



The Complete Chloroplast Genomes of Six *Ipomoea* Species and Indel Marker Development for the Discrimination of Authentic Pharbitidis Semen (Seeds of *I. nil* or *I. purpurea*)

Inkyu Park¹, Sungyu Yang¹, Wook J. Kim¹, Pureum Noh¹, Hyun O. Lee² and Byeong C. Moon^{1*}

¹ K-herb Research Center, Korea Institute of Oriental Medicine, Daejeon, South Korea, ² Phyzen Genomics Institute, Seongnam, South Korea

OPEN ACCESS

Edited by:

Rogerio Margis,
Universidade Federal do Rio Grande
do Sul (UFRGS), Brazil

Reviewed by:

Andreia Carina Turchetto Zolet,
Federal University of Rio Grande do
Sul, Brazil

Peter Poczai,
University of Helsinki, Finland

*Correspondence:

Byeong C. Moon
bcmoon@kiom.re.kr

Specialty section:

This article was submitted to
Evolutionary and Population Genetics,
a section of the journal
Frontiers in Plant Science

Received: 28 March 2018

Accepted: 15 June 2018

Published: 05 July 2018

Citation:

Park I, Yang S, Kim WJ, Noh P,
Lee HO and Moon BC (2018)
The Complete Chloroplast Genomes
of Six *Ipomoea* Species and Indel
Marker Development
for the Discrimination of Authentic
Pharbitidis Semen (Seeds of *I. nil* or
I. purpurea). *Front. Plant Sci.* 9:965.
doi: 10.3389/fpls.2018.00965

Ipomoea L. is the largest genus within the Convolvulaceae and contains 600–700 species. *Ipomoea* species (morning glories) are economically valuable as horticultural species and scientifically valuable as ecological model plants to investigate mating systems, molecular evolution, and both plant–herbivore and plant–parasite interactions. Furthermore, the dried seeds of *I. nil* or *I. purpurea* are used in Korean traditional herbal medicines. In this study, chloroplast (cp) genomes were sequenced from six *Ipomoea* species, namely, *I. nil* and *I. purpurea* and, for the first time, *I. triloba*, *I. lacunosa*, *I. hederacea*, and *I. hederacea* var. *integriscula*. The cp genomes were 161,354–161,750 bp in length and exhibited conserved quadripartite structures. In total, 112 genes were identified, including 78 protein-coding regions, 30 transfer RNA genes, and 4 ribosomal RNA genes. The gene order, content, and orientation of the six *Ipomoea* cp genomes were highly conserved and were consistent with the general structure of angiosperm cp genomes. Comparison of the six *Ipomoea* cp genomes revealed locally divergent regions, mainly within intergenic spacer regions (*petN-psbM*, *trnI-CAU-ycf2*, *ndhH-ndhF*, *psbC-trnS*, and *ccsA-ndhD*). In addition, the protein-coding genes *accD*, *cemA*, and *ycf2* exhibited high sequence variability and were under positive selection ($K_a/K_s > 1$), indicating adaptive evolution to the environment within the *Ipomoea* genus. Phylogenetic analysis of the six *Ipomoea* species revealed that these species clustered according to the APG IV system. In particular, *I. nil* and *I. hederacea* had monophyletic positions, with *I. purpurea* as a sister. *I. triloba* and *I. lacunosa* in the section *Batatas* and *I. hederacea* and *I. hederacea* var. *integriscula* in the section *Quamoclit* were supported in this study with strong bootstrap values and posterior probabilities. We uncovered high-resolution phylogenetic relationships between *Ipomoeae*. Finally, indel markers (IPOTY and IPOYCF) were developed for the discrimination of the important

herbal medicine species *I. nil* and *I. purpurea*. The cp genomes and analyses in this study provide useful information for taxonomic, phylogenetic, and evolutionary analysis of the *Ipomoea* genome, and the indel markers will be useful for authentication of herbal medicines.

Keywords: Ipomoeae, plastid genome, divergent region, phylogenetic relationship, indel marker

INTRODUCTION

Chloroplasts (cp) are among the most important organelles in plants, having important roles in photosynthesis and carbon fixation as well as in the biosynthesis of starch, fatty acids, amino acids, and pigments (Jansen and Ruhlmann, 2012; Daniell et al., 2016). Cp genomes in higher plants are 120–180 kb and, in general, exhibit a quadripartite structure consisting of two single-copy regions, namely, the large single-copy (LSC) and small single-copy (SSC), as well as two copies of a larger inverted repeat (IR) region. Angiosperm cp genomes generally contain 80 protein-coding genes, 4 ribosomal RNA (rRNA) genes, and 30 transfer RNA (tRNA) genes (Wicke et al., 2011). While the majority of cp genomes exhibit highly conserved structures, some reveal structural variations, IR loss, and gene loss as a result of adaptation to their environments (Delannoy et al., 2011; Wicke et al., 2013). Next-generation technologies have allowed the rapid sequencing of many cp genomes in recent years. These abundant cp genomes have facilitated the verification of evolutionary relationships and allowed detailed phylogenetic classifications to be conducted at group, family, and even genus level in Plantae (Jansen et al., 2007; Parks et al., 2009). Furthermore, cp genomes can be used for species identification through the use of DNA barcodes and molecular markers that allow morphologically similar species to be distinguished (Kim et al., 2015; Park et al., 2017a,b). Thus, cp genomes can be used for practical applications such as species identification as well as for fundamental research into biological processes and evolutionary relationships.

Ipomoea is the largest genus in the Convolvulaceae family, with 600–700 species (Austin and Huáman, 1996; Wilkin, 1999). *Ipomoea* species are widely distributed across tropical, subtropical, and some temperate regions worldwide (Austin and Huáman, 1996; Wilkin, 1999). *I. nil*, *I. purpurea*, *I. tricolor*, and *I. batatas* are particularly well-known *Ipomoea* species. *I. nil* exhibited spontaneous mutations related to floricultural traits. These mutants have been exploited as ornamental plants in horticulture (Hoshino et al., 2016). *Ipomoea* is an emerging model system for ecological genomics studies (Baucom et al., 2011; Eserman et al., 2014). Ecological studies of *Ipomoea* have answered many diverse questions about the *Ipomoea* mating system, the evolution of floral color pathways, and both plant–herbivore and plant–parasite interactions (Baucom et al., 2011). Visitation by natural pollinators and the selfing rate in various *Ipomoea* species vary in proportion to the number of offspring derived from self-fertilization. These findings are exemplified by extreme differences in floral color in *Ipomoea*, which ranges from white or yellow to red or purple (Ennos and Clegg, 1983; Epperson and Clegg, 1992; Baucom et al., 2011). These flower colors (which are associated with anthocyanin pigments)

have evolved *via* parallel evolution due to various factors, such as enzyme-coding genes (F3'H) or regulatory modifications (Des Marais and Rausher, 2008; Streisfeld and Rausher, 2009). Furthermore, *I. purpurea* and *I. hederacea* are model plants used to study plant–herbivore interactions based on ecological evolution (Tiffin and Rausher, 1999; Baucom et al., 2011). Several studies show that insects have affected natural selection for plant resistance in *Ipomoea* and the tradeoff between resistance and tolerance in plant defense responses. In addition, studies of plant–herbivore interactions point to the coevolution between *I. purpurea* and *I. hederacea* and their competitors (Rausher and Fry, 1993; Simonsen and Stinchcombe, 2007). Studies examining the evolution of *Ipomoea* in response to plant pathogens show that quantitative resistance to *Colletotrichum dematium* is genetically correlated to quantitative resistance to an insect herbivore and that an oomycete exhibits host specialization in *Ipomoea* (Simms and Rausher, 1993; Sato et al., 2009). Therefore, *Ipomoea* species represent highly important resources that have contributed strongly to ecological studies.

In Korean traditional medicine, the dried seeds of *I. nil* or *I. purpurea* are an important herbal medicine, namely, Pharbitidis Semen, which is used to eliminate toxins or heat, as a diuretic, and as a treatment for constipation relief treatment [Korea Institute of Oriental Medicine (KIOM), 2016]. Pharbitidis Semen is designated as a medicine in Korea, and is regulated by the Ministry of Food and Drug Safety due to its pharmaceutical activity and potential toxicity (Korean Food Standard Codex, 2010). Only seeds of *I. nil* or *I. purpurea* are considered to be authentic Pharbitidis Semen. In general, *Ipomoea* seeds are trigonous and are brown to dark-brown in color (McDonald, 1995), and seeds from different *Ipomoea* species are morphologically similar and difficult to distinguish with the unaided eye. As a result, Pharbitidis Semen in Korean and Chinese herbal markets often contains a mixture of seeds from *I. nil* and *I. purpurea* and seeds from other *Ipomoea* species. Indiscriminate use of these adulterated Pharbitidis Semen preparations could cause unforeseen side-effects and threaten its use as a safe and reliable medication. Methods are therefore needed to distinguish good quality Pharbitidis Semen preparations from adulterated preparations.

Molecular tools can be used for accurate species identification and authentication of herbal medicine. In particular, the universal DNA barcode markers ITS, *matK*, and *rbcL* are widely used for species classification and phylogenetic analysis in Plantae (Semagn et al., 2006; Sucher and Carles, 2008; Chen et al., 2010; Hollingsworth et al., 2011). These barcode markers offer rapid and accurate species identification from short DNA sequences. However, some plants, particularly closely related species, cannot be readily distinguished using these markers. The cp genome

has emerged as an alternative to DNA barcoding markers for species identification and phylogenetic studies. Comparison of cp genomes highlighted several variable regions that could be used for the development of markers to allow species discrimination (Kim et al., 2015; Park et al., 2017a). While the cp genome was generally more highly conserved than the nuclear genome, abundant genetic variations such as insertion/deletions (indels) and single nucleotide polymorphisms (SNPs) were identified between species. Several studies developed cp markers for identification of closely related species, including indel and SNP markers for *Panax ginseng* subspecies (Kim et al., 2015), and indel tandem repeat copy number variation markers for *Fagopyrum tataricum* and *F. esculentum* (Cho et al., 2015). In another example, sequence characterized amplified region markers were developed to resolve *Aconitum* species. Two indel markers derived from large variable regions were used to distinguish three *Aconitum* species, *A. pseudolaeve*, *A. longecassidatum*, and *A. barbatum*, and a small species-specific 6 bp insertion was used to distinguish *A. coreanum* (Park et al., 2017a,b). *Chenopodium quinoa* and *C. album* were distinguished using indel tandem repeat copy number variation markers (Hong et al., 2017). These examples illustrate the utility of the cp genome for plant species identification and for the authentication and identification of herbal medicines.

