The complete chloroplast genomes of three *polygala* species and indel marker development for discrimination of authentic Polygalae Radix (Roots of *P. tenuifolia*)

P04-001

Sumin Jeong[1)†], Jong Won Han[2)†], Yeseul Kim[1)], , Eunjeong Bak[3)], Kyung Ho Ma[4)], Jeong Hoon Lee[4)], Jin Tae Jung[3)] and Inkyu Park*[1)]
1) Department of Biology and Chemistry, Changwon National University, Changwon 51140, Korea.
2) Research management Division, Rural Development Administration, Jeonju 54875, Korea
3) Department of Horticulture, Kongju National University,Yesan 32439, Korea.
4)Department of Herbal Crop Research, National Institute of Horticultural & Herbal Science, Rural Development Administration, Eumseong 27709, Korea

ABSTRACT

The genus *Polygala* L., belonging to the Polygalaceae family, is a large genus containing approximately 700 species distributed throughout the world. The dried roots of *P. tenuifolia*, which have been designated Polygalae Radix, is used as traditional Korean herbal medicine. Despite being a valuable medicinal resource, genomic information of the genus is still limited. In this study, we sequenced three *Polygala*, *P. tenuifolia*, *P. japonica*, and *P. sibirica*, which distributed in Korea. The chloroplast genome (cp) size ranged from 165,246 to 165,420 bp, with a typical quaternary structure, and the three cp genomes GC content was 36.9%. A total of 111 genes were identified, including 79 protein coding regions, 28 tRNA genes, and four rRNA genes. Nucleotide diversity (Pi) analysis results revealed hotspots such as *trnS-trnR*, *rpl32-trnL*, *ccsA-ndhD*, and *ndhG-ndhI*. We developed nine molecular markers to facilitate species identification, and as a result of commercial product verification, 11 products were confirmed to be authentic and were verified to be *P. tenuifolia* through breeding line verification. Phylogenetic analysis using the whole cp genome data was supported by strong bootstrap values and posterior probabilities, showing that *P. tenuifolia* is clearly clustered and has a monophyletic sister group with *P. sibirica* and *P. japonica*. In this study, the cp genomes of three *Polygala* species were determined, and phylogenetic tree analysis and marker development were performed. Indel markers were developed for the discrimination of the important herbal medicine species *P. tenuifolia*. The cp genomes and analyses in this study provide valuable information on further research as a medical resource and the accurate identification of *P. tenuifolia*.

Material and Methods

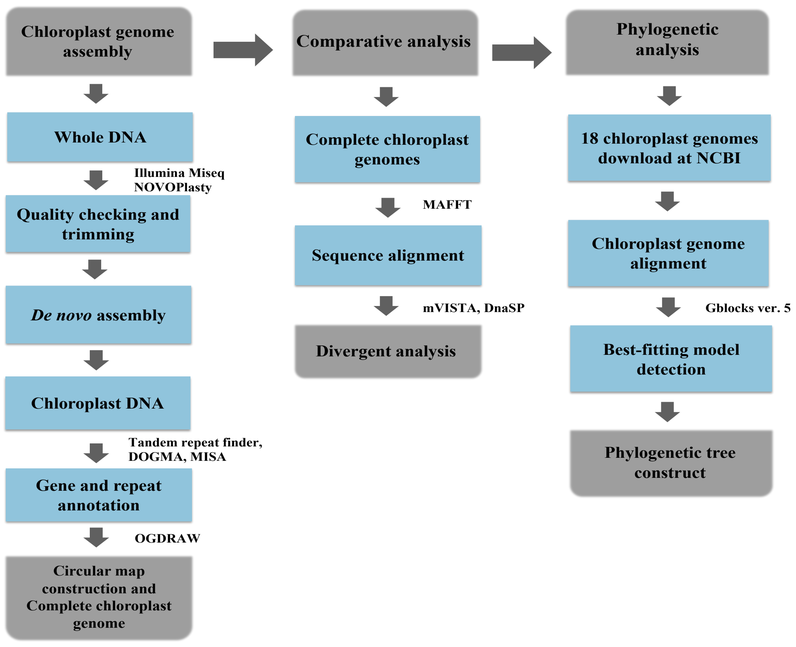

Figure NGS read assembly and chloroplast genome analysis Steps and Tools.

