Saudi Journal of Biological Sciences 27 (2020) 3489–3498

Contents lists available at ScienceDirect

Saudi Journal of Biological Sciences

journal homepage: www.sciencedirect.com

Original article

Plastome comparison and evolution within the tribes of Plantaginaceae: Insights from an Asian gypsyweed

Satish Maurya [a], Ashwini M. Darshetkar [a], Dong-Keun Yi [b], Jinki Kim [c], Changyoung Lee [b], M. Ajmal Ali [d], Sangho Choi [b,1], Ritesh Kumar Choudhary [a,1], Soo-Yong Kim [b,*]

[a] Biodiversity & Palaeobiology Group, Agharkar Research Institute, Pune 411 004, India
[b] International Biological Material Research Center, Korea Research Institute of Bioscience and Biotechnology, Daejeon 34141, Republic of Korea
[c] Seed Vault Center, Baekdudaegan National Arboretum, Gyeongsangbuk-do 36209, Republic of Korea
[d] Department of Botany and Microbiology, College of Science, King Saud University, Riyadh 11451, Saudi Arabia

ARTICLE INFO

Article history:
Received 17 August 2020
Revised 17 September 2020
Accepted 20 September 2020
Available online 25 September 2020

Keywords:
Antirrhinum majus
Nucleotide diversity
Phylogenomics
Plastome evolution
SSRs
Veronica ovata subsp. *kiusiana*

ABSTRACT

In spite of availability of several plastomes representing different tribes of Plantaginaceae, sparse attempts have been made to understand the plastome structure, evolution, and phylogenomics. In the present study, we have made an effort to understand the gene content and plastome evolution in the family Plantaginaceae using the newly generated plastome sequence of *Veronica ovata* subsp. *kiusiana*, a taxon native to SE Asia. In the first-ever attempt, plastomes of seven out of 10 tribes of Plantaginaceae have been compared to understand the evolution across the tribes of Plantaginaceae. The size of the plastome of *V. ovata* subsp. *kiusiana* is 152,249 bp, showing a typical quadripartite structure containing LSC, SSC, and two IRs with the sizes of 83,187, 17,704, and 25,679 respectively. The plastome comparison revealed the unique deletions in *ycf2* and *ndhF* genes of members of different tribes, and also revealed high nucleotide variable hotspots. The study also revealed six highly variable genes and intergenic spacer viz. *rps16*, *rps15-ycf1*, *ccsA-ndhD*, *ndhC-trnV*, *petN-psbM*, and *ycf1-trnN* as potential DNA barcodes for the genus *Veronica*. The phylogenomic study revealed the sister relationship between *V. ovata* subsp. *kiusiana* and *V. persica* and also suggested the tentative placement of seven tribes in the family Plantaginaceae.

1. Introduction

The family Plantaginaceae (formerly Veronicaceae) is divided into 10 tribes viz., Plantagineae, Angelonieae, Gratioleae, Antirrhineae, Callitricheae, Sibthorpieae, Digitalideae, Cheloneae, Veroniceae, and Russelieae (Stevens, 2001) and considered to be a monophyletic group closely allied to the Scrophulariaceae (Rahn, 1996; Olmstead et al., 2001; Stevens, 2001; Albach et al., 2005a). Earlier studies on Plantaginaceae (Rahn, 1996) found it difficult to find a sister group for the monophyletic Plantaginaceae which often appeared as sister to Buddlejaceae, Oleaceae, Globulariaceae, and Lentibulariaceae. Plantaginaceae was further separated from its allied families based on advanced morphological characters like phyllodial and parallel-veined leaves, presence of hairs in the leaf axil, protogynous flowers, scarious corolla, absence of a disc, etc. and it was considered to be a monogeneric family with the only genus *Plantago* L. divided into six subgenera. Later, The Angiosperm Phylogeny Group, (1998) and Olmstead et al. (2001) revealed that the family is not monogeneric and segregated various genera under the family Plantaginaceae. Olmstead et al. (2001) also revealed the presence of a total of nine tribes in family Plantaginaceae (Veronicaceae group), however, the tribal placement was not confirmed due to poor support, polytomy and paraphyletic placement of few tribes. Albach et al. (2005a) further classified Plantaginaceae into 12 different tribes quoting the family highly heterogeneous which includes plants ranging from aquatic to alpine habitats. Lately, APG-IV (Chase et al., 2016), reclassified Plantaginaceae with 10 tribes. None of the earlier studies, however, made efforts to understand the phylogenetic placement of several

* Corresponding author.
E-mail addresses: decoy0@kribb.re.kr (S. Choi), rkchoudhary@aripune.org (R.K. Choudhary), soodole@kribb.re.kr (S.-Y. Kim).
[1] Co-corresponding author.

Peer review under responsibility of King Saud University.

Production and hosting by Elsevier

https://doi.org/10.1016/j.sjbs.2020.09.040
1319-562X/

genera which remained unresolved. Very few studies used the molecular markers to resolve the complexity within the family Plantaginaceae (Olmstead et al., 2001; Albach et al., 2005a & b; Estes & Small, 2008; Refulio-Rodriguez & Olmstead, 2014; Yousefi et al., 2016). Olmstead et al. (2001) carried out molecular systematic study of the family Scrophulariaceae based on three chloroplast markers viz. *rbcL*, *ndhF*, and *rps2*. They also considered Plantaginaceae (formerly Veronicaceae) in the study. Albach et al. (2005a) analyzed the relationships within Plantaginaceae. They circumscribed the family into 12 tribes based on four nuclear and plastid markers. Albach et al. (2005b) attempted to unravel the diversification of tribe Veroniceae in New Guinea using the markers ITS and *trnL-F*. Yousefi et al. (2016) tried to resolve the relationships within the tribe Antirrhinae for the Flora Iranica region based on markers ITS and *rpl32-trnL*. Moreover, the efficacy of the markers was never tested in any of the phylogenetic studies. Recently, Choi et al. (2016) reported plastome sequences of two species of *Veronica* viz. *V. nakaiana* Ohwi and *V. persica* Poir. The study also reported plastome sequence of *Veronicastrum sibiricum* (L.) Pennell, an allied genus of *Veronica*.

In the present study, we have sequenced the plastome of *Veronica ovata* subsp. *kiusiana* (Furumi) Albach a southeast Asian endemic species and compared it with the *V. nakaiana* Ohwi and *V. persica* Poir. plastomes to understand their nucleotide diversity and codon usage. Further, the data was used to compare the plastome structure and evolution across the tribes of Plantaginaceae.

