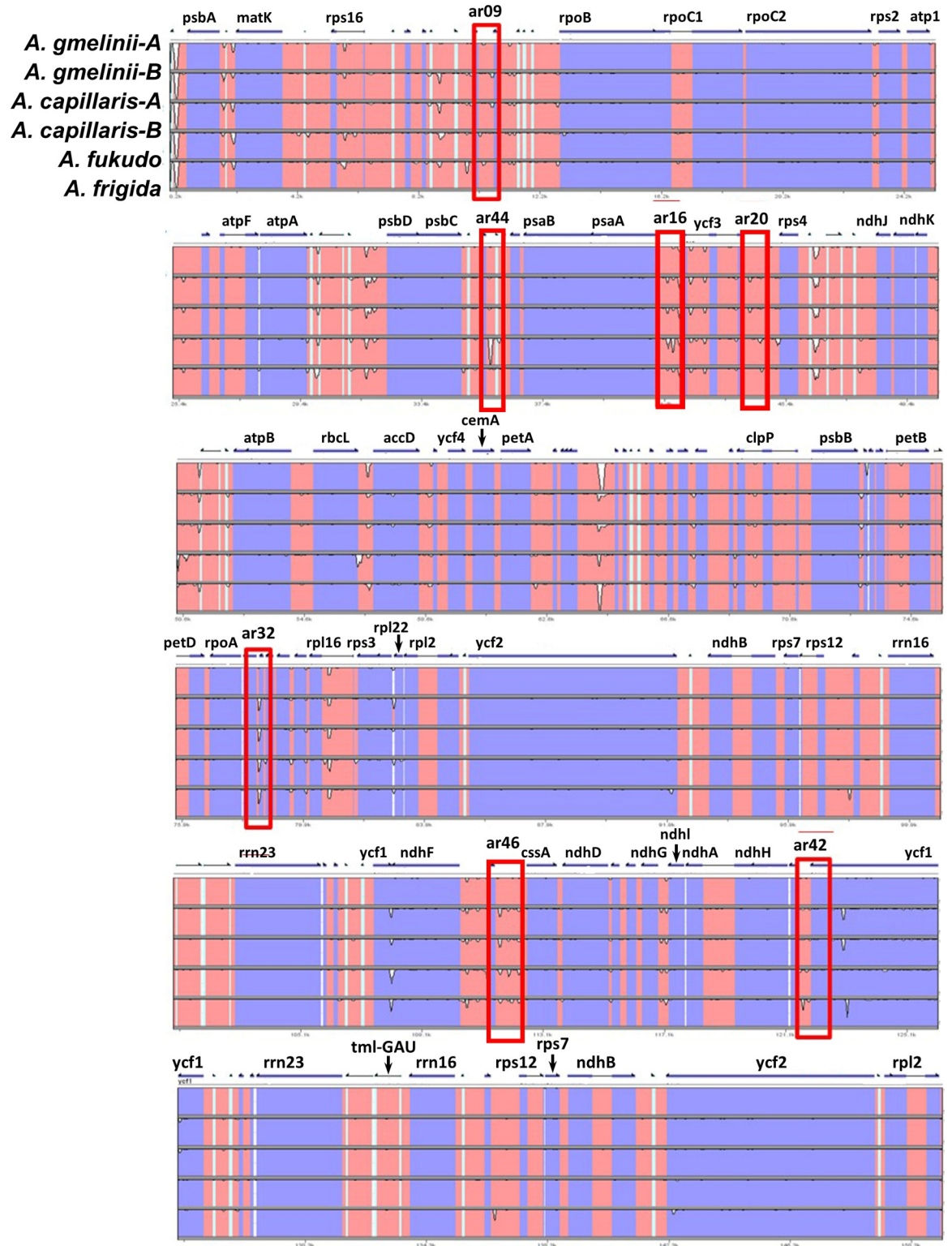


Fig 3. Chloroplast sequence variation between *A. gmelinii*, *A. capillaris*, *A. fukudo*, and *A. frigida*. The molecular markers were developed from the highly variable regions indicated by a red rectangle. The names of markers are shown above the red boxes.

<https://doi.org/10.1371/journal.pone.0264576.g003>

Table 2. The number of SNPs and InDels found in chloroplast genomes of three *Artemisia* species.

Samples	<i>A. gmelinii</i> -A	<i>A. gmelinii</i> -B	<i>A. capillaris</i> -A	<i>A. capillaris</i> -B	<i>A. fukudo</i>
<i>A. gmelinii</i> -A		43	102	104	119
<i>A. gmelinii</i> -B	107 (0.07)		91	91	107
<i>A. capillaris</i> -A	448 (0.30)	456 (0.30)		10	114
<i>A. capillaris</i> -B	458 (0.30)	462 (0.31)	30 (0.02)		113
<i>A. fukudo</i>	453 (0.30)	463 (0.31)	544 (0.36)	548 (0.36)	

The upper triangle shows the number of InDels, and the lower triangle indicates the total nucleotide substitutions. The percentage ratios of nucleotide variations to chloroplast genome sequences are given in brackets.

