

Development of Molecular Markers to Discriminate the Species and Origin of *Gynostemma pentaphyllum* (Cucurbitaceae) Based on the Plastid Genome Data

Daewon Kang¹, Jong-Soo Kang¹, Woohyeon Cho¹, Jee Young Park¹, Tae-Young Kim², Tae-Jin Yang^{1*}

¹ Department of Agriculture, Forestry and Bioresources, Plant Genomics & Breeding Institute, Research Institute of Agriculture and Life Sciences, College of Agriculture & Life Sciences, Seoul National University, 1 Gwanak-ro, Gwanak-gu, Seoul 08826, Korea

² Technology Development Center, BTC Corporation, Ansan 15588, Korea

Abstract

Background : *Gynostemma pentaphyllum* is considered a valuable plant in East Asia, including South Korea, China, and Japan, where it is used in traditional medicine and consumed as tea. The morphology of *Gynostemma* species is not easy to distinguish one another due to high morphological similarity. The only way to differentiate them by morphology is by comparison of their flower structure and fruits. With the steady increase in demand for *G. pentaphyllum* as a herbal remedy and dietary supplement, there is a need for species identification within *Gynostemma* and for determining its origin. Thus, in this study, we aimed to develop molecular markers for identifying *G. pentaphyllum* and its origin through analyzing plastid genome sequences.

Methods and Results : A total of seven *G. pentaphyllum* accessions, including three from South Korea and four from China, were collected and sequenced for generating NGS data. In addition, 22 *Gynostemma* plastid genome sequences were downloaded from the NCBI database. The plastid genomes of seven newly sequenced accessions were assembled using the dnaLCW method (de novo assembly of Low Coverage Whole-genome sequencing). The assembled plastid genomes were used for finding sequence variation and phylogenetic analysis. *G. pentaphyllum* was clearly distinct from other *Gynostemma* species in sequence comparison and phylogenetic analysis. Among the *G. pentaphyllum* accessions, while the genetic diversity within the plastid genomes of *G. pentaphyllum* from each country (South Korea and China) is low, a relative difference was observed between the Korean and Chinese *G. pentaphyllum* accessions. Consequently, we developed three molecular markers based on sequence variations in their plastid genomes to identify *G. pentaphyllum* from other *Gynostemma* and *G. pentaphyllum* from other *Gynostemma* species and determine its origin.

Conclusion : Our study reveals notable genetic differentiation in the plastid genomes both between *G. pentaphyllum* and other *Gynostemma* species, and within *G. pentaphyllum* accessions from Korea and China. We believe that these newly developed markers will enhance the accuracy of *G. pentaphyllum* identification and its origin determination.

Materials and Method

Materials



Figure 1. Morphology of *G. pentaphyllum*.

A specimen of *G. pentaphyllum* was collected from Chungcheongnam-do and deposited at the National Institute of Biological Resources (NIBR).

Table 1. Information of 30 accessions used for marker development and validation.

No.	ID	Collection site	Condition	Source
1	KOR1	Geochang, South Korea	Farm collection	BTC
2	KOR2	Geochang, South Korea	Farm collection	BTC
3	KOR3	Ulleungdo, South Korea	Farm collection	BTC
4	KOR4	Ulleungdo, South Korea	Farm collection	BTC
5	KOR5	Guryongpo-eup, Pohang, South Korea	Wild collection	NIBR
6	KOR6	Sanghyo-dong, Seogwipo, South Korea	Wild collection	NIBR
7	KOR7	North Gyeongyeolbiyeoldo, Taean, South Korea	Wild collection	NIBR
8	KOR8	Jeongbang-dong, Seogwipo, South Korea	Wild collection	NIBR
9	KOR9	Gwangnyeong-ri, Aewol-eup, Jeju, South Korea	Wild collection	NIBR
10	KOR10	Odo, Boryeong, South Korea	Wild collection	NIBR
11	FJ1	Fujian, China	Farm collection	BTC
12	FJ2	Fujian, China	Farm collection	BTC
13	FJ3	Fujian, China	Farm collection	BTC
14	FJ4	Fujian, China	Farm collection	BTC
15	FJ5	Fujian, China	Farm collection	BTC
16	FJ6	Fujian, China	Farm collection	BTC
17	FJ7	Fujian, China	Farm collection	BTC
18	FJ8	Fujian, China	Farm collection	BTC
19	HU	Hunan, China	Farm collection	BTC
20	GX1	Jinxiu, Guangxi, China	Farm collection	BTC
21	GX2	Jinxiu, Guangxi, China	Farm collection	BTC
22	GX3	Jinxiu, Guangxi, China	Farm collection	BTC
23	GX4	Jinxiu, Guangxi, China	Farm collection	BTC
24	GX5	Beihai, Guangxi, China	Farm collection	BTC
25	GX6	Chongzuo, Guangxi, China	Farm collection	BTC
26	GX7	Chongzuo, Guangxi, China	Farm collection	BTC
27	GX8	Chongzuo, Guangxi, China	Farm collection	BTC
28	MF152730	Pingli, Shaanxi, China		NCBI
29	KX014626	Pingli, Shaanxi, China		NCBI
30	KX852298	Xi'an, Shaanxi, China		NCBI