2. Material and methods

2.1. Sampling, DNA extraction, and sequencing

The voucher specimens of *Veronica ovata* subsp. *kiusiana* were collected from Seorak Mountain, Yangyang-gun, Gangwon Province of the Republic of Korea in July 2014. The specimens were deposited at the Herbarium of Korea Research Institute of Bioscience and Biotechnology (Voucher number KRIB 0004193). Seeds from the same specimen were germinated and DNA was extracted from fresh young leaves. DNA extraction was carried out using the DNeasy Plant Mini Kit (QIAGEN, Cat. No. 69104) according to the manufacturer's protocol.

TruSeq DNA PCR-Free Library Prep Kit (Illumina) was used for library preparation. Total genomic DNA was sheared and short inserts (550 bp) paired-end libraries prepared by adding unique barcodes. After the library preparation, the DNA sample was run in a single lane of an Illumina HiSeq 10X with a read length of 151 bp. The final raw data output was 13.72 GB including both forward and reverse reads.

2.2. Plastome assembly and annotation

The raw reads obtained after Illumina sequencing were analyzed using FastQC V0.11.7 (Andrews, 2010) software to ensure the quality of reads and Phred score. The adaptors and low-quality bases were trimmed by using Trimmomatic V0.38 (Bolger et al., 2014). The high-quality reads without adaptors were extracted and were used for the final assembly of the plastome. Both forward and reverse reads with a read length of 150 bp and insert size of 300 bp, with seed sequence were imported in NOVOPlasty V3.8.2 (Dierckxsens et al., 2017). The De-novo as well as the reference-based assembly was performed to confirm the accuracy of the assembled plastome. The closest allied reference of *Veronica nakaiana* (KT633216) was selected for reference-based assembly. The assembly and orientation of Inverted Repeats (IRs), Large Single Copy (LSC), and Small Single Copy (SSC) regions were confirmed by NCBI blast and graphic view using Geneious prime 2020.1.2 (https://www.geneious.com). The assembled plastome was annotated using Geneious prime 2020.1.2 (https://www.geneious.com) and GeSeq (Tillich et al., 2017, https://chlorobox.mpimp-golm.mpg.de/geseq.html). Also, *Veronica nakaiana* was used as a reference for the annotation of the plastome. The circular map of the plastome was constructed using OGDRAW (Lohse et al., 2013).

2.3. Genome comparison

Representative plastomes of seven tribes of Plantaginaceae were considered and aligned using a Geneious prime 2020.1.2 plugin MAFFT v7.450 (Katoh and Standley, 2013). The representative plastomes of seven tribes were considered and presented in Table 1. All the sequences were available on the NCBI database except *V. ovata* subsp. *kiusiana* (generated during this study) and *Antirrhinum majus* L. which was received from the Chinese Academy of Sciences, Beijing, China (Courtesy: Snapdragon Genome Database-©Yongbiao, 2020). The compared plastomes were analyzed for conserved and variable regions using MultiPIPMaker (http://pipmaker.bx.psu.edu/pipmaker/) (Schwartz et al., 2000).

2.4. Simple sequence repeats (SSRs) and repeat analysis

The assembled plastome of *Veronica ovata* subsp. *kiusiana* was used to find out the Simple Sequence Repeats (SSRs) by using the online server MISA (Beier et al., 2017) (http://webblast.ipk-gatersleben.de/misa/index.php?action= 1). The threshold for mono, di, tri, tetra, penta, and hexa nucleotides was kept as 10, 5, 4, 3, 3, and 3 respectively. The number, position, and size of SSRs were compared with other Plantaginaceae members. While REPuter (Kurtz et al., 2001) (https://bibiserv2.cebitec.uni-bielefeld.de/reputer), an online server was used to detect the position of forward, reverse, complement, and palindromic repeat sequences of *V. ovata* subsp. *kiusiana* in plastome. The threshold for repeat length was set to 30 with more than 90% similarity and hamming distance was set to 3. The repeat sequences of *V. ovata* subsp. *kiusiana* was compared with the other Plantaginaceae members.

2.5. Nucleotide diversity and interspecific variation in Veronica

The Single Nucleotide Polymorphisms (SNPs), and codon usages were analyzed separately using DnaSP v6.12.01 software (Rozas et al., 2017) for three plastomes of *Veronica*. Three plastomes of *Veronica* spp. were aligned using MAFFT v7.450 (Katoh and Standley, 2013). The aligned sequences were imported in DnaSP v6.12.01 software to perform Nucleotide diversity (Pi) and sliding window analysis.

2.6. Codon usage

The Protein-coding genes of three species of *Veronica* viz. *V. ovata* subsp. *kiusiana*, *V. nakaiana*, and *V. persica* were imported in DnaSP v6.12.01 software (Rozas et al., 2017) and were analyzed separately. The percent codon usage was further confirmed using Geneious Prime 2020.1.2 (https://www.geneious.com). To examine the frequency and uniformity of Synonymous codon and codon biases, the Relative Synonymous Codon Usage (RSCU) was also determined in DnaSP v6.12.01 software (Rozas et al., 2017).

2.7. Phylogenetic analysis

To understand the phylogenetic placement of *Veronica ovata* subsp. *kiusiana* in Plantaginaceae, samples were selected based on its allied genera and tribes from the family. The ingroup and

Table 1
Representative species and tribes included in this study.

No.	Genus	Species	Tribe	GenBank Accession No.
1.	*Veronica*	*Veronica ovata* subsp. *kiusiana*	Veroniceae	MT671999
		Veronica nakaiana		NC031153
		Veronica persica		NC031344
2.	*Veronicastrum*	*Veronicastrum sibiricum* (L.) Pennell	Veroniceae	NC031345
3.	*Hemiphragma*	*Hemiphragma heterophyllum* Wall.	Sibthorpieae	NC045398
4.	*Antirrhinum*	*Antirrhinum majus* L.	Antirrhineae	http://bioinfo.sibs.ac.cn/Am/.
5.	*Penstemon*	*Penstemon fruticosus* (Pursh) Greene	Cheloneae	MG201976
6.	*Bacopa*	*Bacopa monnieri* (L.) Wettst.	Gratioleae	NC047469
7.	*Digitalis*	*Digitalis lanata* Ehrh.	Digitalideae	NC034688
8.	*Plantago*	*Plantago ovata* Forssk.	Plantagineae	MH165324

outgroup consisted of 14 and three taxa respectively from the closely allied Scrophulariaceae family. A total of 17 plastomes were aligned using a Geneious prime 2020.1.2 plugin MAFFT v7.450 (Katoh and Standley, 2013). The aligned sequences were further used for phylogenetic tree construction using the Maximum Likelihood (ML) method in IQtree 1.6.12-MacOSX (Nguyen et al., 2015). The best fit model was tested using the IQtree Model test pipeline, and GTR + F + R4 was selected as the best fit model using the Bayesian Information Criterion (BIC). To construct the phylogenetic tree, 1000 Bootstrap replications were used of which the opti-

Fig. 1. Plastome map of *Veronica ovata* subsp. *kiusiana*. Genes drawn inside the circle are transcribed clockwise and those outside are counter-clockwise. Genes belonging to different functional groups are shown in different colors. The innermost circle denotes GC content across the plastome.