Results

Species	*P. tenuifolia*	*P. japonica*	*P. sibirica*
Total cp genome size (bp)	165,420	165,246	165,247
Large single copy (LSC) region (bp)	83,696	83,642	83,668
Inverted repeat (IR) region (bp)	36,840	36,750	36,742
Small single copy (SSC) region (bp)	8,044	8,084	8,095
Total number of genes (unique)	111	111	111
Protein-coding gene (unique)	79	79	79
tRNA (unique)	28	28	28
rRNA (unique)	4	4	4
GC content (%)	36.70%	36.70%	36.70%
LSC (%)	34.70%	34.80%	34.80%
IR (%)	39.70%	39.70%	39.70%
SSC (%)	29.40%	29.50%	29.50%

Table Characteristics of three *Polygala* chloroplast genomes.

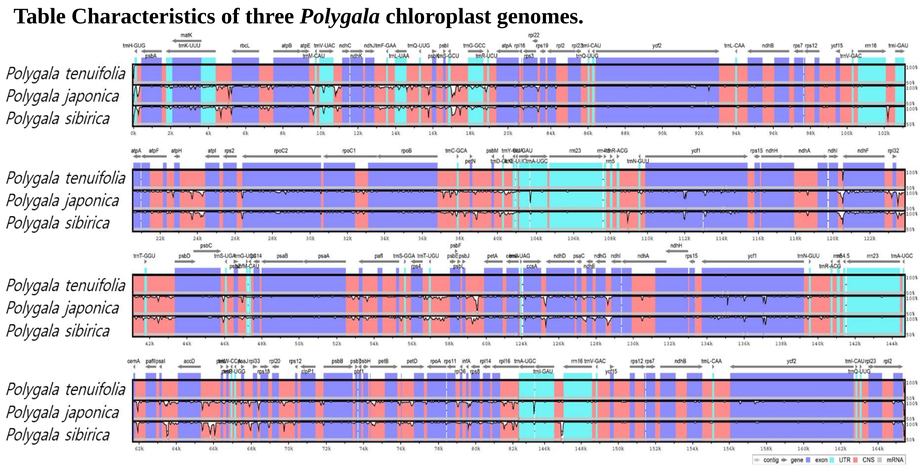

Figure Comparison of *Polygala* complete chloroplast genomes using mVISTA. The area where sequence variation among the three *Polygala* are present in white. The gray arrow above the alignment indicates the direction of the gene. The blue bars represent exons, and the pink bars represent non-coding sequences.

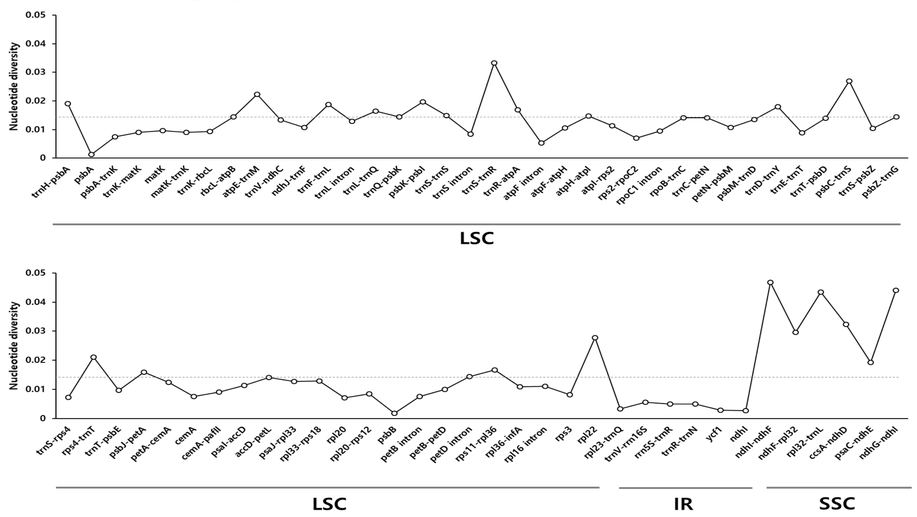

Figure Comparison of nucleotide diversity (Pi) values in three species of *Polygala*. Areas with a Pi value of 0 were excluded.