Previous phylogenetic analysis of the genus *Ipomoea* was unclear, with unresolved monophyly at the subgenera level (Manos et al., 2001; Stefanovie et al., 2002). Previously, molecular phylogenetic relationships within the Convolvulaceae were evaluated using ITS and four cp loci (Stefanovie et al., 2002; Miller et al., 2004), but this analysis identified only monophyletic or weak relationships in tribe Ipomoeae. A separate analysis of four cp loci divided tribe Ipomoeae into two clades, Astripomoeinae and Argyreiinae, but morphological features were not considered (Stefanovie et al., 2003). Recently, Eserman et al. (2014), described high-resolution phylogenetic relationships in tribe Ipomoeae and the Astripomoeinae and Argyreiinae clades and identified similar divergence times (23–26 MYA) based on whole cp genomes. Analysis of 32 cp genomes from magnoliids, monocots, and eudicots verified phylogenetic relationships for sweet potato (*I. batatas*) (Yan et al., 2015). Analysis of the completed nuclear genome identified a whole-genome duplication event in *I. nil* and showed divergence from Solanaceae at 75.25 MYA (Hoshino et al., 2016). Examination of *Ipomoea*, which contains hundreds of species, identified a range of useful genomic information, but this was not sufficient for high-resolution determination of phylogenetic relationships in Ipomoeae. Further research is needed to understand the evolutionary relationships within tribe Ipomoeae as well as the *Ipomoea* genus.

Here, samples of herbal medicine species *I. nil* and *I. purpurea* and four closely related *Ipomoea* species with similar seed structures were collected and their cp genomes were compared. This study aimed to (1) characterize six *Ipomoea* cp genomes and identify genetically variable regions by comparison of their global structures, (2) develop novel molecular markers for use in authentication of herbal medicine species, and (3) understand evolutionary relationships within tribe Ipomoeae through

enhanced phylogenetic studies in conjunction with previously reported cp genomes.

MATERIALS AND METHODS

Plant Materials

Fresh leaves of six *Ipomoea* species were collected from native habitats in Korea and used for cp genome sequencing. *I. nil*, *I. purpurea*, *I. hederacea*, *I. hederacea* var. *integriuscula*, *I. lacunosa*, and *I. triloba* were assigned identification numbers, and specimens were registered in the Korean Herbarium of Standard Herbal Resources (Index Herbariorum code KIO) at the KIO. The plant samples used for cp genome analysis and indel validation in this study are listed in Supplementary Table S1.

Genome Sequencing and Assembly

DNA was extracted using a DNeasy Plant Maxi Kit (Qiagen, Valencia, CA, United States) according to the manufacturer's instructions. Illumina short-insert paired-end sequencing libraries were constructed and generated using the NextSeq platform (Illumina, San Diego, CA, United States). *De novo* assembly was used to construct cp genomes from low-coverage whole-genome sequences. Trimmed paired-end reads (Phred scores ≥ 20) were assembled using CLC genome assembler (ver. 4.06 beta, CLC Inc., Aarhus, Denmark) with default parameters. SOAP *de novo* gap closer was used to fill gaps based on alignment of paired-end reads (Luo et al., 2012). Principal contigs representing the cp genome were retrieved from total contigs using Nucmer (Delcher et al., 2003), and aligned contigs were ordered using the cp genome sequence of *I. nil* (AP017304) as a reference (Hoshino et al., 2016).

Genome Annotation and Comparative Analysis

Gene annotation of the six *Ipomoea* cp genomes was performed using GeSeq (Tillich et al., 2017), and the annotation results were concatenated using an in-house script pipeline. Protein-coding sequences were manually curated and confirmed using Artemis (Carver et al., 2008), and checked against the NCBI protein database. The tRNAs were confirmed with tRNAscan-SE 1.21 (Lowe and Eddy, 1997). IR region sequences were confirmed using IR finder and RepEx (Warburton et al., 2004; Gurusaran et al., 2013). Circular maps of the six *Ipomoea* cp genomes were obtained using OGDRAW (Lohse et al., 2007). GC content and relative synonymous codon usages (RSCU) were analyzed using MEGA6 software (Tamura et al., 2013). The mVISTA program in Shuffle-LAGAN mode was used to compare the six *Ipomoea* cp genomes using the *I. nil* cp genome as a reference. DnaSP version 5.1 (Librado and Rozas, 2009) was used to calculate nucleotide variability (Pi) among the six *Ipomoea* cp genomes. Substitution rates Ka and Ks were estimated with PAL2NAL (Suyama et al., 2006). LSC/IR, IR/SSC, SSC/IR, and IR/LSC regions of completed cp genomes were validated using PCR-based sequencing. Primer information and sequence alignment results are listed in Supplementary Tables S2, S3.

Repeat Analysis

SSRs in six *Ipomoea* cp genomes were detected using MISA (Thiel, 2003) with the minimum number of repeat parameters set to 10, 5, 4, 3, 3, and 3 for mono-, di-, tri-, tetra-, penta-, and hexa-nucleotides, respectively. Tandem repeats were ≥ 20 bp with minimum alignment score and maximum period size of 50 and 500, respectively, and the identity of repeats was set to $\geq 90\%$ (Benson, 1999).

Phylogenetic Analysis

A total of 38 cp genomes, including 36 from Convolvulaceae, were used for phylogenetic analyses, along with *Nicotiana tabacum* (GenBank acc. NC_001879.2) and *Capsicum annuum* var. *glabriusculum* (GenBank acc. KJ619462.1) as outgroups. Of these, 32 cp genome sequences were downloaded from the NCBI GenBank (Supplementary Table S4). MAFFT (Katoh et al., 2002) was used to construct molecular phylogenetic trees from alignments of 48 conserved protein-coding genes, and the sequences were manually adjusted using Bioedit (Hall, 1999). The best-fitting model of nucleotide substitutions was determined using Akaike Information Criterion in JModeltest V2.1.10 (Darriba et al., 2012). The GTR + I + G model was used in both. Maximum likelihood (ML) analysis was performed using RaxML v 8.0.5 (Stamatakis, 2014) with 1000 bootstrap replicates. Bayesian Inference (BI) analysis was performed using MrBayes 3.2.2 (Ronquist et al., 2012) with two-independent runs and four chains using Markov Chain Monte Carlo run simultaneously for one million generations. Trees were sampled every 5,000,000 generations, with the first 25% discarded as burn-in. Trees were determined from 50% majority-rule consensus trees to estimate posterior probabilities (PP). The reconstructed trees were visualized using Fig tree V.1.4.2 (Rambaut, 2012).

InDel Marker Development and Validation for *I. nil* and *I. purpurea*

Indel regions were detected by the alignment of six *Ipomoea* cp genome sequences and comparison of mVISTA similarities. Primers for indel markers were designed using NCBI Primer-BLAST. Specificity of indel markers was confirmed using PCR amplification with 20 ng of genomic DNA extracted from 23 samples of 6 *Ipomoea* species in a 20 μ l PCR mixture with 10 pmol of IPOTY or IPOYCF indel primers. Amplification of both IPOTY and IPOYCF was conducted on a Pro Flex PCR system (Applied Biosystems, Waltham, MA, United States) with the following amplification parameters: initial denaturation at 95°C for 2 min; 35 cycles at 95°C for 50 s, 62°C for 50 s, and 72°C for 50 s; and final extension at 72°C for 5 min. PCR products were separated on a 2% agarose gel for 40 min at 150V. DNA fragments were extracted from agarose using a Gel Extraction Kit, subcloned into the pGEM-T Easy vector (Promega, WI, United States), and sequenced on a DNA sequence analyzer (ABI 3730, Applied Biosystems Inc., CA, United States). The six *Ipomoea* species germplasms used are listed in Supplementary Table S1.