Table 2
Comparison of major features of *V. ovata* subsp. *kiusiana* plastome with allied Plantaginaceae members.

Plastome feature	*Veronica ovata* subsp. *kiusiana*	*Veronica nakaiana*	*Veronica persica*	*Veronicastrum sibiricum*	*Hemiphragma heterophyllum*	*Antirrhinum majus*	*Penstemon fruticosus*	*Bacopa monnieri*	*Digitalis lanata*	*Plantago ovata*
Genome length	152,249	152,319	150,198	152,930	152,700	152,690	152,704	152,495	153,108	162,116
% GC content	38	37.9	37.9	38.3	38.1	37.9	37.9	37.6	38.6	38.2
LSC	83,187	83,195	81,850	83,615	83,243	85,045	83,784	83,764	83,934	82,084
SSC	17,704	17,702	17,418	17,801	17,831	17,929	17,823	17,395	17,688	5272
IRs	25,679	25,711	25,465	25,757	25,808	24,858	25,549	25,668	25,743	37,380
Genes										
number of genes	130	133	130	131	133	132	133	132	130	149
CDS	85	88	86	86	86	87	87	88	85	98
tRNA	37	37	36	37	39	37	38	36	37	43
rRNA	8	8	8	8	8	8	8	8	8	8
% GC LSC	36.1	36.0	36.0	36.5	36.2	36.0	36.0	35.5	36.8	36.9
% GC SSC	32.0	31.7	31.6	32.3	32.1	32.0	31.7	31.4	32.7	31.7
% GC IR	43.2	43.2	43.2	43.3	43.2	42.9	43.2	42.8	43.4	40.0
No. of genes duplicated in IR	19	19	19	19	19	21	19	19	19	31
No. of genes having introns	24	24	24	24	24	24	24	24	24	28

Table 3
Gene composition of the chloroplast genome of *Veronica ovata* subsp. *kiusiana*.

Category	Group	Name
Photosynthesis-related genes	Rubisco	*rbcL*
	Photosystem 1	*psaA, psaB, psaC, psaI, psaJ*
	Photosystem 2	*psbA, psbB, psbC, psbD, psbE, psbF, psbH, psbI, psbJ, psbK, psbL, psbM, psbN, psbT, psbZ*
	APT synthase	*atpA, atpB, atpE, atpF†, atpH, atpI*
	Cytochrome *b*/f complex	*petA, petB, petD, petG, petL, petN*
	NADPH Dehydrogenase	*ndhA†, ndhB†*, ndhC, ndhD, ndhE, ndhF, ndhG, ndhH, ndhI, ndhJ, ndhK*
Transcription and translation related genes	Transcription	*rpoA, rpoB, rpoC1†, rpoC2*
	Ribosomal proteins	*rps2, prs3, rps4, rps7*, rps8, rps11, rps12†*, rps14, rps15, rps16†, rps18, rps19 rpl2†*, rpl14, rpl16, rpl20, rpl22, rpl23*, rpl33, rpl36*
	Translation initiation factor	*infA*
RNA genes	Ribosomal RNA	*rrn5*, rrn4.5*, rrn16*, rrn23**
	Transfer RNA	*trnA-UGC†*, trnC-ACA†*, trnC-GCA, trnD-GUC, trnE-UUC, trnF-GAA, trnfM-CAU, trnG-UCC*, trnH-AUG†, trnI-GAU†*, trnK-UUU†, trnL-CAA*, trnL-UAA, trnL-UAG†, trnM-CAU*, trnN-GUU†*, trnP-GGG, trnP-UGG, trnQ-UUG, trnR-ACG*, trnR-UCU, trnS-GCU, trnS-GGA, trnS-UGA†, trnT-GGU, trnT-UGU, trnV-GAC*, trnW-CCA, trnY-GUA*
Other genes	RNA processing	*matK*
	Carbon metabolism	*cemA*
	Fatty acid synthesis	*accD*
	Proteolysis	*clpP*[†]
Genes of unknown function	Conserved reading frame	*ycf1, ycf2, ycf3†, ycf4, ycf15*

† Genes with Intron.
* duplicated genes.

Table 4
Details of the deletions (in bp) shown by MultiPIP analysis in various taxa of Plantaginaceae.

Taxa name	Deletion		
	gene/spacer region	bp deletion	Position in plastome
Veronica persica	*atpH-atpI*	241	LSC
	psbE-petL	289	LSC
Antirrhinum majus	*trnC GCA-petN*	470	LSC
	rpl32-ccsA	278	SSC
Bacopa monnieri	*matk-rps16*	399	LSC
	rpoB-petN	351	LSC
	rpl32-ccsA	664	SSC
Plantago ovata	*trnT GGU-psbD*	283	LSC
	ycf3-rps4	592	LSC
	ndhC-trnV AUC	765	LSC
	accD	425	LSC
	clpP	337	LSC
	ycf2	1124	IRa
	ycf2	1124	IRb

mal log-likelihood value −460554.508 representing the consensus tree was selected. To visualize the output of IQtree, the tree file was imported in Figtree V1.4.4 (https://github.com/rambaut/figtree/releases) (Rambaut, 2018).

3. Results

3.1. Plastome assembly

The Illumina sequencing generated 36,725,388 reads. The average organelle coverage was 1,527X. The final trimmed reads obtained from Trimmomatic V0.38 were assembled in NOVOPlasty which resulted in two circular contigs. Both were manually checked and mapped with the reference sequence in Geneious Prime 2020.1.2. The final assembled plastome length was 152,249 bp.

3.2. Plastome characteristics

The plastome of *Veronica ovata* subsp. *kiusiana* exhibits a typical plastome structure, having LSC (Large Single Copy), SSC (Small Single Copy), and a pair of IRs (Inverted Repeats). The total length of the plastome is 152,249 bp while that of LSC, SSC, and IR regions is 83,187, 17,704, and 25,679 bp respectively (Fig. 1, Table 2). The GC content of the whole plastome is 38%. The plastome sequence generated has been deposited to NCBI (Accession no.

MT671999). The plastome of *Veronica ovata* subsp. *kiusiana* is characterized by the presence of a total of 130 genes, comprising 85 CDS, 37 tRNAs, and 8 rRNAs (Table 2). The list of genes is presented in Table 3.