<https://doi.org/10.1371/journal.pone.0264576.t002>

successfully distinguished *A. gmelinii* and *A. capillaris* (S2 Fig). We found that *A. annua* and *A. apiacea* exhibited the same band pattern as *A. gmelinii* and *A. capillaris* on the gel, except for the result from the ar16 marker. This suggests that the two herbs might outcross, and multiple barcode markers are necessary for accurate validation [12]. This result demonstrated that InDel-based barcode markers could be utilized in the identification and authentication of *Artemisia* species.

LC-MS/MS-based analysis on specialized metabolome

Although chemical markers are useful for authentication of herbal products from which DNA extraction is impossible, it is difficult to develop valid chemical markers for identification, because the contents of specialized metabolites are heavily influenced by environmental factors. Nevertheless, we analyzed the *Artemisia* samples using LC-MS/MS, as a preliminary effort to gather background knowledge on their specialized metabolome. To minimize environmental effects on the accumulation or induction of specialized metabolites, all the plants were

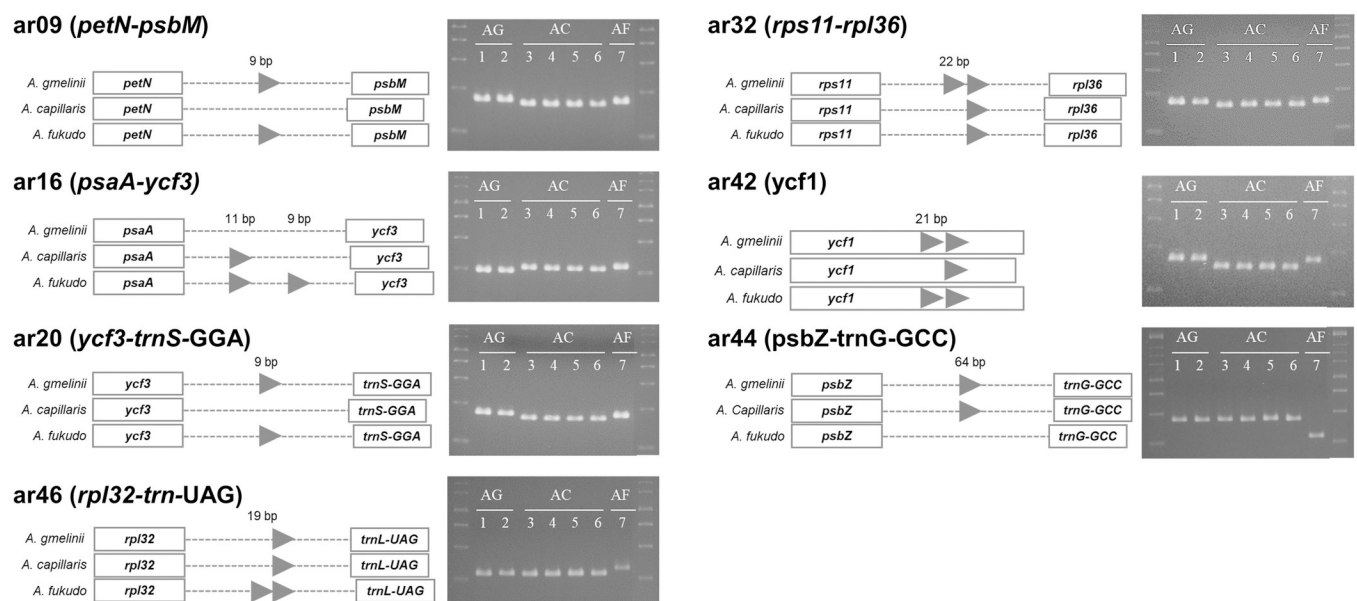


Fig 4. Discrimination success for *A. gmelinii*, *A. capillaris*, and *A. fukudo* using InDel-based barcode markers. Rectangles and dashed lines indicate the genes and intergenic regions, respectively. Arrows represent the inserted sequences within the intergenic regions. The amplified DNA fragments using the InDel-based barcode markers were separated on an agarose gel as shown on the right side (AG, *A. gmelinii*; AC, *A. capillaris*; AF, *A. fukudo*). The detailed sequence variations between the species are presented in S1 Fig.