Table 2. Information of all sequence data used for phylogenetic analysis.

No.	Scientific name	accession name	Source
1	<i>G. laxum</i>	ON872374	NCBI
2	<i>G. pubescens</i>	MF152732	NCBI
3	<i>G. simplicifolium</i>	ON872376	NCBI
4	<i>G. burmanicum</i>	MN542403	NCBI
5	<i>G. burmanicum</i>	ON872370	NCBI
6	<i>G. burmanicum</i>	MF152731	NCBI
7	<i>G. longipes</i>	MF152730	NCBI
8	<i>G. pentaphyllum</i>	KX014626	NCBI
9	<i>G. pentaphyllum</i>	KX852298	NCBI
10	<i>G. pentaphyllum</i>	KOR1	This study
11	<i>G. pentaphyllum</i>	KOR2	This study
12	<i>G. pentaphyllum</i>	KOR3	This study
13	<i>G. pentaphyllum</i>	FJ1	This study
14	<i>G. pentaphyllum</i>	GX1	This study
15	<i>G. pentaphyllum</i>	GX2	This study
16	<i>G. compressum</i>	GX6	This study
17	<i>G. compressum</i>	KY817143	NCBI
18	<i>G. pentaphyllum</i>	MN583314	NCBI
19	<i>G. compressum</i>	ON872372	NCBI
20	<i>G. longipes</i>	ON872375	NCBI
21	<i>G. caulopterum</i>	ON872371	NCBI
22	<i>G. guangxiense</i>	ON186498	NCBI
23	<i>G. caulopterum</i>	MF136487	NCBI
24	<i>G. guangxiense</i>	ON872373	NCBI
25	<i>G. pentagynum</i>	KY670737	NCBI
26	<i>G. laxiflorum</i>	MF136486	NCBI
27	<i>G. yixingense</i>	MT028489	NCBI
28	<i>G. cardiospermum</i>	KX852299	NCBI
29	<i>G. microsporum</i>	MZ286581	NCBI
30	<i>Hemsleya lijianensis</i>	MG733988	NCBI
31	<i>Hemsleya zhejiangensis</i>	MN414239	NCBI
32	<i>Gomphogyne cissiformis</i> var. <i>cissiformis</i>	MH256801	NCBI
33	<i>Gomphogyne cissiformis</i> var. <i>vilosa</i>	MF784515	NCBI

- Among the 33 plastid genome sequences, 26 were previously reported plastid genome sequences of *Gynostemma*, *Hemsleya* and *Gomphogyne* (3 genera) downloaded from NCBI GenBank, and 7 were newly assembled in this study

Method

DNA extraction

- Geneall SV Plant Mini kit (GeneAll Biotechnology Co., Songpa-gu, Seoul)

Genome sequencing & assembly

- Illumina Miseq (Illumina Inc., San Diego, CA)
- dnaLCW method (using CLC assembly cell)

Gene annotation

- GeSeq (<https://chlorobox.mpimp.golm.mpg.de/geseq>)
- tRNAscan-SE 2.0 (Chan et al., 2021)
- Artemis (Carver et al., 2012)

Phylogenetic analysis & visualization

- MAFFT version 7 for alignment (Katoh et al., 2013)
- IQ-TREE2 (Minh et al., 2020)
- iTOL v5 (<https://itol.embl.de/>)

Comparative analysis with plastid genome

Table 3. Summary of seven newly assembled plastid genomes in this study.