The genome alignment using MultiPIP resulted in the highly conserved plastomes among the seven tribes of Plantaginaceae except for few variable regions in *Veronica persica*, *Antirrhinum majus*, *Bacopa monnieri,* and *Plantago ovata* (Table 4, Fig. 2). *Veronica persica*, *Antirrhinum majus*, and *Bacopa monnieri* exhibited deletions in intergenic spacer of various genes, whereas *Plantago ovata* revealed deletions in both intergenic spacers as well as in three genes (Table 4).

3.3. Repeat and SSR analysis

In the present study, we determined SSRs of *Veronica ovata* subsp. *kiusiana* as well as nine other Plantaginaceae members. The type, distribution, and size of SSRs were examined. Based on the SSR analysis, a total of 47 microsatellites were detected. Six compounds, 29 mono, six di, two tri, and four tetra-nucleotide repeats were detected. However, no penta, or *hexa*-nucleotide repeats were detected in the plastome of *V. ovata* subsp. *kiusiana* (Fig. 3a). 43, 1, and 2 SSRs were found in the LSC, IR, and SSC regions respectively (Fig. 3b). All the SSRs found in *V. ovata* subsp. *kiusiana* were AT-rich. The size of the SSRs ranged from 10 to 121 bp (Fig. 4).

The number of SSRs in the compared plastomes varied from 14 (*Antirrhinum majus*) to 57 (*Hemiphragma heterophyllum*). Also, the highest numbers of SSRs were located in the LSC region followed by SSC and IRs in all the compared plastomes (Fig. 3a & b).

The reputer analysis yielded a total of 34 repeat sequences in the plastome of *Veronica ovata* subsp. *kiusiana*. 16 forward, 17 palindromic, and one reverse repeat were detected, however, no complement type of repeats was observed. Other compared plastomes also exhibited a high number of forward and palindromic repeats and a low number of reverse and complement repeats (Fig. 5a). Most of the repeats ranged in the size of 30–49 in all compared plastomes except in *Plantago* and *Antirrhinum* where the repeat sizes varied from 30 to 8000 (Fig. 5b).

3.4. Nucleotide diversity and interspecific variation in Veronica

The alignment of the plastomes of *Veronica ovata* subsp. *kiusiana*, *V. nakaiana,* and *V. persica* was also studied. 4979 SNPs and 5255 indel mutations were observed. Sliding window analysis between the three plastomes exhibited some hotspot regions. High nucleotide diversity and Pi distance was observed in two genes and four intergenic spacer regions viz., *rps16*, *rps15-ycf1*, *ccsA-ndhD*, *ndhC-trnV*, *petN-psbM*, and *ycf1-trnN* (Fig. 6).

3.5. Codon usage

The three *Veronica* plastomes were compared for their codon usage. The plastomes of *Veronica ovata* subsp. *kiusiana*, *V. nakaiana*, and *V. persica* are characterized by 26050, 26791, and 26,528 codons and 78150, 80373, and 79,584 nucleotides respectively. The most abundant amino acid is Leucine while the least abundant is Cysteine in all the three compared plastomes (Fig. 7). Codon usage is biased towards A and T in all the three plastomes. The high codon preference (RSCU) is 1.78, 1.82, and 1.90 while the low codon preference is 0.46, 0.31, and 0.31 in the plastomes of *V. ovata* subsp. *kiusiana*, *V. nakaiana,* and *V. persica* respectively (Supplementary Table 1).

3.6. Phylogenomic analysis

This is the first-ever study to understand relationships between tribes of Plantaginaceae using plastome data. The phylogenomic analysis included representation of seven tribes of Planataginaceae viz., Plantagineae, Digitalideae, Veroniceae, Sibthorpieae, Antirrhineae, Cheloneae, and Gratioleae. *Veronica ovata* subsp. *kiusiana* appeared sister to *V. nakaiana* (BP = 100/100) (Fig. 8). *V. persica* appeared sister to the clade of *V. ovata* subsp. *kiusiana* and *V. nakaiana*. Tribes Plantagineae and Digitalideae appeared sister (BP = 99.6/98) while tribes Veroniceae and Sibthorpieae appeared sister (BP = 100/100). Single representatives are available for the tribes Antirrhineae, Cheloneae, and Gratioleae which appeared sister to the clade of Plantagineae/Digitalideae and Veroniceae/Sibthorpieae (BP = 100/100).

4. Discussion

This is the first-ever study on the characterization and comparison of plastomes of seven tribes of Plantaginaceae. The plastome alignment revealed significant deletions in genes and intergenic spacers of the seven tribes of Plantaginaceae (Table 4, Fig. 2). The genus *Veronica* showed highly conserved plastome and depicted

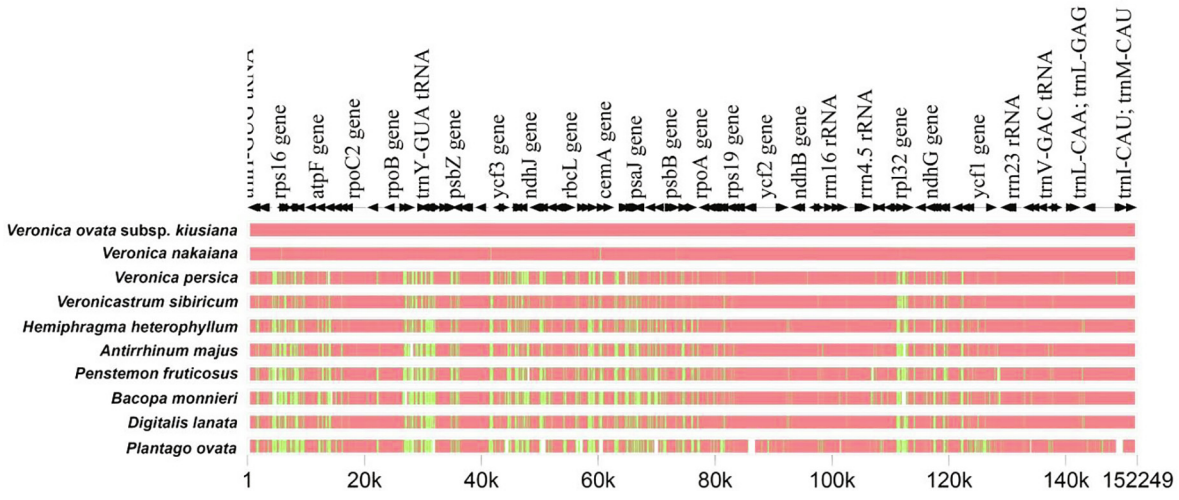


Fig. 2. Plastome alignment of seven tribes of Plantaginaceae using MultiPIP.