<https://doi.org/10.1371/journal.pone.0264576.g004>

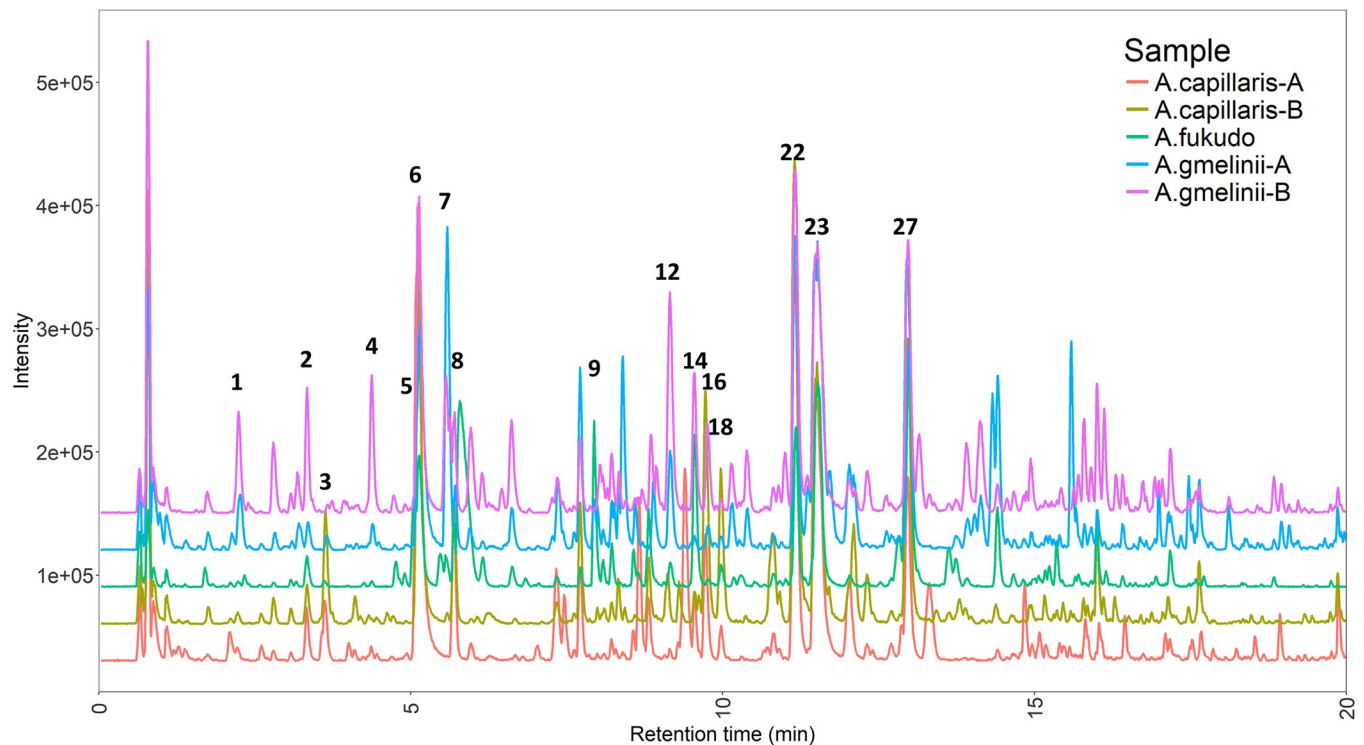


Fig 5. Representative LC-MS base peak ion (BPI) chromatograms of the *Artemisia* extracts. Tentatively annotated chromatographic peaks are annotated with peak numbers. Gaps between the chromatograms were added to help visualize the differences, so the y-axis values do not equal the absolute intensities.

<https://doi.org/10.1371/journal.pone.0264576.g005>

grown in a same greenhouse. As shown in Fig 5, the chemical profiles of the *Artemisia* extracts showed qualitative and quantitative interspecific variety. Twenty-eight peaks were putatively identified in the different Metabolomics Standards Initiative (MSI) confidence levels [34] (Table 3). Six peaks (2, 6, 12, 16, 18, and 25) were identified at MSI level 1 using commercial standards, seventeen peaks were annotated at MSI level 2 by MS/MS spectral matching to the GNPS spectral library [35] (S3 Fig), and five peaks were annotated at level 3 by comparison of relative retention time and MS data with previous studies [16,36]. Most of the identified metabolites were polyphenols such as phenylpropanoids, flavonoids, and coumarins.

For the comparative analysis between the specialized metabolome of three *Artemisia* species, we extracted and normalized 1,763 MS features from the LC-MS data and applied them to principal component analysis (PCA). After technical validation with QC samples, the PCA biplot was visualized to reveal the chemical markers showing the most significantly different content between samples. The PC1 score, accounting for 42.3% of the total variance, discriminated *A. capillaris* from the other two species, while the PC2 score discriminated *A. gmelinii* and *A. fukudo*, accounting for 28.3% of the total variance (Fig 6A). PCA loading vectors revealed which MS features show relatively large ion intensity in each species. Caffeic acid (5, marked as MS feature 4 in Fig 6A) was abundant in *A. fukudo*, while *p*-coumaric acid hexoside (7, MS feature 357) and scopoletin (12, MS feature 42) showed relatively high contents in *A. gmelinii*. Multiple compounds such as 3-*O*-caffeoylquinic acid (6, MS features 9 and 10), 3,4-*O*-dicafeoylquinic acid (22, MS feature 6), 1,5-*O*-dicafeoylquinic acid (23, MS features 371 and 374), and hyperoside (16, MS feature 1226) showed relatively high ion abundances in the two *A. capillaris* samples compared to the others.