Scientific name	ID	Plastid genome Length (bp)	GC content (%)
<i>G. pentaphyllum</i>	KOR1	157,699	36.9
<i>G. pentaphyllum</i>	KOR2	157,693	36.9
<i>G. pentaphyllum</i>	KOR3	157,726	36.9
<i>G. pentaphyllum</i>	FJ1	157,709	36.9
<i>G. pentaphyllum</i>	GX1	157,686	36.9
<i>G. pentaphyllum</i>	GX2	157,708	36.9
<i>G. compressum</i>	GX6	157,417	36.9

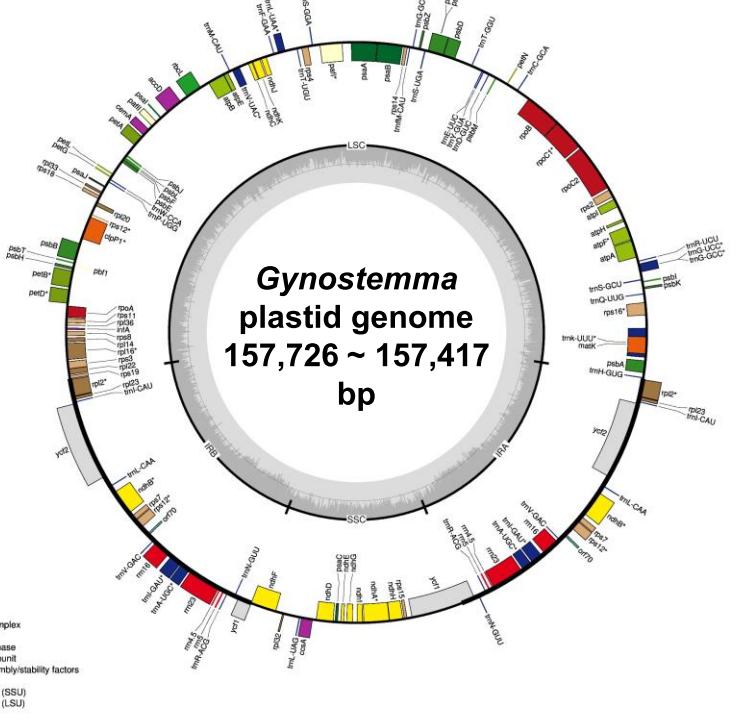


Figure 2. Circular map of *Gynostemma* plastid genomes.

Table 4. Variation information including SNPs and InDels among ten *Gynostemma* plastid genomes.

SNP/InDel	KOR1	KOR2	KOR3	GX1	GX2	FJ1	MF152730	KX014626	KX852298	GX6
KOR1	-	4	8	73	77	77	72	79	72	263
KOR2	5	-	10	72	75	75	72	78	72	264
KOR3	6	5	-	69	74	74	69	75	69	263
GX1	124	123	122	-	16	17	18	22	22	264
GX2	121	120	119	11	-	1	15	22	21	265
FJ1	121	120	119	11	0	-	16	23	21	265
MF152730	127	126	125	18	13	13	-	21	20	262
KX014626	134	133	132	22	21	21	26	-	26	267
KX852298	212	211	212	112	107	107	112	118	-	265
GX6	1,104	1,103	1,103	1,115	1,114	1,114	1,118	1,122	1,196	-