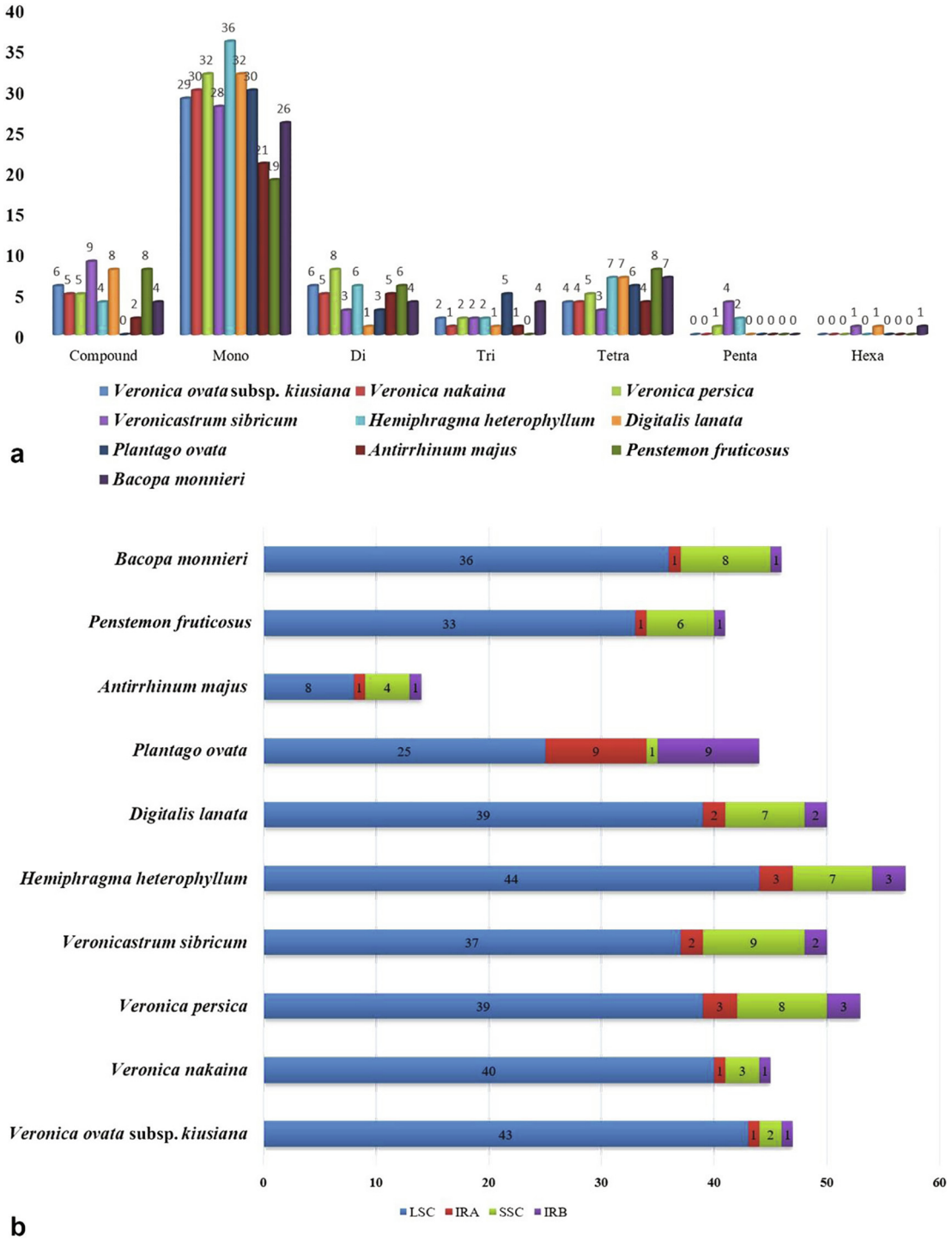


Fig. 3. a. Comparison of types of SSRs across 10 plastomes. b. Comparison of the position of SSRs across 10 plastomes.

few deletions in the intergenic spacers of *Veronica persica*. While *Plantago ovata* possesses high variability among the seven tribes of Plantaginaceae due to the expansion of IRs and genome size (Fig. 2). *Plantago ovata* exhibited huge variations in terms of gene composition, genome size, and rearrangements (Asaf et al., 2020). While we found that there was a 1,124 bp long deletion in *ycf2* gene of *Plantago ovata* and this could be due to the rearrangement and expansion of IRs.

Simple Sequence Repeats (SSRs) are the short stretches of 1–6 nucleotides found in coding as well as non-coding regions of the genome (Schlötterer, 2000; Ellegren, 2004). In this study, we compared the type and size of SSRs across all the seven tribes. The least number of SSRs were reported in *Antirrhinum majus* of tribe Antirrhinae. The compared plastomes exhibited the highest number of mononucleotide repeats while most of the repeats were AT-rich which was in accordance with the earlier available plastomes for the genus (Choi et al., 2016).

The family Plantaginaceae has been studied extensively by various researchers but phylogenetic positions of several genera remained unresolved. Very few studies used the molecular mark-

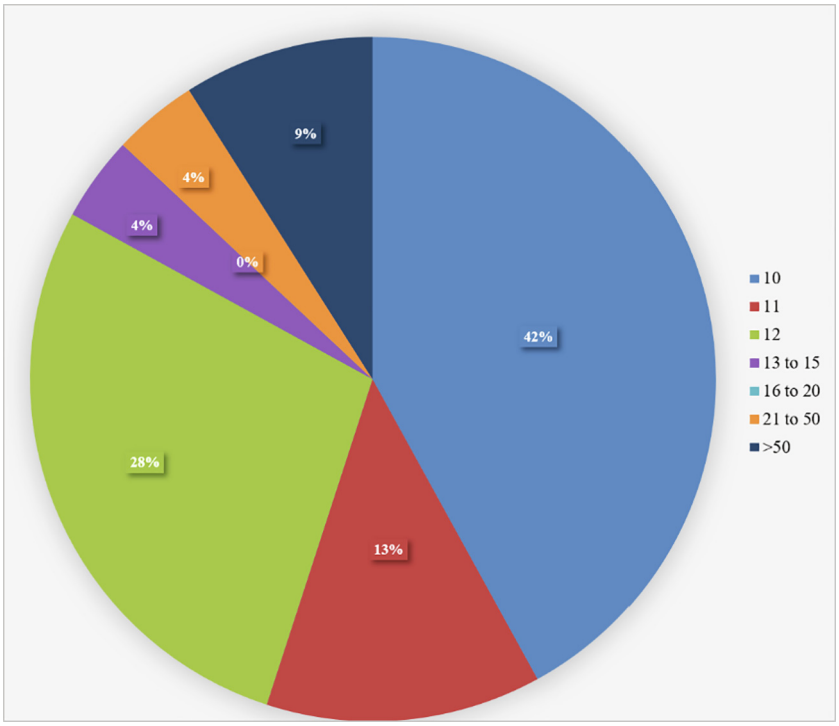


Fig. 4. Size of SSRs in the plastome of *Veronica ovata* subsp. *kiusiana*.