The ion abundances of MS features suggested as discriminating variables in PCA (peaks 5, 6, 7, 12, 16, 22, and 23) were further compared as shown in Fig 6B. Peak 5 showed significantly

Table 3. Chromatographic and spectrometric data of the 28 tentatively identified metabolites from *Artemisia* species.

peak number	identification	t _R (min)	calculated m/z [M-H] ⁻	observed m/z [M-H] ⁻	mass error (ppm)	molecular formula	MSI identification level
1	dihydroxybenzoic acid hexoside	2.2	315.1716	315.0718	0.6	C ₁₃ H ₁₆ O ₉	3
2	neochlorogenic acid	3.3	353.0873	353.0873	0.0	C ₁₆ H ₁₈ O ₉	1
3	dihydroxybenzoic acid pentoside	3.7	285.0610	285.0605	-1.8	C ₁₂ H ₁₄ O ₈	2
4	1,3- <i>O</i> -dicaffeoylquinic acid	4.4	515.1401	515.1401	0.0	C ₂₂ H ₂₈ O ₁₄	3
5	caffeic acid	5.1	179.0344	179.0338	-3.4	C ₉ H ₈ O ₄	2
6	3- <i>O</i> -caffeoylquinic acid	5.1	353.0873	353.0871	-0.6	C ₁₆ H ₁₈ O ₉	1
7	<i>p</i> -coumaric acid hexoside	5.6	325.0923	325.0919	-1.2	C ₁₅ H ₁₈ O ₈	2
8	4- <i>O</i> -caffeoylquinic acid	5.7	353.0873	353.0869	-1.1	C ₁₆ H ₁₈ O ₉	2
9	apigenin 6,8-di- <i>C</i> -hexoside	7.9	593.1506	593.1497	-1.5	C ₂₇ H ₃₀ O ₁₅	2
10	3- <i>O</i> -feruloylquinic acid	8.2	367.1029	367.1024	-1.4	C ₁₇ H ₂₀ O ₉	2
11	methyl chlorogenate	8.7	367.1029	367.1021	-2.2	C ₁₇ H ₂₀ O ₉	2
12	scopoletin ^a	9.2	191.0344	191.0340	-2.1	C ₁₀ H ₈ O ₄	1
13	feruloyl hexose	9.4	355.1029	355.1015	-3.9	C ₁₆ H ₂₀ O ₉	2
14	quercetin 3- <i>O</i> -neohesperidoside	9.6	609.1456	609.1470	2.3	C ₂₇ H ₃₀ O ₆	2
15	monohydroxy-dimethoxycoumarin	9.8	221.0450	221.0442	-2.3	C ₁₁ H ₁₀ O ₅	3
16	hyperoside	9.8	463.0877	463.0875	-0.4	C ₂₁ H ₂₀ O ₁₂	1
17	isoquercitrin	10.0	463.0877	463.0874	-0.6	C ₂₁ H ₂₀ O ₁₂	2
18	luteolin 7- <i>O</i> -glucoside	10.3	447.0927	447.0932	1.1	C ₂₁ H ₂₀ O ₁₁	1
19	methoxy-pentahydroxy(iso)flavone- <i>O</i> -hexoside	10.4	493.0982	493.0981	-0.2	C ₂₂ H ₂₂ O ₁₃	3
20	quercetin 3- <i>O</i> -acetylhexoside	10.8	505.0982	505.0982	0.0	C ₂₃ H ₂₂ O ₁₃	2
21	kaempferol-3- <i>O</i> -rutinoside	11.0	593.1506	593.1509	0.5	C ₂₇ H ₃₀ O ₁₅	2
22	3,4- <i>O</i> -dicaffeoylquinic acid	11.2	515.1190	515.1202	2.3	C ₂₅ H ₂₄ O ₁₂	2
23	1,5- <i>O</i> -dicaffeoylquinic acid	11.4	515.1190	515.1202	2.3	C ₂₅ H ₂₄ O ₁₂	3
24	3,5- <i>O</i> -dicaffeoylquinic acid	11.5	515.1190	515.1205	2.9	C ₂₅ H ₂₄ O ₁₂	2
25	isorhamnetin 3- <i>O</i> -β- <i>D</i> -glucoside	12.1	477.1030	477.1028	0.4	C ₂₂ H ₂₂ O ₁₂	1
26	quercetin 3- <i>O</i> -acetylhexoside	12.3	505.0982	505.0983	0.2	C ₂₃ H ₂₂ O ₁₃	2
27	4,5- <i>O</i> -dicaffeoylquinic acid	13.0	515.1190	515.1184	-1.2	C ₂₅ H ₂₄ O ₁₂	2
28	monohydroxy-dimethoxycoumarin	14.2	221.0450	221.0443	-3.2	C ₁₁ H ₁₀ O ₅	2

<https://doi.org/10.1371/journal.pone.0264576.t003>

higher contents in *A. fukudo* than in the others, which supported **5** as a potential chemical marker for *A. fukudo*. Peak **16** also showed species-specific occurrence in *A. capillaris*, although the relative abundance of **16** showed 2-fold differences between biological replicates. Peak **12** were comparatively abundant in *A. gmelinii* while **23** showed high concentration in *A. fukudo*; however, *A. fukudo* also contained both compounds, so further data should be acquired from more individuals to validate if this trend is general. In contrast, peaks **6** and **22** were ubiquitous in the three species, and their interspecific variations were not significantly larger than the intraspecific variations within *A. gmelinii* and *A. capillaris* (S4 Fig).