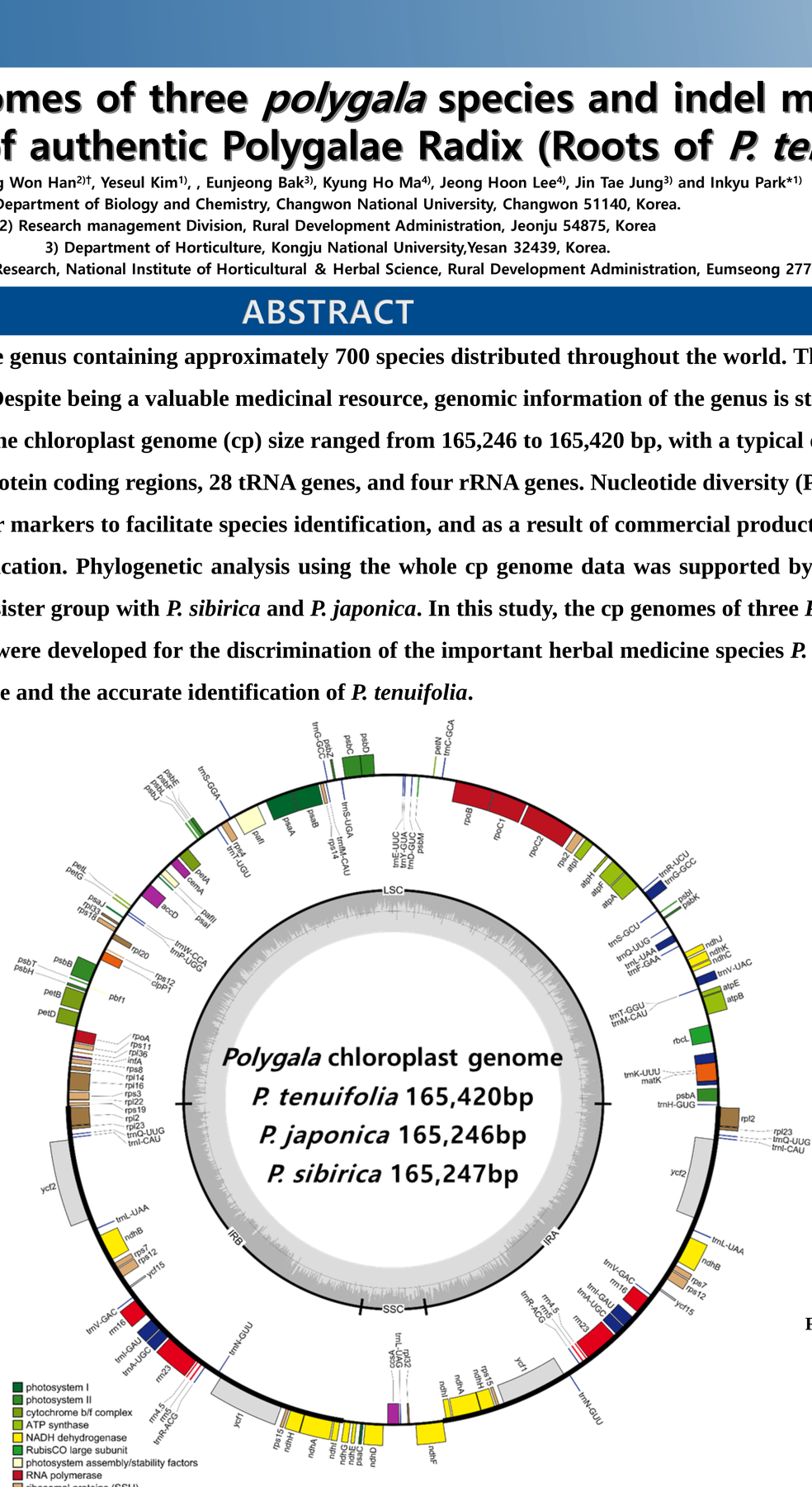

Figure Gene map of *Polygala* chloroplast genomes. The genes that are drawn inside the circle are transcribed clockwise, and outside the circle are transcribed counterclockwise. Gene colors indicate that they belong to different functional groups. Thick lines represent the inverted repeats (IRa and IRb), which are separated by the large