ers (nuclear - ITS, plastid - *atpB*, *matK*, *ndhF*, *psbB*, *rbcL*, *rpl32-trnL*, *rps3*, *rps4*, *rps16*, *trnLF*, *trnK*, and *trnV-atpE*) to resolve the complexity within family Plantaginaceae (Olmstead et al., 2001; Albach et al., 2005ab; Estes & Small, 2008; Refulio-Rodriguez & Olmstead, 2014; Yousefi et al., 2016). Later, Choi et al. (2016) studied the plastome structure and nucleotide diversity in *Veronica nakaiana* and *V. persica* and proposed five probable markers (*trnG-trnM*, *trnT-trnL*, *ycf4-cemA*, *petD-rpoA*, and *rpl2-trnL*). Though, the efficacy of the markers was never tested in any of the phylogenetic study. The makers proposed by Choi et al. (2016), was selected from the 114 highly variable genes (Coding, noncoding regions, and Introns). However, our study used the complete plastome of three species of *Veronica* viz. *V. ovata* subsp. *kiusiana*, *V. nakaiana*, and *V. persica*, and revealed few more high variable regions and mutational sites. Here we propose six highly variable genes and intergenic spacer viz. *rps16*, *rps15-ycf1*, *ccsA-ndhD*, *ndhC-trnV*, *petN-psbM*, and *ycf1-trnN* (Fig. 6). These markers possess high nucleotide diversity and will be helpful in the future to resolve the complexes within genera of family Plantaginaceae. The plastome possesses high variability in LSC, SSC, and followed by IRs (Li & Zheng, 2018), similar was observed in *Veronica*. The study of Olmstead et al. (2001) highlighted the deletion/ insertion of a few bases in *ndhF* gene of *Digitalis*, *Hemiphragma*, *Plantago,* and *Veronica*, which was supported by our study. We observed the deletion of three bases (TTC) at 713 bp to the upstream of 3′ of *ndhF* gene in the tribe Veroniceae, Sibthorpieae, Digitalideae, and Plantagineae. To confirm this deletion in the future, we need robust sampling from all the tribes of Plantaginaceae.

The Leucine and Cysteine were the most and least common amino acids in the plastome of the genus *Veronica*, respectively (Fig. 7). Our study was also supported by the previous findings on the plastomes of *Plantago ovata* (Asaf et al., 2020). The study also supports the presence of AT codon biases towards the third position of codons, similar to most of the Angiosperms (Jiang et al., 2017; Asaf et al., 2020).

Our phylogenetic analysis confirmed the position of *Veronica ovata* subsp. *kiusiana* in the genus *Veronica*, which appeared sister to *V. nakaiana* (Fig. 8). The phylogenetic placement of genus *Veronica* was found to be in congruence with the previous studies (Choi et al., 2016; Asaf et al., 2020). The phylogenetic tree also revealed the placement of seven tribes within the family Plantaginaceae (Fig. 8) but due to the sampling limitation, we could not confirm the phylogenetic position. In the future, the inclusion of robust sampling will confirm the phylogenetic position of all ten tribes in the family Plantaginaceae.

Declaration of Competing Interest

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

Acknowledgements

SM, AMD, and RKC are thankful to the Director, Agharkar Research Institute for facilities. AMD is thankful to the CSIR for SRF fellowship support.

Funding

This research was supported by the grant from the KRIBB Initiative Program of the Republic of Korea and the Bio and Medical Technology Development Program of the National Research Foundation (NRF) and funded by the Korean government (MSIT) (NRF-2016K1A1A8A01939075).

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.sjbs.2020.09.040.

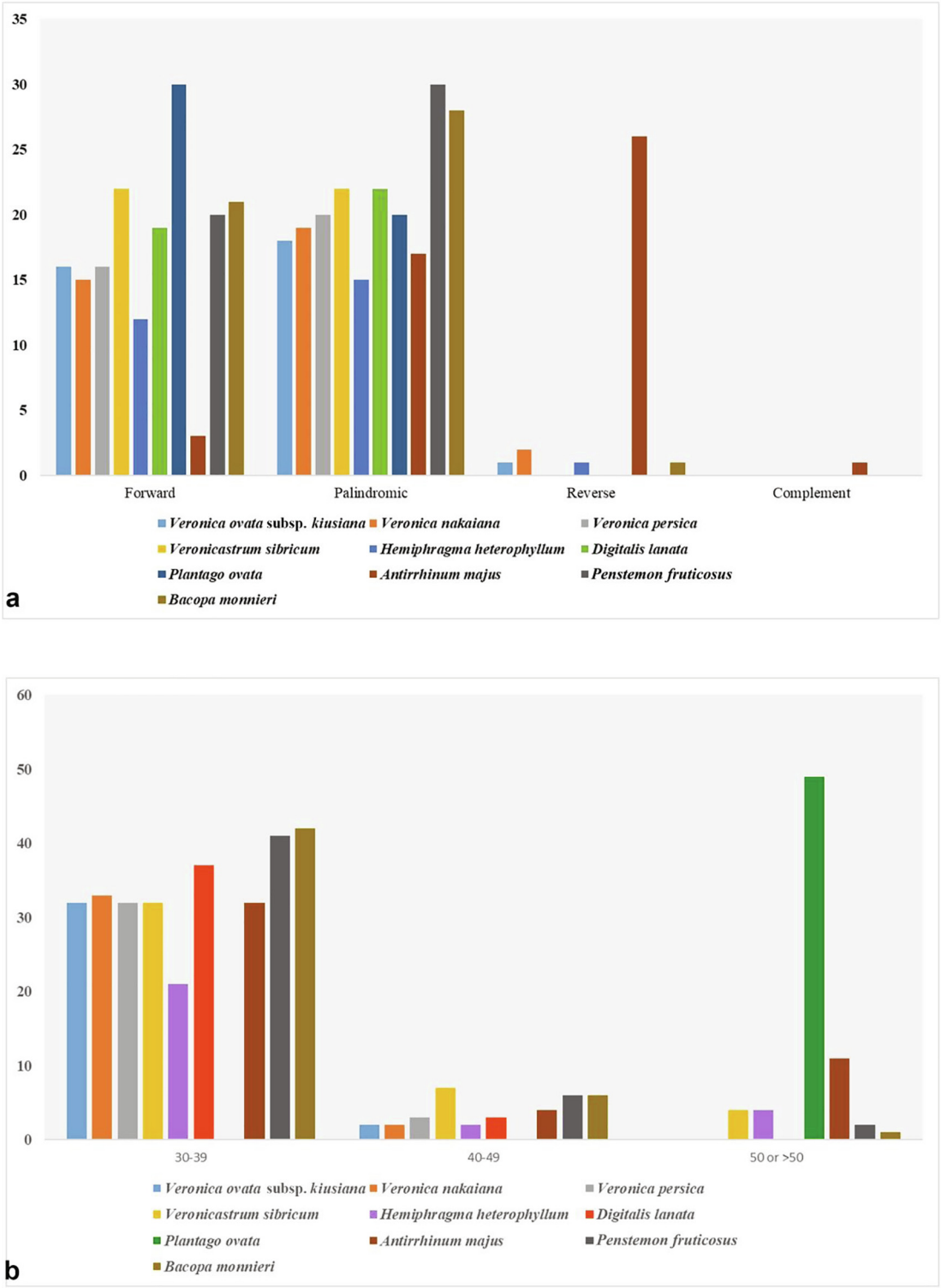


Fig. 5. Comparison of a. types of repeats and b. the number of repeats across 10 compared plastomes.