Phylogenetic analysis of plastid genome

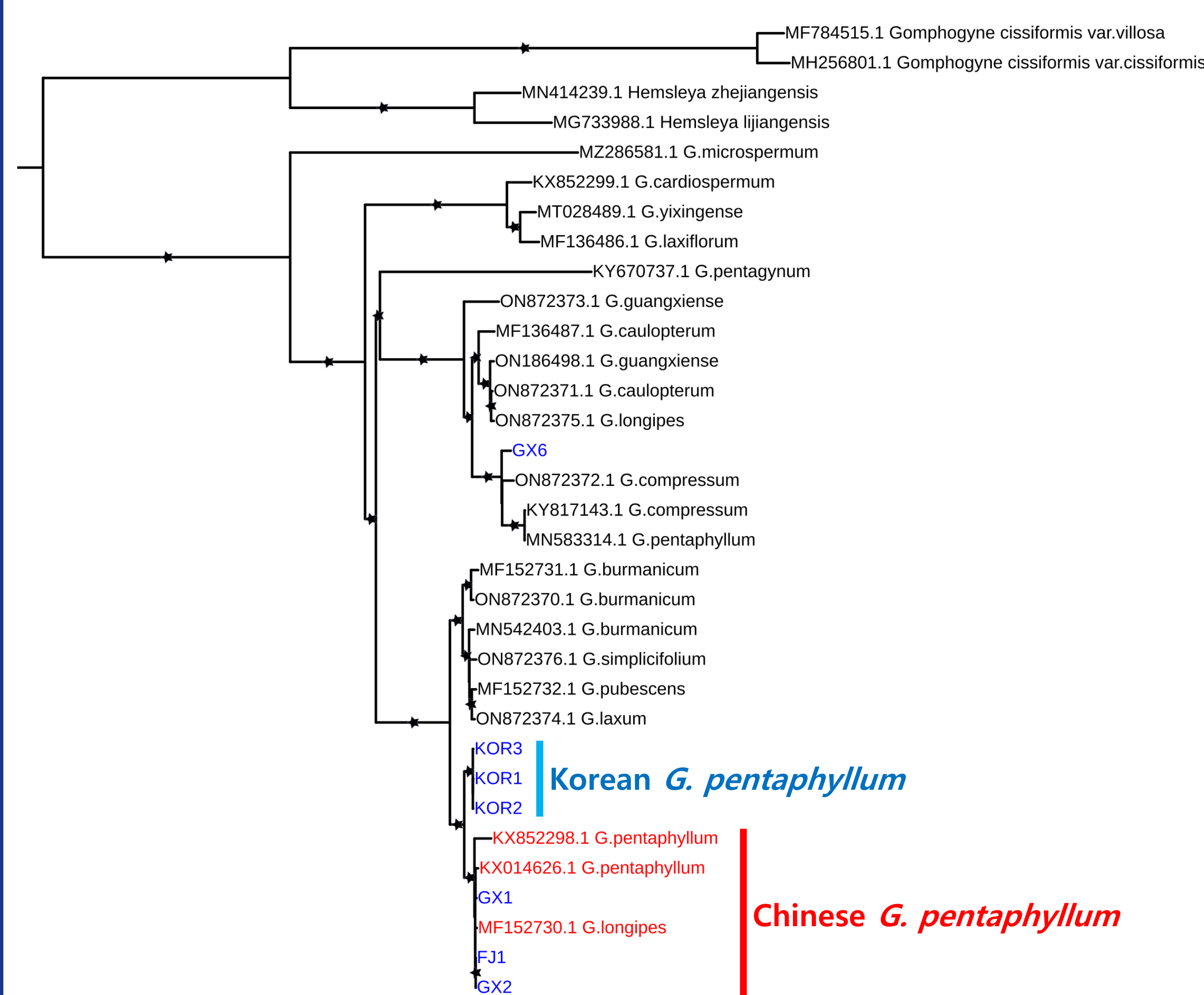
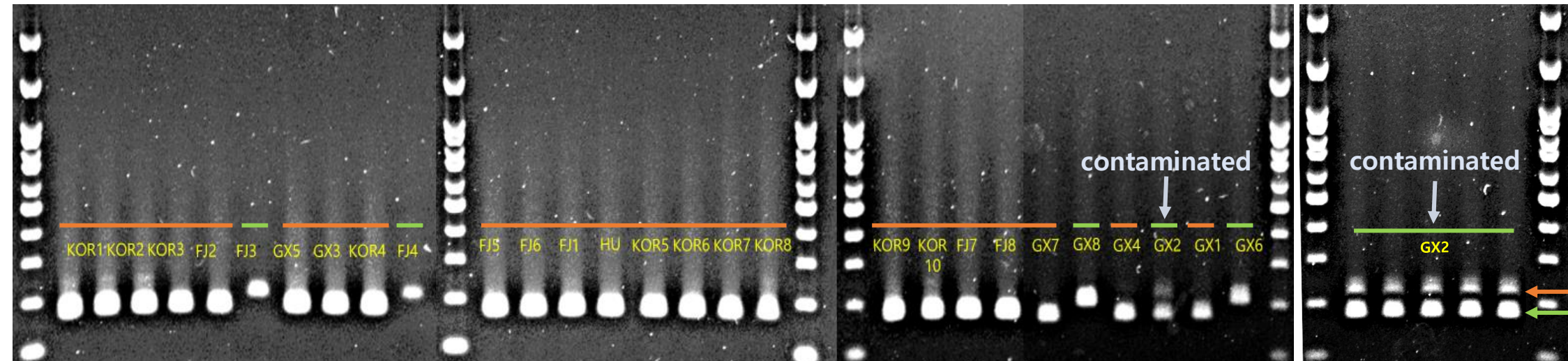


Figure 3. Phylogenetic tree based on plastid genome sequences of the genus *Gynostemma*.