Peak **7** highlighted the reason why a large population of sample should be considered for developing chemical markers. Despite its abundance in *A. gmelinii*-A, it showed only a trace amount in *A. gmelinii*-B. This result suggests that the intraspecific genetic variation can cause about 100-fold difference on a level of a metabolite, as we controlled the environmental factors by growing these samples within a same greenhouse. Thus, more data from samples acquired from larger genetic pools, in addition to larger environmental and geographical varieties, should be inspected to develop valid chemical markers for *Artemisia* species.

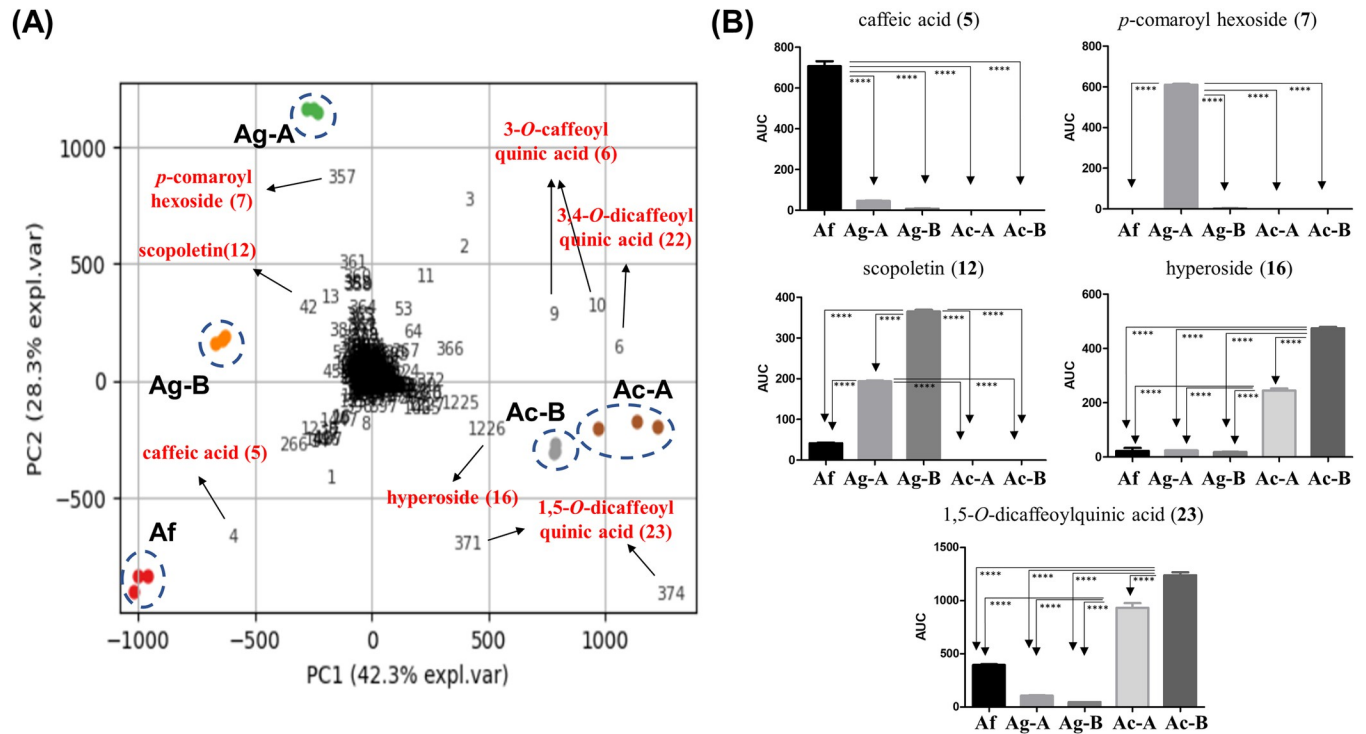


Fig 6. (A) The PCA biplot of the *Artemisia* samples represented by group marks (Af: *A. fukudo*; Ag-A: *A. gmelinii*-A; Ag-B: *A. gmelinii*-B; Ac-A: *A. capillaris*-A; Ac-B: *A. capillaris*-B). The numbers denote PCA loadings of each MS feature. Important marker MS features are labeled with their annotation and peak numbers enlisted in Table 3. Most MS features shown here are $[M-H]^-$ ions, but MS features 9 and 372 are in-source fragments of peaks 6 and 23, respectively. (B) Bar plots showing the ion intensities of selected marker peaks (5, 7, 12, 16, and 23) in analyzed *Artemisia* samples (**** $p < 0.0001$).