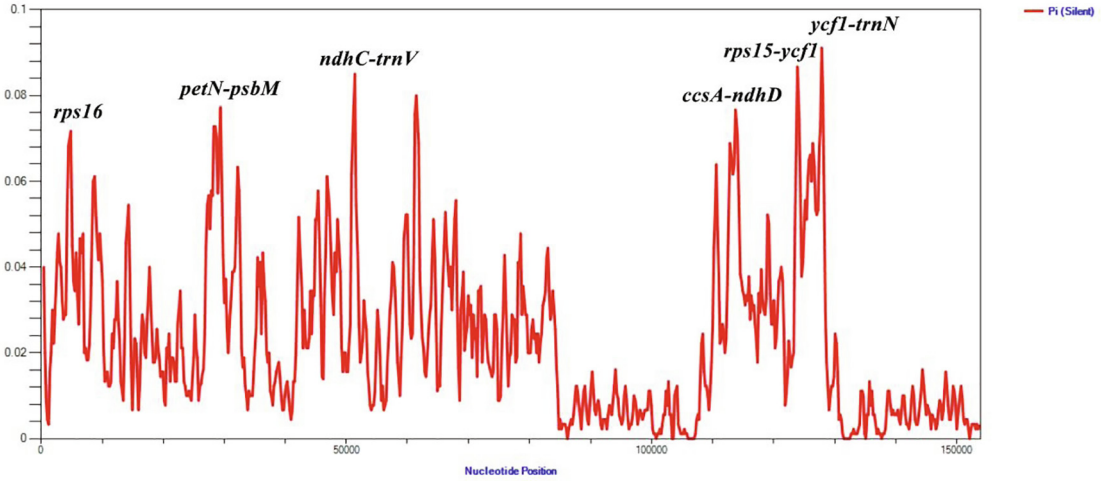
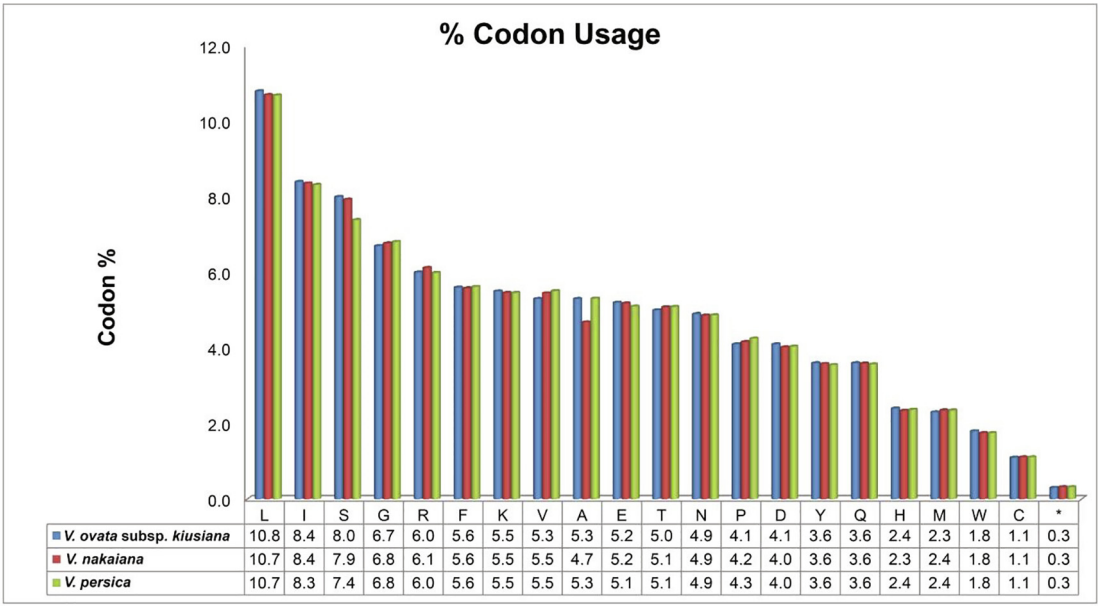


Fig. 6. Nucleotide diversity and hotspot regions between *Veronica* plastomes. The X-axis represents the nucleotide position and Y-axis represents nucleotide diversity (Pi).



	L	I	S	G	R	F	K	V	A	E	T	N	P	D	Y	Q	H	M	W	C	*
V. ovata subsp. *kiusiana*	10.8	8.4	8.0	6.7	6.0	5.6	5.5	5.3	5.3	5.2	5.0	4.9	4.1	4.1	3.6	3.6	2.4	2.3	1.8	1.1	0.3
V. nakaiana	10.7	8.4	7.9	6.8	6.1	5.6	5.5	5.5	4.7	5.2	5.1	4.9	4.2	4.0	3.6	3.6	2.3	2.4	1.8	1.1	0.3
V. persica	10.7	8.3	7.4	6.8	6.0	5.6	5.5	5.5	5.3	5.1	5.1	4.9	4.3	4.0	3.6	3.6	2.4	2.4	1.8	1.1	0.3

Fig. 7. Codon usage of three *Veronica* taxa. X-axis: Amino acid, Y-axis: codon usage in percentage. * indicates the stop codon.