single-copy (LSC) region and small single-copy (SSC) region. Gray bars in the inner circle indicates GC content.

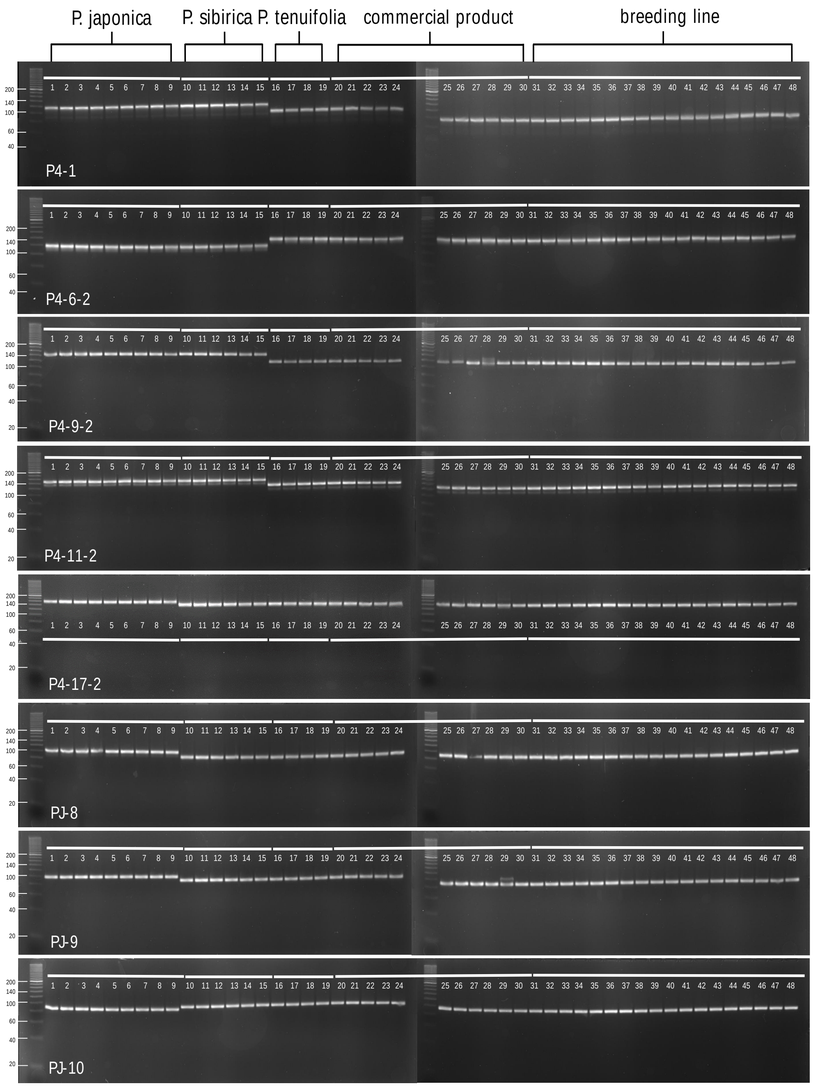

Figure Gel images of indel markers for *Polygala*. Eight primers were tested by control samples (1-9: *P. japonica*; 10-15: *P. sibirica*; 16-19: *P. tenuifolia*), commercial products (20-30), and breeding lines (21-48).