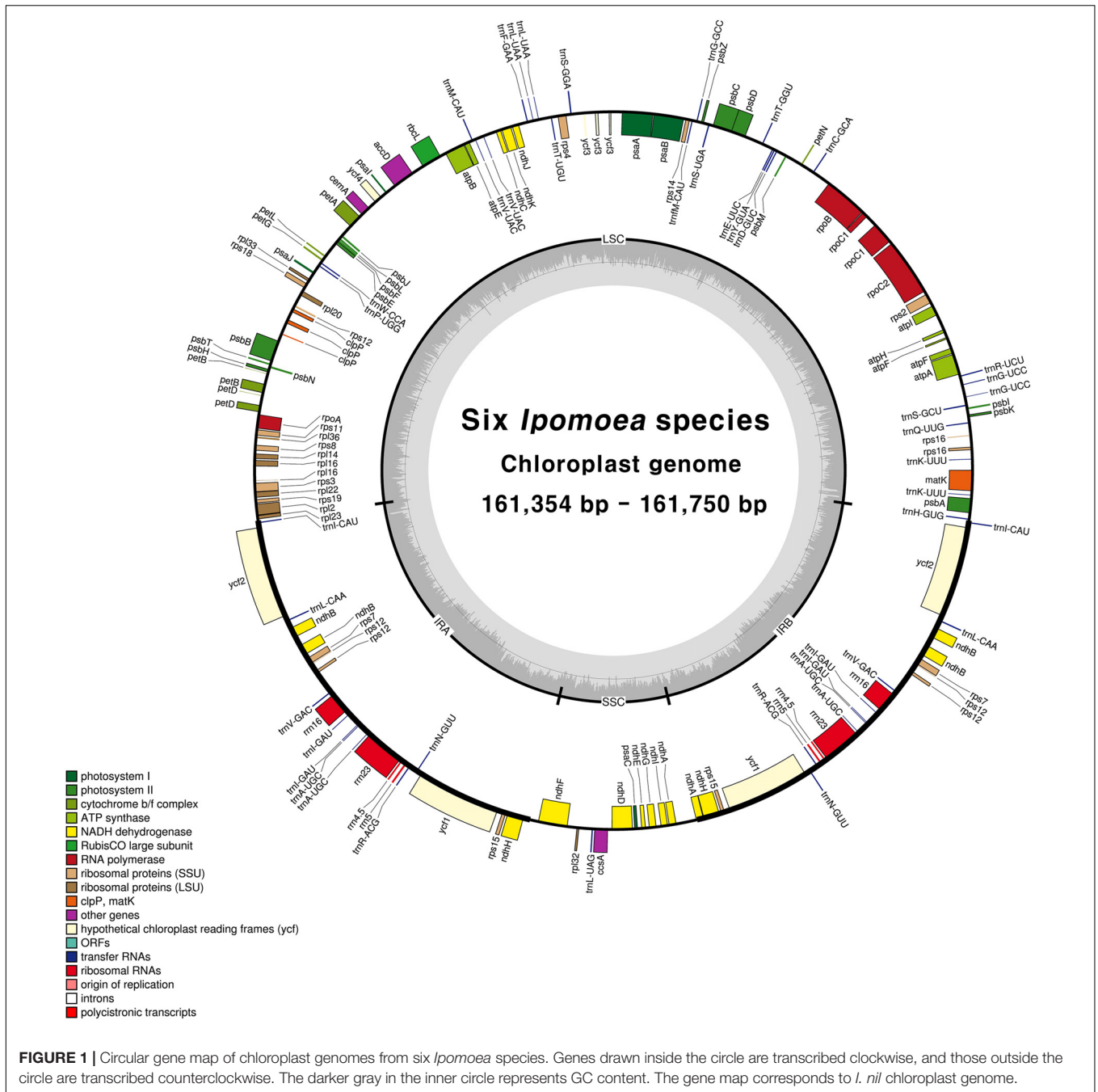
RESULTS AND DISCUSSION

Chloroplast Genome Organization of Six *Ipomoea* Species

Illumina sequencing generated 1.3–1.6 Gb of trimmed paired-end reads from six *Ipomoea* species (Supplementary Table S5). The six species yielded complete circular chloroplast cp genomes of 161,354–161,750 bp, with 384–611 \times coverage (Supplementary Table S6). As in most land plants, the *Ipomoea* cp genomes exhibited quadripartite structures consisting of a pair of IRs (61,220–62,122 bp) separated by LSC (87,579–88,134 bp) and SSC (12,039–12,101 bp) regions (Figure 1 and Table 1). The six *Ipomoea* cp genomes had similar GC contents, with higher GC contents observed in the IR regions (41%) than in the single-copy regions (LSC, 36% and SSC, 32%), consistent with previously reported cp genomes (Eserman et al., 2014; Yan et al., 2015). The gene content, order, and orientation were similar in the six *Ipomoea* cp genomes. The 112 unique genes consisted of 78 protein-coding genes and 30 tRNAs, with 17 duplicated genes in the 6 genomes (Table 2). Seventeen of the genes contained introns, fourteen with a single intron and two (*yec3* and *clpP*) with two introns (Supplementary Table S7). The genes *psbL* and *ndhD* had the alternative start codon ACG, and *rps19* started with GTG. Use of ACG and GTG as start codons is common for several genes in the cp genomes of land plants (Sasaki et al., 2003; Kahlau et al., 2006; Gao et al., 2009; Sanchez-Puerta and Abbona, 2014). The codon usage and anticodon recognition patterns of the six *Ipomoea* cp genomes are shown in Supplementary Figure S1A. Protein-coding genes comprised 28,280 codons in *I. hederacea* to 28,434 codons in *I. triloba*, which was consistent with other plant cp genomes. Codons for leucine, isoleucine, and serine were the most abundant, whereas those for cysteine and tryptophan were found least often (Supplementary Figure S1B). RSCU values revealed synonymous codon usage bias, with a high proportion of synonymous codons having A or T in the third position. As expected, codons for arginine, leucine, and serine had abundant synonymous codons and higher RSCU values due to their importance as components of cp genes related to biosynthetic processes (Wang et al., 2016). The RSCU values of the six *Ipomoea* cp genomes were consistent with those of other higher plants. This phenomenon is indicative of stable cp evolution, which protects important cp genes against harmful mutations and adaptive selective pressures in important cp genes (Wang et al., 2016; Ivanova et al., 2017; Zuo et al., 2017). The six *Ipomoea* genomes exhibited typical features of *Ipomoea* cp genomes and had similar genome structures, gene orders, and gene contents, including introns and base composition, to one another.

SSR and Tandem Repeat Analysis in Six *Ipomoea* Chloroplast Genomes

SSRs (1–6 nucleotide repeats) were distributed abundantly across the cp genome. SSRs from cp genomes can be used for analysis of phylogenetic relationships and population genetics due to their high polymorphism rates and stable reproducibility (Powell et al., 1995; Dong et al., 2013; Yang et al., 2016). Most SSRs



contained A or T units, contributing to the overall AT richness of the cp genome (Qian et al., 2013). Here, MISA software identified 191–202 SSRs in the six *Ipomoea* cp genomes. Most SSRs were found in single-copy regions (LSC and SSC) and non-coding regions (Figure 2). Mononucleotide motifs were the most abundant repeat type, with dinucleotide motifs, the second most abundant in the six *Ipomoea* cp genomes (Figure 2C). SSRs were compared between the genomes to identify common and species-specific SSRs. *I. purpurea* contained seven specific SSRs, whereas no *I. hederacea*-specific SSRs were identified. The *Ipomoea* cp SSRs encompassed abundant variation and

will be useful genomic resources for marker development and population genetic studies of *Ipomoea* species. Tandem repeat sequences influence genome structure with respect to genome size, genome rearrangement, and gene duplication (Nie et al., 2012). Here, 27–33 tandem repeats of >20 bp were identified in the six *Ipomoea* genomes, averaging 56–74 bp in length (Figure 3). Of these, most were located in non-coding LSC and SSC regions (Figure 3A). On average, long tandem repeats (>70 bp) constituted 27% of all tandem repeats in the six *Ipomoea* cp genomes. *I. triloba* contained the longest repeat (484 bp). While most of the tandem repeats were conserved,

TABLE 1 | Features of six *Ipomoea* chloroplast genomes.

Species	<i>I. nil</i>	<i>I. purpurea</i>	<i>I. hederacea</i>	<i>I. hederacea</i> var. <i>integriscula</i>	<i>I. lacunosa</i>	<i>I. triloba</i>
Accession number	MG973745	MG973746	MG973747	MG973748	MG973749	MG973750
Total cp genome size (bp)	161,747	161,629	161,354	161,714	161,492	161,750
Large single-copy (LSC) region (bp)	88,097	88,134	88,041	88,083	87,579	87,589
Inverted repeat (IR) region (bp)	61,564	61,394	61,220	61,538	61,862	62,122
Small single-copy (SSC) region (bp)	12,086	12,101	12,093	12,093	12,051	12,039
Total number of genes (unique)	112	112	112	112	112	112
Protein-coding gene (unique)	78	78	78	78	78	78
rRNA (unique)	4	4	4	4	4	4
tRNA (unique)	30	30	30	30	30	30
GC content (%)	38	38	38	38	38	38
LSC (%)	36	36	36	36	36	36
IR (%)	41	41	41	41	41	41
SSC (%)	32	32	32	32	32	32

TABLE 2 | Genes in the chloroplast genomes of six *Ipomoea* species.

Gene groups	Gene names
Photosystem I	<i>psaA, B, C, I, J, ycf3², ycf4</i>
Photosystem II	<i>psbA, B, C, D, E, F, H, I, J, K, L, M, N, T, Z</i>
Cytochrome b6/f	<i>petA, B¹, D¹, G, L, N</i>
ATP synthase	<i>atpA, B, E, F¹, H, I</i>
Rubisco	<i>rbcL</i>
NADH oxidoreductase	<i>ndhA¹, B^{1,3}, C, D, E, F, G, H³, I, J, K</i>
Large subunit ribosomal proteins	<i>rpl2, 14, 16¹, 20, 22, 23, 32, 33, 36</i>
Small subunit ribosomal proteins	<i>rps2, 3, 4, 7³, 8, 11, 12²⁻⁴, 14, 15³, 16¹, 18, 19</i>
RNA polymerase	<i>rpoA, B, C1¹, C2</i>
Unknown function protein-coding gene	<i>ycf1³, 2³</i>
Other genes	<i>accD, ccsA, cemA, clpP², matK</i>
Ribosomal RNAs	<i>rrn16³, 23³, 4.5³, 5³</i>
Transfer RNAs	<i>trnA-UGC^{1,3}, trnC-GCA, trnD-GUC, trnE-UUC, trnF-GAA, trnM-CAU, trnG-GCC, trnG-UCC¹, trnH-GUG, trnI-CAU³, trnI-GAU^{1,3}, trnK-UUU¹, trnL-CAA³, trnL-UAA, trnL-UAG, trnM-CAU, trnN-GUU³, trnP-UGG, trnQ-UUG, trnR-ACG³, trnR-UCU, trnS-GCU, trnS-GGA, trnS-UGA, trnT-GGU, trnT-UGU, trnV-GAC, trnV-UAC, trnW-CCA, trnY-GUA</i>