<https://doi.org/10.1371/journal.pone.0264576.g006>

Conclusion

This study identified molecular markers that can be used for the authentication of three medicinal plants similar in appearance: *A. capillaris*, *A. gmelinii*, and *A. fukudo*. Comparative analysis of the complete chloroplast genomes of these plants suggested seven InDel-based DNA barcodes, and these markers successfully identified *Artemisia* species in the market. In addition, we analyzed specialized metabolites of the *Artemisia* species grown in a same environment using LC-MS/MS. It revealed multiple metabolites showing significant interspecific variation. These metabolites can be potential chemical markers, although further intensive efforts are required to validate those as chemical markers. By integrating DNA barcoding and chemical profiling, these species can be identified, authenticated, and monitored throughout the production process of botanical goods i.e., cultivation, harvest, circulation, and extraction.

Supporting information

S1 Fig. Sequence variation in the developed DNA barcode markers between species. AG, *A. gmelinii*; AC, *A. capillaris*; AF, *A. fukudo*.
(PDF)

S2 Fig. The result of the blind test on commercial *Artemisia* samples, using five InDel-based barcode markers for identification.
(PDF)

S3 Fig. The mirror plot of experimental MS/MS spectra (upper, black) and GNPS reference spectra (lower, green) for the metabolite annotation.

(PDF)

S4 Fig. Bar plots showing the ion intensities of peaks 6 and 22 in analyzed *Artemisia* samples.

(PDF)

S1 Table. Detailed information of the plant materials used in this study.

(PDF)

S2 Table. Accession numbers of the chloroplast genomes used in this study.

(PDF)

S3 Table. Primer sets used in this study.

(PDF)

S4 Table. Detailed parameters used for the feature finding in MZmine 2.

(PDF)

S1 File. Detailed MS acquisition parameters.

(PDF)

S1 Raw images. Raw gel images.

(PDF)

Author Contributions

Conceptualization: Yun Sun Lee, Sang Hyun Sung, Tae-Jin Yang.

Data curation: Yun Sun Lee, Sunmin Woo, Jin-Kyung Kim, Jee Young Park, Nur Kholilatul Izzah, Hyun-Seung Park, Kyo Bin Kang.

Funding acquisition: Sang Hyun Sung, Tae-Jin Yang.

Investigation: Yun Sun Lee, Sunmin Woo, Jin-Kyung Kim, Nur Kholilatul Izzah, Hyun-Seung Park.

Methodology: Yun Sun Lee, Kyo Bin Kang.

Project administration: Jee Young Park.

Resources: Jung Hwa Kang, Taek Joo Lee.

Supervision: Sang Hyun Sung, Kyo Bin Kang, Tae-Jin Yang.

Visualization: Yun Sun Lee, Sunmin Woo, Hyun-Seung Park, Kyo Bin Kang.

Writing – original draft: Yun Sun Lee, Sunmin Woo, Jee Young Park, Kyo Bin Kang.

Writing – review & editing: Yun Sun Lee, Hyun-Seung Park, Kyo Bin Kang, Tae-Jin Yang.

References

1. Garcia S, Sanz M, Garnatje T, Kreitschitz A, McArthur ED, Vallès J. Variation of DNA amount in 47 populations of the subtribe Artemisiinae and related taxa (Asteraceae, Anthemideae): karyological, ecological, and systematic implications. *Genome*. 2004 Dec; 47(6): 1004–14. <https://doi.org/10.1139/g04-061> PMID: 15644958.
2. Vallès J, McArthur ED. *Artemisia* systematics and phylogeny: cytogenetic and molecular insights. In: McArthur E. Duran; Fairbanks, Daniel J. comps. *Shrubland ecosystem genetics and biodiversity*:

- proceedings; 2000 June 13–15; Provo, UT. Proc. RMRS-P-21. Ogden, UT: US Department of Agriculture, Forest Service, Rocky Mountain Research Station. 2001. 67–74.
3. Lee JH, Lee JW, Sung JS, Bang KH, Moon SG. Molecular authentication of 21 Korean *Artemisia* species (Compositae) by polymerase chain reaction-restriction fragment length polymorphism based on trnL-F region of chloroplast DNA. *Biol. Pharm. Bull.* 2009 Nov. 32(11): 1912–16. <https://doi.org/10.1248/bpb.32.1912> PMID: 19881307.
 4. Bisht D, Kumar D, Kumar D, Dua K, Chellappan DK. Phytochemistry and pharmacological activity of the genus *Artemisia*. *Arch. Pharm. Res.* 2001 Apr. 44(5): 439–74. <https://doi.org/10.1007/s12272-021-01328-4> PMID: 33893998.
 5. Miller LH, Su X. Artemisinin: Discovery from the Chinese herbal garden. *Cell.* 2011 Sep. 146(6): 855–8. <https://doi.org/10.1016/j.cell.2011.08.024> PMID: 21907397.
 6. Torrell M, Garcia-Jacas N, Susanna A, Vallès J. Infrageneric phylogeny of the genus *Artemisia* L. (Asteraceae, Anthemideae) based on nucleotide sequences of nuclear ribosomal DNA internal transcribed spacers (ITS). *Taxon.* 1999. 48: 721–36.
 7. Chen S, Yao H, Han J, Liu C, Song J, Shi L, et al. Validation of the ITS2 region as a novel DNA barcode for identifying medicinal plant species. *PLoS One.* 2010 Jan. 5(1): e8613. <https://doi.org/10.1371/journal.pone.0008613> PMID: 20062805.
 8. Lee MY, Doh EJ, Park CH, Kim YH, Kim ES, Ko BS, et al. Development of SCAR marker for discrimination of *Artemisia princeps* and *A. argyi* from other *Artemisia* herbs. *Biol. Pharm. Bull.* 2006 Apr. 29(4): 629–33. <https://doi.org/10.1248/bpb.29.629> PMID: 16595892.
 9. Song M, Li J, Xiong C, Liu H, Liang J. Applying high-resolution melting (HRM) technology to identify five commonly used *Artemisia* species. *Sci. Rep.* 2016 Oct. 6: 34133. <https://doi.org/10.1038/srep34133> PMID: 27698485.
 10. Mei Q, Chen X, Xiang L, Liu Y, Su Y, Gao Y, et al. DNA barcode for identifying folium *Artemisiae Argyi* from counterfeits. *Biol. Pharm. Bull.* 2016. 39(9): 1531–7. <https://doi.org/10.1248/bpb.b16-00336> PMID: 27582332.
 11. Liu G, Ning H, Ayidaerhan N, Aisa HA. Evaluation of DNA barcode candidates for the discrimination of *Artemisia* L. Mitochondrial DNA Part A 2017, 28(6):956–964. <https://doi.org/10.1080/24701394.2016.1219729> PMID: 27607516.
 12. Park HS, Jayakodi M, Lee SH, Jeon JH, Lee HO, Park JY et al. Mitochondrial plastid DNA can cause DNA barcoding paradox in plants. *Sci. Rep.* 2020 Apr. 10(1): 6112. <https://doi.org/10.1038/s41598-020-63233-y> PMID: 32273595.
 13. Yang H, Lee DY, Jeon M, Suh Y, Sung SH. Determination of five active compounds in *Artemisia princeps* and *A. capillaris* based on UPLC-DAD and discrimination of two species with multivariate analysis. *Arch. Pharm. Res.* 2014 May. 37(5): 617–25. <https://doi.org/10.1007/s12272-013-0204-5> PMID: 23835834.
 14. Alaerts G, Pieters S, Logie H, Van Erps J, Merino-Arévalo M, Dejaegher, et al. Exploration and classification of chromatographic fingerprints as additional tool for identification and quality control of several *Artemisia* species. *J. Pharm. Biomed. Anal.* 2014 Jul. 95: 34–46. <https://doi.org/10.1016/j.jpba.2014.02.006> PMID: 24637050.
 15. Guo L, Zhang D, Wang L, Xue Z, Guo M, Duan L, et al. Comparison and discrimination of *Artemisia argyi* and *Artemisia lavandulifolia* by gas chromatography-mass spectrometry-based metabolomic approach. *J. AOAC Int.* 2019 Nov. 102(6): 1814–21. <https://doi.org/10.5740/jaoacint.19-0080> PMID: 31288892.
 16. Olennikov DN, Chirikova NK, Kashchenko NI, Nikolaev VM, Kim SW, Vennos C. Bioactive phenolics of the genus *Artemisia* (Asteraceae): HPLC-DAD-ESI-TQ-MS/MS profile of the Siberian species and their inhibitory potential against α -amylase and α -glucosidase. *Front. Pharmacol.* 2018 Jul. 9:756 <https://doi.org/10.3389/fphar.2018.00756> PMID: 30050443.
 17. Kim GB, Lim CE, Kim JS, Kim K, Lee JH, Yu HJ, et al. Comparative chloroplast genome analysis of *Artemisia* (Asteraceae) in East Asia: insights into evolutionary divergence and phylogenomic implications. *BMC Genomics* 2020. 21(1):415. <https://doi.org/10.1186/s12864-020-06812-7> PMID: 32571207.
 18. Allen GC, Flores-Vergara MA, Krasynanski S, Kumar S, Thompson WF. A modified protocol for rapid DNA isolation from plant tissues using cetyltrimethylammonium bromide. *Nat. Protoc.* 2006. 1(5): 2320–5. <https://doi.org/10.1038/nprot.2006.384> PMID: 17406474.
 19. Kim K, Lee SC, Lee J, Lee HO, Joh HJ, Kim NH et al. Comprehensive survey of genetic diversity in chloroplast genomes and 45S rDNAs within *Panax ginseng* species. *PLoS One.* 2015 Jun. 10(6): e0117159. <https://doi.org/10.1371/journal.pone.0117159> PMID: 26061692.
 20. Kim K, Lee SC, Lee J, Yu Y, Yang K, Choi BS et al. Complete chloroplast and ribosomal sequences for 30 accessions elucidate evolution of *Oryza* AA genome species. *Sci. Rep.* 2015 Oct. 5: 15655. <https://doi.org/10.1038/srep15655> PMID: 26506948.