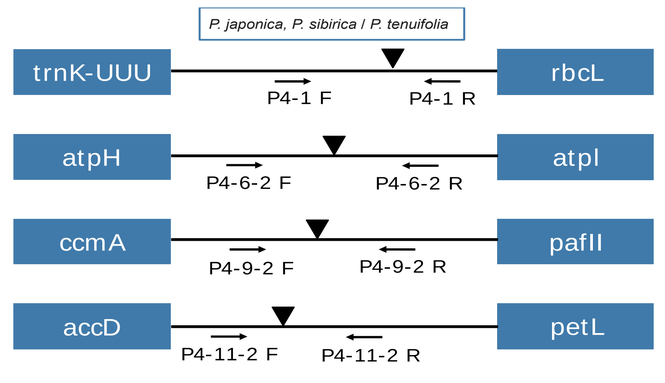

Figure Schematic representation of indel markers for *Polygala* species.

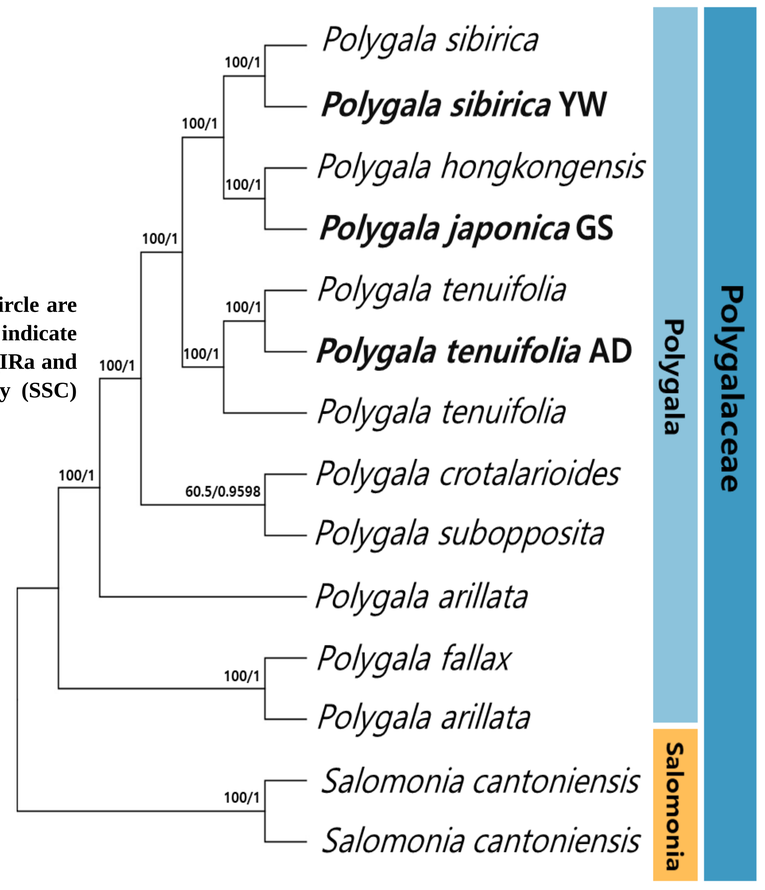

Figure A phylogenetic tree of *Polygala* with sister genus *Sanguisorba* representing the maximum likelihood (ML) bootstrap and Bayesian posterior (BI) probability based on whole plastomes. The number of branches above or below represents the ML bootstrap and BI probability support values.

Conclusions

In this study, we determined *Polygala* cp genomes and performed phylogenetic analysis and molecular marker development. The cp genomes of *Polygala* were highly conserved in terms of their gene contents, gene orientations, GC contents, and local variations. Most of the differences were detected in the non-coding area. These hotspots were used to develop new indel markers for efficient and rapid species identification. The Indel marker successfully identified an important herbal species, *P. tenuifolia*. The cp genomes and analyses in this study provide valuable information on further research as a medical resource and on the accurate identification of *P. tenuifolia*.

This work was carried out with the support of "Cooperative Research Program for Agriculture Science and Technology Development (Project No. RS-2022-RD010267)" Rural Development Administration, Republic of Korea.