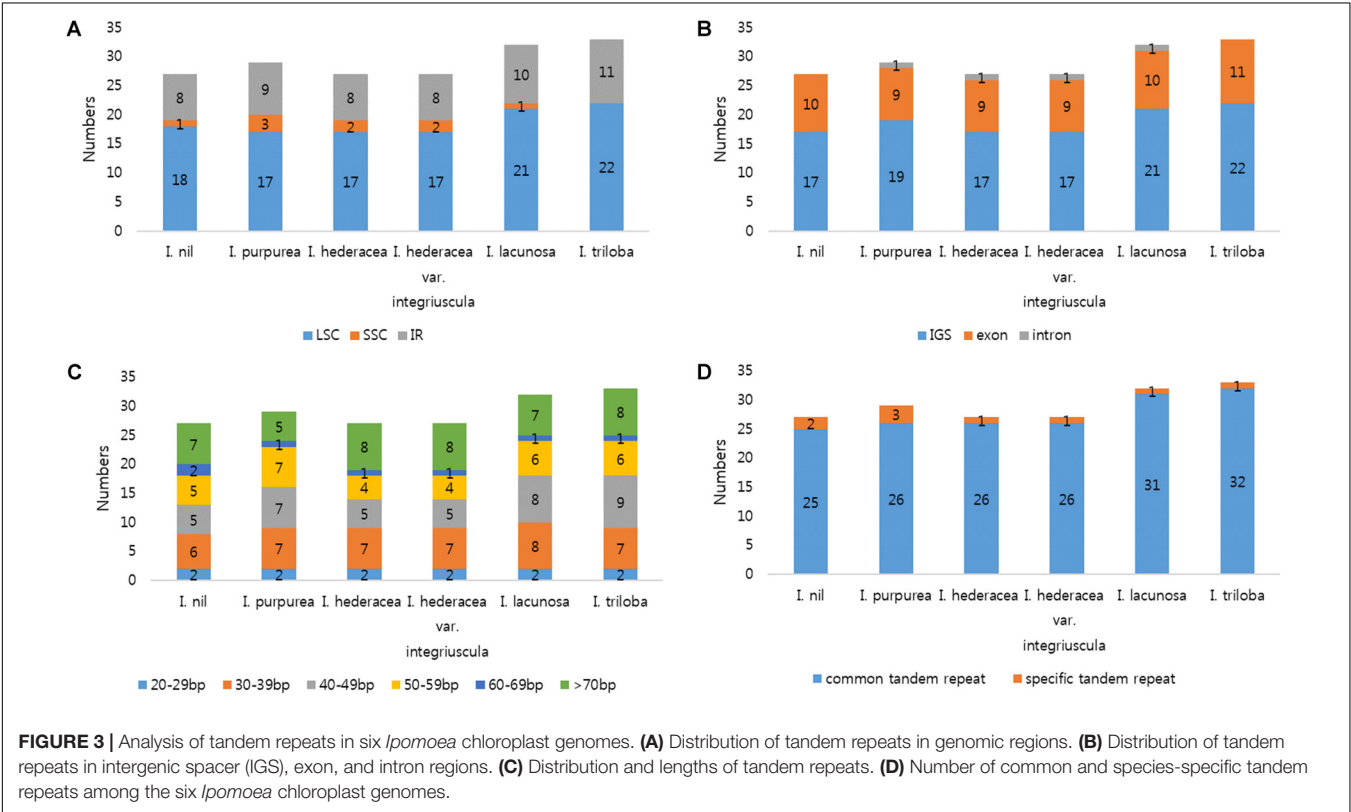
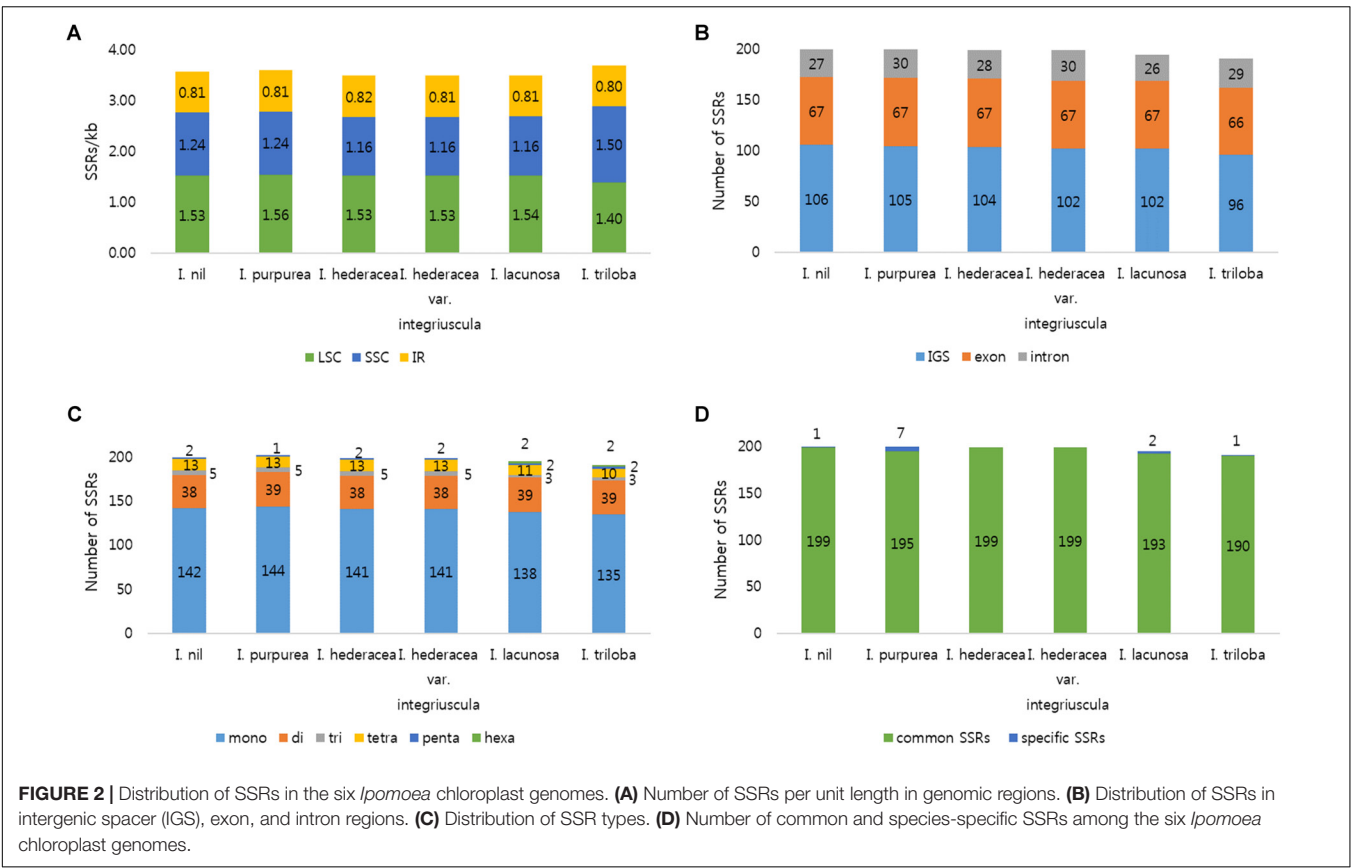
¹Gene containing a single intron; ²Gene containing two introns; ³Two gene copies in IRs; ⁴Trans-splicing gene.

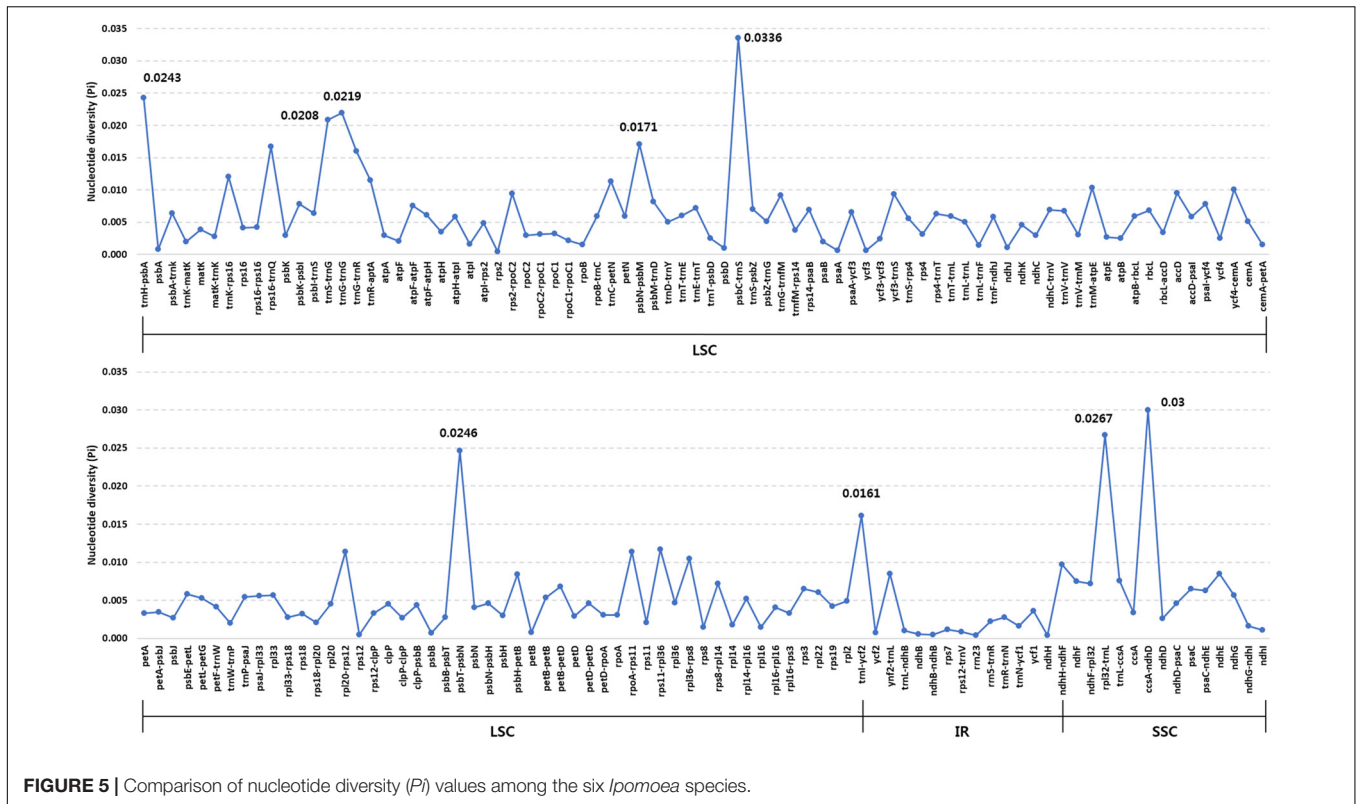
a small number of specific tandem repeats were detected in the six *Ipomoea* cp genomes (**Figure 3D**). In particular, indel regions with different repeat copy numbers were identified for *I. nil* in *trnN-ycf1* (68 bp × one copy in *I. nil* / 68 bp × two copies in other *Ipomoea*) and *I. purpurea* in *ycf1* (18 bp × two copies in *I. purpurea* / 18 bp × three copies in other *Ipomoea*). These characteristics allowed indel markers for distinguishing *I. nil* and *I. purpurea* from other *Ipomoea* species to be developed in this study.