Phylogenetic tree was constructed using 33 plastid genomes. Two *Hemsleya* and two *Gomphogyne* plastid genomes were included as outgroups. Asterisks indicate bootstrap values greater than 90%. Plastid genomes of *G. pentaphyllum* from previous studies are indicated in red, while the newly assembled *Gynostemma* plastid genomes in this study are indicated in blue.

Molecular marker development and application

A) Gyno_sp_1



B) Gyno_ori_1



C) Gyno_ori_2

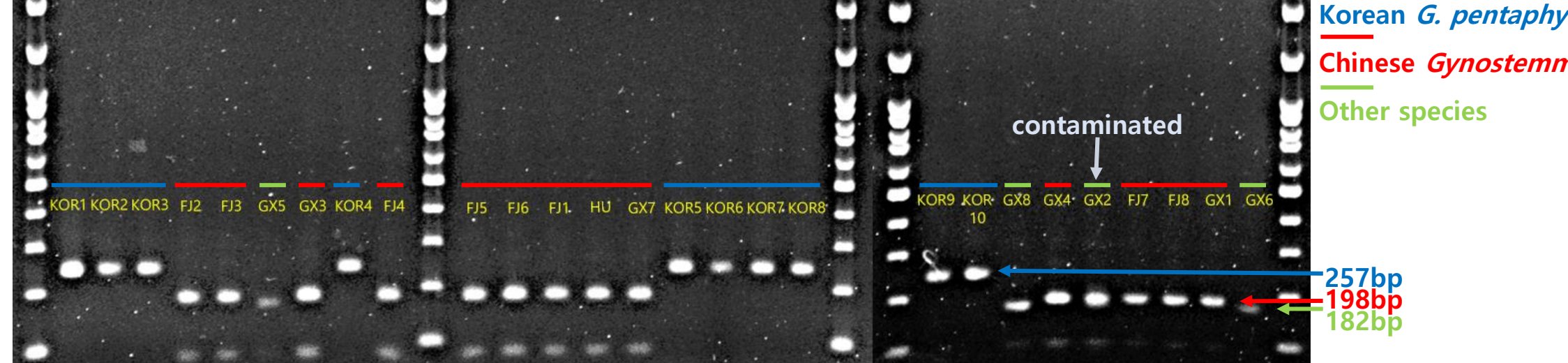


Table 5. Brief information of the developed markers and expected results.

Primer ID	Location	Product Size (bp)
Gyno_sp_1	<i>trnT-trnL</i>	Korean & Chinese <i>Gp</i> 191
		Other species 228
Gyno_ori_1	<i>trnT-trnL</i>	Korean <i>Gp</i> 235
		Other species & Chinese <i>Gp</i> -
Gyno_ori_2	<i>ndhA</i> intron	Korean <i>Gp</i> 257
		Other species & Chinese <i>Gp</i> 198
		Other species 182

Gp: *Gynostemma pentaphyllum*

Figure 4. Validation of three molecular markers on 27 *Gynostemma* accessions.

Gel electrophoresis results of PCR products derived from three markers were presented. (A) *G. pentaphyllum* specific InDel marker. GX2 with double bands indicates contamination by other *Gynostemma* species. (B) Korean *G. pentaphyllum* specific dominant marker. (C) CAPs marker, which can identify Korean *G. pentaphyllum* and other species including Chinese *G. pentaphyllum*.

Conclusion

Our study demonstrates that *G. pentaphyllum* is clearly distinct from other *Gynostemma* species based on plastid genome data and that Korean and Chinese *G. pentaphyllum* accessions have a significant difference. Based on these results, a total of three molecular markers were developed to identify *G. pentaphyllum* from other *Gynostemma* and determine the origin of *G. pentaphyllum*. By applying these markers, it becomes possible to verify whether it is *G. pentaphyllum* or not and whether it is domestically produced or imported. Moreover, if other species are mixed in, it can be detected, similar to the case of GX2 in this study.