21. Liu Y, Huo N, Dong L, Wang Y, Zhang S, Young HA et al. Complete chloroplast genome sequences of *Mongolia* medicine *Artemisia frigida* and phylogenetic relationships with other plants. *PLoS One*. 2013. 8(2): e57533. <https://doi.org/10.1371/journal.pone.0057533> PMID: 23460871.
22. Wyman SK, Jansen RK, Boore JL. Automatic annotation of organellar genomes with DOGMA. *Bioinformatics*. 2004 Nov. 20(17): 3252–5. <https://doi.org/10.1093/bioinformatics/bth352> PMID: 15180927.
23. Greiner S, Lehwark P, Bock R. OrganellarGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes. *Nucleic Acids Res*. 2019 Jul. 47(W1): W59–W64. <https://doi.org/10.1093/nar/gkz238> PMID: 30949694.
24. Frazer KA, Pachter L, Poliakov A, Rubin EM, Dubchak I. VISTA: computational tools for comparative genomics. *Nucleic Acids Res*. 2004 Jul. 32: W273–9. <https://doi.org/10.1093/nar/gkh458> PMID: 15215394.
25. Kurtz S, Choudhuri JV, Ohlebusch E, Schleiermacher C, Stoye J, Giegerich R. REPuter: the manifold applications of repeat analysis on a genomic scale. *Nucleic Acids Res*. 2001 Nov. 29(22): 4633–42. <https://doi.org/10.1093/nar/29.22.4633> PMID: 11713313.
26. Katoh K, Misawa K, Kuma K, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res*. 2002 Jul. 30(14): 3059–66. <https://doi.org/10.1093/nar/gkf436> PMID: 12136088.
27. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol. Biol. Evol*. 2013 Dec. 30(12): 2725–9. <https://doi.org/10.1093/molbev/mst197> PMID: 24132122.
28. Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M et al. Primer3—new capabilities and interfaces. *Nucleic Acids Res*. 2012 Aug. 40(15): e115–e115. <https://doi.org/10.1093/nar/gks596> PMID: 22730293.
29. Pluskal T, Castillo S, Villar-Briones A, Oresic M et al. MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC Bioinformatics* 2010 Jul. 11: 395. <https://doi.org/10.1186/1471-2105-11-395> PMID: 20650010.
30. Lee YS, Park JY, Kim JK, Lee HO, Park HS, Lee SC et al. Complete chloroplast genome sequence of *Artemisia fukudo* Makino (Asteraceae). *Mitochondrial DNA B Resour*. 2016 Jun. 1(1): 376–7. <https://doi.org/10.1080/23802359.2016.1155426> PMID: 33490395.
31. Lee YS, Park JY, Kim JK, Lee HO, Park HS, Lee SC et al. The complete chloroplast genome sequences of *Artemisia gmelinii* and *Artemisia capillaris* (Asteraceae). *Mitochondrial DNA B Resour*. 2016 Jul. 1(1): 410–1. <https://doi.org/10.1080/23802359.2016.1176880> PMID: 33473501.
32. Kang SH, Kim K, Lee JH, Ahn BO, Won SY, Sohn SH et al. The complete chloroplast genome sequence of medicinal plant, *Artemisia argyi*. *Mitochondrial DNA B Resour*. 2016 Mar. 1(1): 257–8. <https://doi.org/10.1080/23802359.2016.1159926> PMID: 33473468.
33. Chen J, Hao Z, Xu H, Yang L, Liu G, Sheng Y et al. The complete chloroplast genome sequence of the relict woody plant *Metasequoia glyptostroboides* Hu et Cheng. *Front. Plant Sci*. 2019 Jun. 6: 447. <https://doi.org/10.3389/fpls.2015.00447> PMID: 26136762.
34. Sumner LW, Amberg A, Barrett D, Beale MH, Beger R, Daykin CA et al. Proposed minimum reporting standards for chemical analysis Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI). *Metabolomics*. 2007 Sep. 3(3): 211–21. <https://doi.org/10.1007/s11306-007-0082-2> PMID: 24039616.
35. Wang M, Carver JJ, Phelan VV, Sanchez LM, Garg N, Peng Y et al. Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nat. Biotechnol*. 2016 Aug. 34(8): 828–837. <https://doi.org/10.1038/nbt.3597> PMID: 27504778.
36. Ferrante C, Zengin G, Menghini L, Diuzheva A, Jekó J, Cziáky Z et al. Qualitative fingerprint analysis and multidirectional assessment of different crude extracts and essential oil from wild *Artemisia santonicum* L. *Processes (Basel)*. 2019 Aug. 7(8): 522. <https://doi.org/10.3390/pr7080522>.