Comparative Analysis of Six *Ipomoea* Chloroplast Genomes

Overall, the six *Ipomoea* cp genomes were highly conserved, with 98.5–99.8% similarity, conserved genomic structure, and conserved gene order and orientation (Supplementary Table S8). Pairwise determination of divergent regions was conducted using mVISTA (**Figure 4**). In general, non-coding regions were more diverged than coding regions. Five non-coding regions, *petN-psbM*, *trnI-CAU-ycf2*, *ndhH-ndhF*, *psbC-trnS*, and

ccsA-ndhD, exhibited high divergence among the six *Ipomoea*. Coding regions were generally more conserved, with the exception of *ycf1*, *matK*, and *rbcL*, which are commonly used as representative plant DNA barcoding regions (CBOL Plant Working Group, 2009). Previous phylogenetic analysis of divergent non-coding regions allowed identification of potential molecular markers and DNA barcoding analysis (Shaw et al., 2007; Xu et al., 2017). *Pi* in the six *Ipomoea* cp genomes was calculated to show divergence at the sequence level (**Figure 5**). As expected, IR regions were more conserved than the LSC and SSC regions, with average *Pi* values of 0.003 for IR and 0.006 in SC (for regions other than those with a *Pi* value = 0). The average *Pi* value for coding regions was 0.00315 (range, 0.00038–0.00955; *accD* = 0.00955). The *Pi* value for intron-containing IGS averaged 0.00752 (range, 0.0005–0.00336; *psbC-trnS* in LSC = 0.00336). In the SSC, *ccsA-ndhD* exhibited a *Pi* value of 0.03, higher by an order of magnitude. Although the six *Ipomoea* cp genomes were generally highly conserved, the intergenic regions were particularly divergent. This is consistent



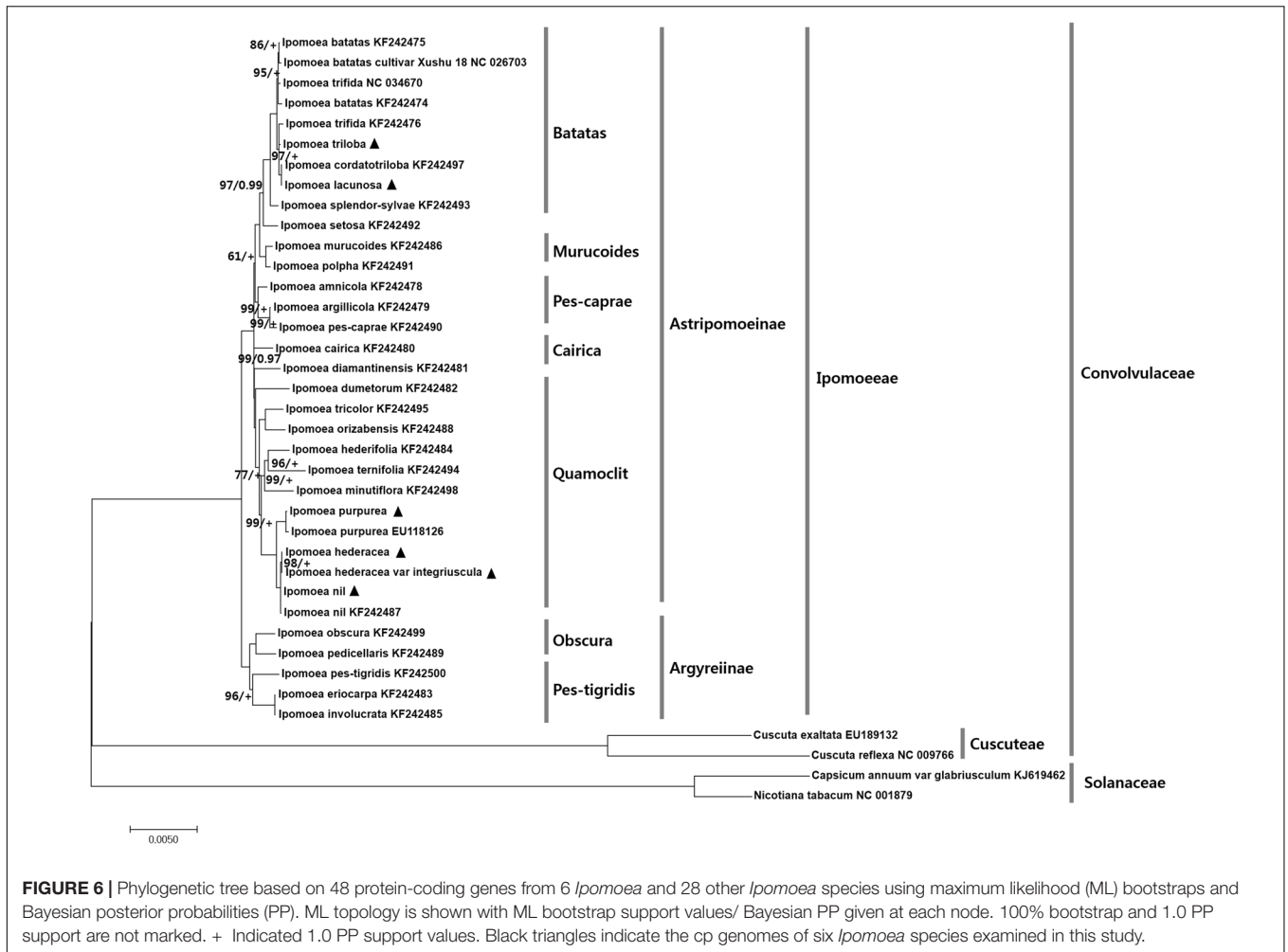


reported that *accD*, *cemA*, and *ycf2* genes in both *Ipomoea* and *Cuscuta* were under purifying selection pressure. However, in the current study, we found that these genes were under positive selection pressure. Although plants within the *Cuscuta* genus express markedly diverse genes to adapt to life as parasitic plants, these genes might be rapidly evolving in the *Ipomoea* genus. Several studies show that nuclear genes in *Ipomoea* are under positive selection pressure. In particular, genes encoding dihydroflavonol-4 reductase (DFR) and chalcone synthase (CHS) in *Ipomoea* are under positive selection pressure (Yang et al., 2004; Des Marais and Rausher, 2008). DFR is an important factor in the anthocyanin biosynthetic pathway. Des Marais and Rausher (2008) demonstrated that escape from adaptive conflict *via* repeated positive selection occurred after DFR genes duplicated in *I. purpurea*. Thus, DFR genes exhibit adaptive evolutionary changes. The CHS genes (which function in flavonoid biosynthesis) experienced selective pressure to promote divergence *via* increasing gene duplication in *Ipomoea*. *Ipomoea* has extremely diverse flower colors, a rapid generation time, and various growth forms. We suggest that these ecological characteristics of *Ipomoea* reflect their remarkable adaptability to various environments due to diverse positive selection pressure on genes in the nucleus or plastid.

Phylogenetic Relationships of Six *Ipomoea* Within Ipomoeae

Cp genomes are valuable genomic resources for reconstruction of accurate and high-resolution phylogenies, and have been used as such in several studies (Jansen et al., 2007; Moore et al., 2007),

for example, in angiosperms (Wu et al., 2010; Nie et al., 2012). To identify the phylogenetic positions of the six *Ipomoea* species within the Convolvulaceae, we aligned 48 protein-coding sequences shared by 38 cp genomes (Figure 6 and Supplementary Figure S4). The alignment length was 38,229 bp. All except two nodes were supported by a Bayesian PP of 1.0. *Ipomoea* and *Cuscuta* had the closest phylogenetic relationship within the Convolvulaceae. Consistent with previous analysis, the tribe Ipomoeae was divided into 2 major clades, Atstipomoeinae and Atgyreinae, with 28 *Ipomoea* species within 7 sections (Stefanović et al., 2003). Most *Ipomoea* species were within Quamoclit and Batatas. The positions of the six *Ipomoea* examined in this study were strongly supported with BI and PP values. *I. nil* and *I. hederacea* formed a monophyletic cluster as a sister to *I. purpurea* within Quamoclit. *I. lacunosa* and *I. triloba* formed a monophyletic cluster in Batatas. Previous analysis of 28 *Ipomoea* cp genomes clarified the evolutionary relationships within the 2 major clades of Ipomoeae (Eserman et al., 2014). Quamoclit species were divided into two clades (Miller et al., 2004), whereas our phylogenetic results revealed their monophyly. In this study, *I. hederacea* and *I. hederacea* var. *integriscula* were clustered with *I. nil* in a monophyletic relationship, but that *I. purpurea* was paraphyletic with these species. A previous study indicated that *I. nil* and *I. purpurea* share a monophyletic relationship. Here, we obtained more accurate information about the relationship between *I. nil* and *I. hederacea* and *I. hederacea* var. *integriscula* and *I. purpurea*. Furthermore, *I. triloba* and *I. lacunosa* were positioned in Batatas as a monophyletic group with *I. trifida* and *I. cordatotriloba*.



Therefore, we performed high-resolution phylogenetic analysis of the positions of *Ipomoea* species in Batatas in the phylogenetic tree. The reconstructed phylogenetic trees were clearly consistent with previous studies according to the APG IV system (Austin, 1978; Eserman et al., 2014; The Angiosperm Phylogeny Group, 2016). The results of this study are strongly supported by those of previous studies; however, we further clarified the phylogenetic relationships within the Ipomoeaceae. Based on the phylogenetic positions of the six *Ipomoea* species determined in the present study, *I. nil* and *I. purpurea* (whose seeds are used for Pharbitidis Semen) share closer relationships with *I. hederacea* and *I. hederacea* var. *integrariuscula* than with *I. triloba* and *I. lacunosa*. Thus, there is a strong possibility for confusion between *I. hederacea* and *I. hederacea* var. *integrariuscula*. We suspect that the most frequent adulterations of Pharbitidis Semen are seeds of *I. hederacea* and *I. hederacea* var. *integrariuscula*.

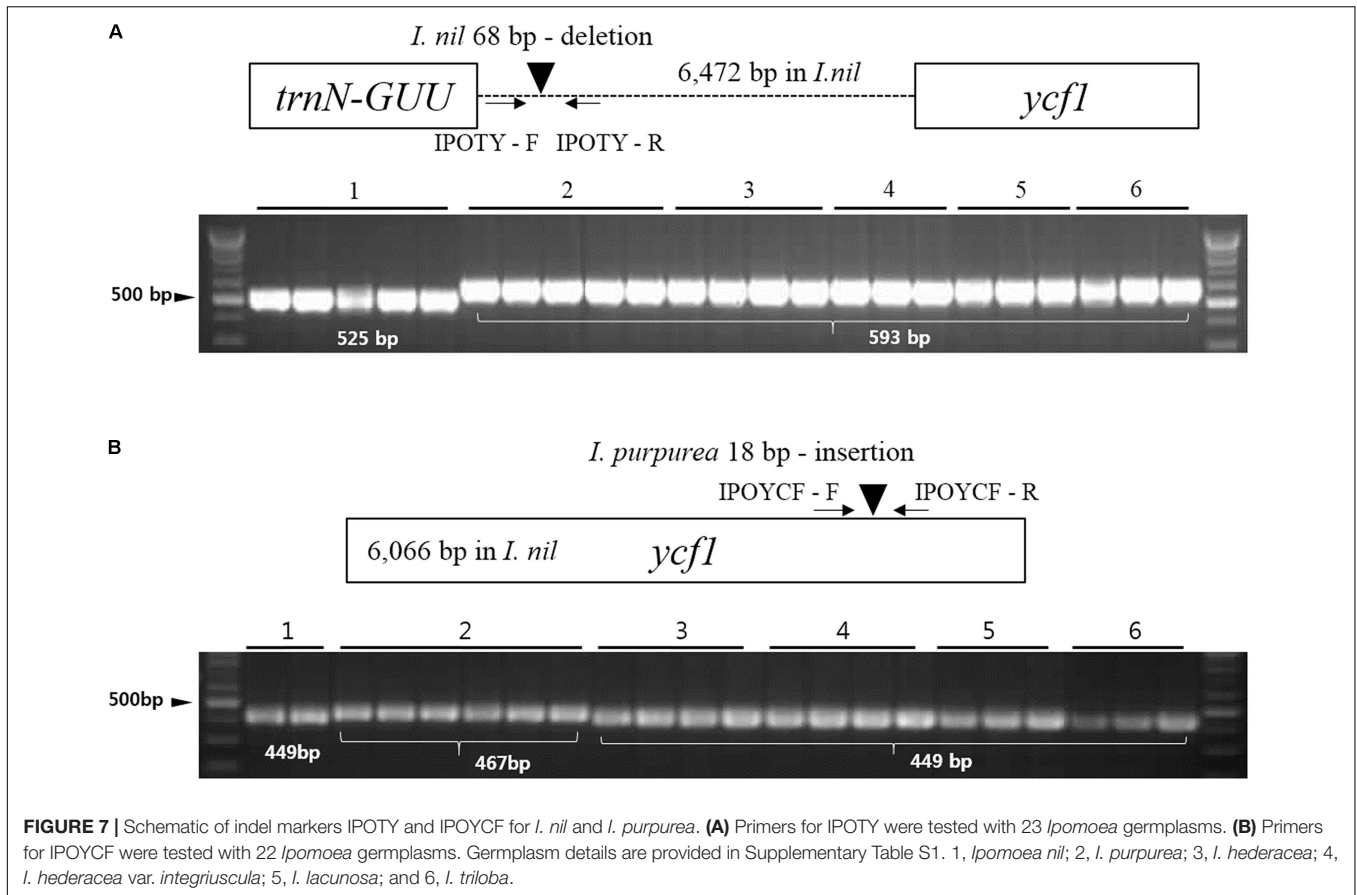
New Indel Markers for Distinguishing Herbal Medicine Plants

Dried seeds of *I. nil* and *I. purpurea* are used in traditional herbal medicine in Korea (Korea Institute of Oriental Medicine

[KIOM], 2016). However, seeds from other *Ipomoea* strongly resemble those of *I. nil* and *I. purpurea* and are often inappropriately included in herbal preparations. Although the phylogenetic analysis in this study indicated that *I. triloba* and *I. lacunosa* were phylogenetically distant from *I. nil* and *I. purpurea*, the highly similar seed shapes present a challenge for identifying authentic herbal medicines, and a molecular approach would be beneficial. Here, DNA barcode analysis was performed for *I. nil* and *I. purpurea* and four related *Ipomoea* species, *I. triloba*, *I. lacunosa*, *I. hederacea*, and *I. hederacea* var. *integrariuscula*, with similar seed shapes. *I. purpurea* was distinguished from other *Ipomoea* species with the ITS2 and *matK* regions (Supplementary Figure S5). However, the sequence of *I. nil* at ITS2 was the same as that of *I. hederacea*, and the

TABLE 3 | Primer information for indel markers IPOTY and IPOYCF.

Primer name	Primer sequence (5' > 3')	Position
IPOTY_F	TAACGGTCAAAGCGAGCCCC	<i>trnN-GUU-ycf1</i>
IPOTY_R	AAGTCCAGCCGCAAGAAGCTGA	
IPOYCF_F	GGTCGCGGTAATCCCAGCA	<i>ycf1</i>
IPOYCF_R	TCTTCCCAGAATTTGTGCGGC	



sequence of *I. nil* at *matK* was the same as that of *I. lacunosa* and *I. triloba*, highlighting the limitations of universal DNA barcode sequences for distinguishing species. To resolve this problem, divergent regions within the cp genome were examined with the aim of distinguishing *I. nil* and *I. purpurea*. This analysis revealed species-specific divergent regions at *trnN-ycf1* and *ycf1* for *I. nil* and *I. purpurea*, respectively, with respect to copy number variation in tandem repeats. To develop indel markers, specific primers were designed against conserved regions of *trnN-ycf1* and *ycf1* (Table 3). The primer pairs, respectively named IPOTY and IPOYCF, successfully amplified sequences from *I. nil* and *I. purpurea* (Figure 7). The markers were tested with other *Ipomoea* germplasms (23 samples for IPOTY and 22 samples for IPOYCF), and the five *I. nil* samples and six *I. purpurea* samples were clearly distinguishable. Amplified fragments from all of the tested *Ipomoea* samples were sequenced to identify exact amplicon size. IPOTY primers yielded a 525 bp amplicon with *I. nil*, and IPOYCF primers yielded a 467 bp fragment with *I. purpurea*. Predicted deletion or insertion sizes from cp genomes were consistent with those resulting from the *Ipomoea* germplasms used in this study. Indel markers for variable copy numbers at tandem repeats were also used previously to distinguish closely related *Fagopyrum* and *Chenopodium* species, indicating the utility of these markers in species identification (Cho et al., 2015; Hong et al., 2017). Copy number variation at tandem repeats in cp genomes may, therefore, prove broadly

useful in distinguishing closely related plant species where universal barcode sequences are non-discriminatory. The IPOTY and IPOYCF indel markers developed in this study will be useful for *Ipomoea* species identification and authentication of herbal medicines.

CONCLUSION

Six *Ipomoea* cp genomes were sequenced in this study. Overall, the cp genomes were highly conserved with respect to gene content, gene orientation, and GC content, but local variations in sequence and structure were observed. Tandem repeats and SSRs were identified with the aim of developing molecular markers for species identification and authentication of herbal medicines. The most divergent regions among the six genomes were found in non-coding regions *petN-psbM*, *trnI-CAU-ycf2*, *ndhH-ndhF*, *psbC-trnS*, and *ccsA-ndhD*, and coding regions *accD*, *cemA*, and *ycf2*. The *accD*, *cemA*, and *ycf2* genes exhibited positive selection. Phylogenetic analysis of cp genome sequences yielded more accurate phylogenetic relationships within the *Ipomoea* genus than previous studies. Novel indel markers based on copy number variation at tandem repeats were developed for identification of *I. nil* and *I. purpurea*. These markers, named IPOTY and IPOYCF, were able to discriminate between authentic *I. nil* and *I. purpurea* and other inauthentic *Ipomoea*

species, respectively, and will be useful for authentication of herbal medicines containing these two species. The cp genomes and analyses in this study are valuable for species identification, clarification of taxonomy, and understanding evolutionary history in the *Ipomoea* genus.

AUTHOR CONTRIBUTIONS

IP designed the experimental framework and drafted and revised the manuscript. SY and BCM collected and identified plant materials. WJK and PN performed the experiments. HOL carried out sequence analysis. BCM revised the manuscript. All authors contributed to the experiments and approved the final manuscript.

FUNDING

This work was supported by a grant of the Development of Foundational Techniques for the Domestic Production of

REFERENCES

- Austin, D. F. (1978). The *Ipomoea batatas* complex-I. Taxonomy. *Bull. Torrey Bot. Club* 105, 114–129. doi: 10.2307/2484429
- Austin, D. F., and Huáman, Z. (1996). A synopsis of *Ipomoea* (Convolvulaceae) in the Americas. *Taxon* 45, 3–38. doi: 10.2307/1222581
- Baucom, R., Chang, S., Kniskern, J., Rausher, M., and Stinchcombe, J. (2011). Morning glory as a powerful model in ecological genomics: tracing adaptation through both natural and artificial selection. *Heredity* 107, 377–385. doi: 10.1038/hdy.2011.25
- Benson, G. (1999). Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* 27, 573–580. doi: 10.1093/nar/27.2.573
- Carver, T., Berriman, M., Tivey, A., Patel, C., Bohme, U., Barrell, B. G., et al. (2008). Artemis and ACT: viewing, annotating and comparing sequences stored in a relational database. *Bioinformatics* 24, 2672–2676. doi: 10.1093/bioinformatics/btn529
- CBOL Plant Working Group (2009). A DNA barcode for land plants. *Proc. Natl. Acad. Sci. U.S.A.* 106, 12794–12797. doi: 10.1073/pnas.0905845106
- Chen, S., Yao, H., Han, J., Liu, C., Song, J., Shi, L., et al. (2010). Validation of the ITS2 region as a novel DNA barcode for identifying medicinal plant species. *PLoS One* 5:e8613. doi: 10.1371/journal.pone.0008613
- Cho, K. S., Yun, B. K., Yoon, Y. H., Hong, S. Y., Mekapogu, M., Kim, K. H., et al. (2015). Complete chloroplast genome sequence of tartary buckwheat (*Fagopyrum tataricum*) and comparative analysis with common buckwheat (*F. esculentum*). *PLoS One* 10:e0125332. doi: 10.1371/journal.pone.0125332
- Daniell, H., Lin, C. S., Yu, M., and Chang, W. J. (2016). Chloroplast genomes: diversity, evolution, and applications in genetic engineering. *Genome Biol.* 17:134. doi: 10.1186/s13059-016-1004-2
- Darriba, D., Taboada, G. L., Doallo, R., and Posada, D. (2012). jModelTest 2: more models, new heuristics and parallel computing. *Nat. Methods* 9:772. doi: 10.1038/nmeth.2109
- Delannoy, E., Fujii, S., des Francs-Small, C., Brundrett, M., and Small, I. (2011). Rampant gene loss in the underground orchid *Rhizanthella gardneri* highlights evolutionary constraints on plastid genomes. *Mol. Biol. Evol.* 28, 2077–2086. doi: 10.1093/molbev/msr028
- Delcher, A. L., Salzberg, S. L., and Phillippy, A. M. (2003). Using MUMmer to identify similar regions in large sequence sets. *Curr. Protoc. Bioinform.* Chapter 10:Unit 10.13. doi: 10.1002/0471250953.bi1003s00
- Des Marais, D. L., and Rausher, M. D. (2008). Escape from adaptive conflict after duplication in an anthocyanin pathway gene. *Nature* 454, 762–765. doi: 10.1038/nature07092
- Authentic Herbal Medicines based on the Establishment of Molecular Authentication Systems (K17403 and K18403) from the Korea Institute of Oriental Medicine (KIOM), South Korea.

ACKNOWLEDGMENTS

The authors would like to thank the “Classification and Identification Committee of the KIOM” for the identification of plant materials and the Herbarium of Korea Standard Herbal Resources (Index Herbariorum code KIOM) for the provision of plant materials.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2018.00965/full#supplementary-material>

- Dong, W., Xu, C., Cheng, T., Lin, K., and Zhou, S. (2013). Sequencing angiosperm plastid genomes made easy: a complete set of universal primers and a case study on the phylogeny of Saxifragales. *Genome Biol. Evol.* 5, 989–997. doi: 10.1093/gbe/evt063
- Ennos, R. A., and Clegg, M. T. (1983). Flower color variation in the morning glory, *Ipomoea purpurea*. *J. Hered.* 74, 247–250. doi: 10.1093/oxfordjournals.jhered.a109778
- Epperson, B., and Clegg, M. (1992). Unstable white flower color genes and their derivatives in morning glory. *J. Hered.* 83, 405–409. doi: 10.1093/oxfordjournals.jhered.a111242
- Eserman, L. A., Tiley, G. P., Jarret, R. L., Leebens-Mack, J. H., and Miller, R. E. (2014). Phylogenetics and diversification of morning glories (tribe Ipomoeae, Convolvulaceae) based on whole plastome sequences. *Am. J. Bot.* 101, 92–103. doi: 10.3732/ajb.1300207
- Gao, L., Yi, X., Yang, Y. X., Su, Y. J., and Wang, T. (2009). Complete chloroplast genome sequence of a tree fern *Alsophila spinulosa*: insights into evolutionary changes in fern chloroplast genomes. *BMC Evol. Biol.* 9:130. doi: 10.1186/1471-2148-9-130
- Gurusaran, M., Ravella, D., and Sekar, K. (2013). RepEx: repeat extractor for biological sequences. *Genomics* 102, 403–408. doi: 10.1016/j.ygeno.2013.07.005
- Hall, T. A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acid Symp. Ser.* 41, 95–98.
- Hansen, D. R., Dastidar, S. G., Cai, Z., Penaflor, C., Kuehl, J. V., Boore, J. L., et al. (2007). Phylogenetic and evolutionary implications of complete chloroplast genome sequences of four early-diverging angiosperms: *Buxus* (Buxaceae), *Chloranthus* (Chloranthaceae), *Dioscorea* (Dioscoreaceae), and *Illicium* (Schisandraceae). *Mol. Phylogenet. Evol.* 45, 547–563. doi: 10.1016/j.ympev.2007.06.004
- Hollingsworth, P. M., Graham, S. W., and Little, D. P. (2011). Choosing and using a plant DNA barcode. *PLoS One* 6:e19254. doi: 10.1371/journal.pone.0019254
- Hong, S. Y., Cheon, K. S., Yoo, K. O., Lee, H. O., Cho, K. S., Suh, J. T., et al. (2017). complete chloroplast genome sequences and comparative analysis of *Chenopodium quinoa* and *C. album*. *Front. Plant Sci.* 8:1696. doi: 10.3389/fpls.2017.01696
- Hoshino, A., Jayakumar, V., Nitasaka, E., Toyoda, A., Noguchi, H., Itoh, T., et al. (2016). Genome sequence and analysis of the Japanese morning glory *Ipomoea nil*. *Nat. Commun.* 7:13295. doi: 10.1038/ncomms13295
- Huang, H., Shi, C., Liu, Y., Mao, S. Y., and Gao, L. Z. (2014). Thirteen *Camellia* chloroplast genome sequences determined by high-throughput sequencing: genome structure and phylogenetic relationships. *BMC Evol. Biol.* 14:151. doi: 10.1186/1471-2148-14-151

- Ivanova, Z., Sablok, G., Daskalova, E., Zahmanova, G., Apostolova, E., Yahubyan, G., et al. (2017). Chloroplast genome analysis of resurrection tertiary relict *Haberlea rhodopensis* highlights genes important for desiccation stress response. *Front. Plant Sci.* 8:204. doi: 10.3389/fpls.2017.00204
- Jansen, R. K., Cai, Z., Raubeson, L. A., Daniell, H., Leebens-Mack, J., Müller, K. F., et al. (2007). Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns. *Proc. Natl. Acad. Sci. U.S.A.* 104, 19369–19374. doi: 10.1073/pnas.0709121104
- Jansen, R. K., and Ruhlman, T. A. (2012). “Plastid genomes of seed plants,” in *Genomics of Chloroplasts and Mitochondria: Advances in Photosynthesis and Respiration (Including Bioenergy and Related Processes)*, Vol. 35, eds R. Bock and V. Knoop (Dordrecht: Springer).
- Kahlau, S., Aspinall, S., Gray, J. C., and Bock, R. (2006). Sequence of the tomato chloroplast DNA and evolutionary comparison of solanaceous plastid genomes. *J. Mol. Evol.* 63, 194–207. doi: 10.1007/s00239-005-0254-5
- Katoh, K., Misawa, K., Kuma, K. I., and Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* 30, 3059–3066. doi: 10.1093/nar/gkf436
- Khakhlova, O., and Bock, R. (2006). Elimination of deleterious mutations in plastid genomes by gene conversion. *Plant J.* 46, 85–94. doi: 10.1111/j.1365-313X.2006.02673.x
- Kim, K., Lee, S. C., Lee, J., Lee, H. O., Joh, H. J., Kim, N. H., et al. (2015). Comprehensive survey of genetic diversity in chloroplast genomes and 45S rDNAs within *Panax ginseng* species. *PLoS One* 10:e0117159. doi: 10.1371/journal.pone.0117159
- Kimura, M. (1989). The neutral theory of molecular evolution and the world view of the neutralists. *Genome* 31, 24–31. doi: 10.1139/g89-009
- Korea Institute of Oriental Medicine [KIOM] (2016). *Defining Dictionary for Medicinal Herbs*. Available at: <http://boncho.kiom.re.kr/codex/>
- Korean Food Standard Codex (2010). *Ministry of Food and Drug Safety (Republic of Korea)*. Available at: http://www.mfds.go.kr/files/upload/eng/9_Foods_Labeling_Standards.pdf
- Librado, P., and Rozas, J. (2009). DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25, 1451–1452. doi: 10.1093/bioinformatics/btp187
- Liu, Y., Huo, N., Dong, L., Wang, Y., Zhang, S., Young, H. A., et al. (2013). Complete chloroplast genome sequences of Mongolia medicine *Artemisia frigida* and phylogenetic relationships with other plants. *PLoS One* 8:e57533. doi: 10.1371/journal.pone.0057533
- Lohse, M., Drechsel, O., and Bock, R. (2007). OrganellarGenomeDRAW (OGDRAW): a tool for the easy generation of high-quality custom graphical maps of plastid and mitochondrial genomes. *Curr. Genet.* 52, 267–274. doi: 10.1007/s00294-007-0161-y
- Lowe, T. M., and Eddy, S. R. (1997). tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* 25, 955–964. doi: 10.1093/nar/25.5.0955
- Luo, R., Liu, B., Xie, Y., Li, Z., Huang, W., Yuan, J., et al. (2012). SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *Gigascience* 1:18. doi: 10.1186/2047-217X-1-18
- Manos, P. S., Miller, R. E., and Wilkin, P. (2001). Phylogenetic analysis of *Ipomoea*, *Argyrea*, *Stictocardia*, and *Turbina* suggests a generalized model of morphological evolution in morning glories. *Syst. Bot.* 26, 585–602.
- McDonald, J. A. (1995). Revision of *Ipomoea* section *Leptocallis* (Convolvulaceae). *Harv. Pap. Bot.* 1, 97–122.
- McNeal, J. R., Kuehl, J. V., Boore, J. L., and de Pamphilis, C. W. (2007). Complete plastid genome sequences suggest strong selection for retention of photosynthetic genes in the parasitic plant genus *Cuscuta*. *BMC Plant Biol.* 7:57. doi: 10.1186/1471-2229-7-57
- Miller, R. E., McDonald, J. A., and Manos, P. S. (2004). Systematics of *Ipomoea* subgenus *Quamoclit* (Convolvulaceae) based on ITS sequence data and a Bayesian phylogenetic analysis. *Am. J. Bot.* 91, 1208–1218. doi: 10.3732/ajb.91.8.1208
- Moore, M. J., Bell, C. D., Soltis, P. S., and Soltis, D. E. (2007). Using plastid genome-scale data to resolve enigmatic relationships among basal angiosperms. *Proc. Natl. Acad. Sci. U.S.A.* 104, 19363–19368. doi: 10.1073/pnas.0708072104
- Nie, X., Lv, S., Zhang, Y., Du, X., Wang, L., Biradar, S. S., et al. (2012). Complete chloroplast genome sequence of a major invasive species, crofton weed (*Ageratina adenophora*). *PLoS One* 7:e36869. doi: 10.1371/journal.pone.0036869
- Park, I., Kim, W. J., Yang, S., Yeo, S. M., Li, H., and Moon, B. C. (2017a). The complete chloroplast genome sequence of *Aconitum coreanum* and *Aconitum carmichaelii* and comparative analysis with other *Aconitum* species. *PLoS One* 12:e0184257. doi: 10.1371/journal.pone.0184257
- Park, I., Yang, S., Choi, G., Kim, W. J., and Moon, B. C. (2017b). The complete chloroplast genome sequences of *Aconitum pseudolaevae* and *Aconitum longecassidatum*, and development of molecular markers for distinguishing species in the *Aconitum* Subgenus *Lycocotnum*. *Molecules* 22:E2012. doi: 10.3390/molecules22112012
- Parks, M., Cronn, R., and Liston, A. (2009). Increasing phylogenetic resolution at low taxonomic levels using massively parallel sequencing of chloroplast genomes. *BMC Biol.* 7:84. doi: 10.1186/1741-7007-7-84
- Powell, W., Morgante, M., McDevitt, R., Vendramin, G., and Rafalski, J. (1995). Polymorphic simple sequence repeat regions in chloroplast genomes: applications to the population genetics of pines. *Proc. Natl. Acad. Sci. U.S.A.* 92, 7759–7763. doi: 10.1073/pnas.92.17.7759
- Qian, J., Song, J., Gao, H., Zhu, Y., Xu, J., Pang, X., et al. (2013). The complete chloroplast genome sequence of the medicinal plant *Salvia miltiorrhiza*. *PLoS One* 8:e57607. doi: 10.1371/journal.pone.0057607
- Raman, G., and Park, S. (2016). The complete chloroplast genome sequence of *Ampelopsis*: gene organization, comparative analysis, and phylogenetic relationships to other angiosperms. *Front. Plant Sci.* 7:341. doi: 10.3389/fpls.2016.00341
- Rambaut, A. (2012). *FigTree v 1.4. 2: Molecular Evolution, Phylogenetics and Epidemiology*. Edinburgh: University of Edinburgh.
- Raubeson, L. A., Peery, R., Chumley, T. W., Dziubek, C., Fourcade, H. M., Boore, J. L., et al. (2007). Comparative chloroplast genomics: analyses including new sequences from the angiosperms *Nuphar advena* and *Ranunculus macranthus*. *BMC Genomics* 8:174. doi: 10.1186/1471-2164-8-174
- Rausher, M. D., and Fry, J. D. (1993). Effects of a locus affecting floral pigmentation in *Ipomoea purpurea* on female fitness components. *Genetics* 134, 1237–1247.
- Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D. L., Darling, A., Höhna, S., et al. (2012). MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* 61, 539–542. doi: 10.1093/sysbio/sys029
- Sanchez-Puerta, M. V., and Abbona, C. C. (2014). The chloroplast genome of *Hyoscyamus niger* and a phylogenetic study of the tribe Hyoscyameae (Solanaceae). *PLoS One* 9:e98353. doi: 10.1371/journal.pone.0098353
- Sasaki, T., Yukawa, Y., Miyamoto, T., Obokata, J., and Sugiura, M. (2003). Identification of RNA editing sites in chloroplast transcripts from the maternal and paternal progenitors of tobacco (*Nicotiana tabacum*): comparative analysis shows the involvement of distinct trans-factors for ndhB editing. *Mol. Biol. Evol.* 20, 1028–1035. doi: 10.1093/molbev/msg098
- Sato, T., Okamoto, J., Degawa, Y., Matsunari, S., Takahashi, K., and Tomioka, K. (2009). White rust of *Ipomoea* caused by *Albugo ipomoeae-panduratae* and *A. ipomoeae-hardwickii* and their host specificity. *J. Gen. Plant Pathol.* 75, 46–51. doi: 10.1007/s10327-008-0142-0
- Semagn, K., Björnstad, Å., and Ndjiondjop, M. (2006). An overview of molecular marker methods for plants. *Afr. J. Biotechnol.* 5, 2540–2568.
- Shaw, J., Lickey, E. B., Schilling, E. E., and Small, R. L. (2007). Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: the tortoise and the hare III. *Am. J. Bot.* 94, 275–288. doi: 10.3732/ajb.94.3.275
- Simms, E. L., and Rausher, M. D. (1993). Patterns of selection on phytophage resistance in *ipomoea purpurea*. *Evolution* 47, 970–976. doi: 10.1111/j.1558-5646.1993.tb01252.x
- Simonsen, A. K., and Stinchcombe, J. R. (2007). Induced responses in *Ipomoea hederacea*: simulated mammalian herbivory induces resistance and susceptibility to insect herbivores. *Arthropod Plant Interact.* 1, 129–136. doi: 10.1007/s11829-007-9009-0
- Song, Y., Dong, W., Liu, B., Xu, C., Yao, X., Gao, J., et al. (2015). Comparative analysis of complete chloroplast genome sequences of two tropical trees *Machilus yunnanensis* and *Machilus balansae* in the family Lauraceae. *Front. Plant Sci.* 6:662. doi: 10.3389/fpls.2015.00662

- Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313. doi: 10.1093/bioinformatics/btu033
- Stefanovia, S., Austin, D. F., and Olmstead, R. G. (2003). Classification of Convolvulaceae: a phylogenetic approach. *Syst. Bot.* 28, 791–806.
- Stefanovia, S., Krueger, L., and Olmstead, R. G. (2002). Monophyly of the Convolvulaceae and circumscription of their major lineages based on DNA sequences of multiple chloroplast loci. *Am. J. Bot.* 89, 1510–1522. doi: 10.3732/ajb.89.9.1510
- Streisfeld, M. A., and Rausher, M. D. (2009). Genetic changes contributing to the parallel evolution of red floral pigmentation among *Ipomoea* species. *New Phytol.* 183, 751–763. doi: 10.1111/j.1469-8137.2009.02929.x
- Sucher, N. J., and Carles, M. C. (2008). Genome-based approaches to the authentication of medicinal plants. *Planta Med.* 74, 603–623. doi: 10.1055/s-2008-1074517
- Suyama, M., Torrents, D., and Bork, P. (2006). PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res.* 34, W609–W612. doi: 10.1093/nar/gkl315
- Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013). MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* 30, 2725–2729. doi: 10.1093/molbev/mst197
- The Angiosperm Phylogeny Group (2016). An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. *Bot. J. Linn. Soc.* 181, 1–20. doi: 10.1111/boj.12385
- Thiel, T. (2003). MISA—*Microsatellite Identification Tool*. Available at: <http://pgrc.ipk-gatersleben.de/misa/>
- Tiffin, P., and Rausher, M. D. (1999). Genetic constraints and selection acting on tolerance to herbivory in the common morning glory *Ipomoea purpurea*. *Am. Nat.* 154, 700–716. doi: 10.1086/303271
- Tillich, M., Lehwark, P., Pellizzer, T., Ulbricht-Jones, E. S., Fischer, A., Bock, R., et al. (2017). GeSeq - versatile and accurate annotation of organelle genomes. *Nucleic Acids Res.* 45, W6–W11. doi: 10.1093/nar/gkx391
- Wang, Y., Zhan, D.-F., Jia, X., Mei, W.-L., Dai, H.-F., Chen, X.-T., et al. (2016). Complete chloroplast genome sequence of *Aquilaria sinensis* (Lour.) Gilg and evolution analysis within the malvales order. *Front. Plant Sci.* 7:280. doi: 10.3389/fpls.2016.00280
- Warburton, P. E., Giordano, J., Cheung, F., Gelfand, Y., and Benson, G. (2004). Inverted repeat structure of the human genome: the X-chromosome contains a preponderance of large, highly homologous inverted repeats that contain testes genes. *Genome Res.* 14, 1861–1869. doi: 10.1101/gr.2542904
- Wicke, S., Muller, K. F., de Pamphilis, C. W., Quandt, D., Wickett, N. J., Zhang, Y., et al. (2013). Mechanisms of functional and physical genome reduction in photosynthetic and nonphotosynthetic parasitic plants of the Broomrape family. *Plant Cell* 25, 3711–3725. doi: 10.1105/tpc.113.11.3373
- Wicke, S., Schneeweiss, G. M., dePamphilis, C. W., Muller, K. F., and Quandt, D. (2011). The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. *Plant Mol. Biol.* 76, 273–297. doi: 10.1007/s11103-011-9762-4
- Wilkin, P. (1999). A morphological cladistic analysis of the Ipomoeae (Convolvulaceae). *Kew Bull.* 853–876. doi: 10.3732/ajb.89.9.1510
- Wu, F. H., Chan, M. T., Liao, D. C., Hsu, C. T., Lee, Y. W., Daniell, H., et al. (2010). Complete chloroplast genome of *Oncidium Gower Ramsey* and evaluation of molecular markers for identification and breeding in *Oncidiinae*. *BMC Plant Biol.* 10:68. doi: 10.1186/1471-2229-10-68
- Xu, C., Dong, W., Li, W., Lu, Y., Xie, X., Jin, X., et al. (2017). Comparative analysis of six *Lagerstroemia* complete chloroplast genomes. *Front. Plant Sci.* 8:15. doi: 10.3389/fpls.2017.00015
- Yan, L., Lai, X., Li, X., Wei, C., Tan, X., and Zhang, Y. (2015). Analyses of the complete genome and gene expression of chloroplast of sweet potato [*Ipomoea batata*]. *PLoS One* 10:e0124083. doi: 10.1371/journal.pone.0124083
- Yang, J., Gu, H., and Yang, Z. (2004). Likelihood analysis of the chalcone synthase genes suggests the role of positive selection in morning glories (*Ipomoea*). *J. Mol. Evol.* 58, 54–63. doi: 10.1007/s00239-003-2525-3
- Yang, Y., Zhou, T., Duan, D., Yang, J., Feng, L., and Zhao, G. (2016). Comparative analysis of the complete chloroplast genomes of five *Quercus* species. *Front. Plant Sci.* 7:959. doi: 10.3389/fpls.2016.00959
- Yukawa, M., Tsudzuki, T., and Sugiura, M. (2006). The chloroplast genome of *Nicotiana sylvestris* and *Nicotiana tomentosiformis*: complete sequencing confirms that the *Nicotiana sylvestris* progenitor is the maternal genome donor of *Nicotiana tabacum*. *Mol. Genet. Genomics* 275, 367–373. doi: 10.1007/s00438-005-0092-6
- Zuo, L.-H., Shang, A.-Q., Zhang, S., Yu, X.-Y., Ren, Y.-C., Yang, M.-S., et al. (2017). The first complete chloroplast genome sequences of *Ulmus* species by de novo sequencing: genome comparative and taxonomic position analysis. *PLoS One* 12:e0171264. doi: 10.1371/journal.pone.0171264

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer AZ and the handling Editor declared their shared affiliation.

Copyright © 2018 Park, Yang, Kim, Noh, Lee and Moon. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.