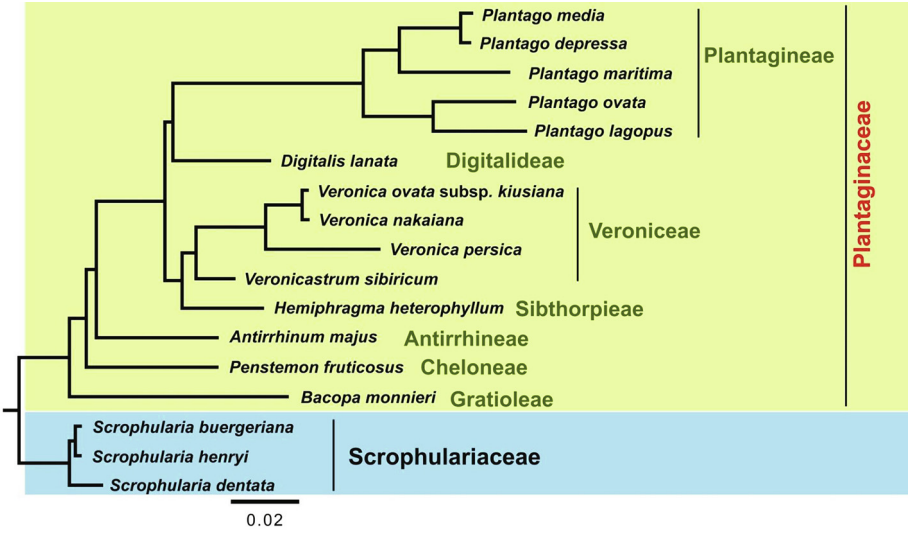


Fig. 8. Phylogenomic tree showing the placement of seven tribes of family Plantaginaceae. All the nodes in the tree have 100% bootstrap support.

References

Albach, D.C., Meudt, H.M., Oxelman, B., 2005a. Piecing together the "new" Plantaginaceae. Am. J. Bot. 92 (2), 297–315. https://doi.org/10.3732/ajb.92.2.297.

Albach, D.C., Utteridge, T., Wagstaff, S.J., 2005b. Origin of Veroniceae (Plantaginaceae, formerly Scrophulariaceae) on New Guinea. Syst. Bot. 30 (2), 412–423. https://doi.org/10.1600/0363644054223666.

Andrews, S., 2010. FastQC: a quality control tool for high throughput sequence data. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Asaf, S., Khan, A.L., Khan, A., Khan, G., Lee, I.J., Al-Harrasi, A., 2020. Expanded inverted repeat region with large scale inversion in the first complete plastid genome sequence of Plantago ovata. Sci. Rep. 10 (1), 1–16. https://doi.org/10.1038/s41598-020-60803-y.

Beier, S., Thiel, T., Münch, T., Scholz, U., Mascher, M., 2017. MISA-web: a web server for microsatellite prediction. Bioinformatics 33 (16), 2583–2585. https://doi.org/10.1093/bioinformatics/btx198.

Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30 (15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170.

Chase, M.W., Christenhusz, M.J.M., Fay, M.F., Byng, J.W., Judd, W.S., Soltis, D.E., Mabberley, D.J., Sennikov, A.N., Soltis, P.S., Stevens, P.F., 2016. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. Botanical Journal of the Linnean Society 181 (1), 1–20. https://doi.org/10.1111/boj.12385.

Choi, K.S., Chung, M.G., Park, S., 2016. The complete chloroplast genome sequences of three Veroniceae species (Plantaginaceae): comparative analysis and highly divergent regions. Front. Plant Sci. 7, 355. https://doi.org/10.3389/fpls.2016.00355.

Dierckxsens, N., Mardulyn, P., Smits, G., 2017. NOVOPlasty: de novo assembly of organelle genomes from whole genome data. Nucleic Acids Res. 45 (4). https://doi.org/10.1093/nar/gkw955.

Ellegren, H., 2004. Microsatellites: simple sequences with complex evolution. Nat. Rev. Genet. 5 (6), 435–445. https://doi.org/10.1038/nrg1348.

Estes, D., Small, R.L., 2008. Phylogenetic relationships of the monotypic genus Amphianthus (Plantaginaceae tribe Gratioleae) inferred from chloroplast DNA sequences. Syst. Bot. 33 (1), 176–182. https://doi.org/10.1600/036364408783887375.

Jiang, D., Zhao, Z., Zhang, T., Zhong, W., Liu, C., Yuan, Q., Huang, L., 2017. The chloroplast genome sequence of *Scutellaria baicalensis* provides insight into intraspecific and interspecific chloroplast genome diversity in *Scutellaria*. Genes 8 (9), 227. https://doi.org/10.3390/genes8090227.

Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. 30 (4), 772–780. https://doi.org/10.1093/molbev/mst010.

Kurtz, S., Choudhuri, J.V., Ohlebusch, E., Schleiermacher, C., Stoye, J., Giegerich, R., 2001. REPuter: the manifold applications of repeat analysis on a genomic scale. Nucleic Acids Res. 29 (22), 4633–4642. https://doi.org/10.1093/nar/29.22.4633.

Li, B., Zheng, Y., 2018. Dynamic evolution and phylogenomic analysis of the chloroplast genome in Schisandraceae. Sci. Rep. 8 (1), 1–11. https://doi.org/10.1038/s41598-018-27453-7.

Lohse, M., Drechsel, O., Kahlau, S., Bock, R., 2013. OrganellarGenomeDRAW—a suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets. Nucleic Acids Res. 41 (W1), W575–W581. https://doi.org/10.1093/nar/gkt289.

Nguyen, L.T., Schmidt, H.A., Von Haeseler, A., Minh, B.Q., 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol. Biol. Evol. 32 (1), 268–274. https://doi.org/10.1093/molbev/msu300.

Olmstead, R.G., de Pamphilis, C.W., Wolfe, A.D., Young, N.D., Elisons, W.J., Reeves, P. A., 2001. Disintegration of the Scrophulariaceae. Am. J. Bot. 88 (2), 348–361. https://doi.org/10.2307/2657024.

Rahn, K., 1996. A phylogenetic study of the Plantaginaceae. Bot. J. Linn. Soc. 120 (2), 145–198. https://doi.org/10.1111/j.1095-8339.1996.tb00484.x.

Rambaut, A., 2018. FigTree v1. 4.2, a graphical viewer of phylogenetic trees. 2014. https://github.com/rambaut/figtree/releases

Refulio-Rodriguez, N.F., Olmstead, R.G., 2014. Phylogeny of lamiidae. Am. J. Bot. 101 (2), 287–299. https://doi.org/10.3732/ajb.1300394.

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

Schlötterer, C., 2000. Evolutionary dynamics of microsatellite DNA. Chromosoma 109 (6), 365–371. https://doi.org/10.1007/s004120000089.

Schwartz, S., Zhang, Z., Frazer, K.A., Smit, A., Riemer, C., Bouck, J., Gibbs, R., Hardison, R., Miller, W., 2000. PipMaker—a web server for aligning two genomic DNA sequences. Genome Res. 10 (4), 577–586. https://doi.org/10.1101/gr.10.4.577.

Stevens, P. F., 2001 onwards. Angiosperm Phylogeny Website. Version 14, July 2017 [and more or less continuously updated since]." will do. http://www.mobot.org/MOBOT/research/APweb/. (accessed 25 July 2020).

The Angiosperm Phylogeny Group, 1998. An Ordinal Classification for the Families of Flowering Plants. Annals of the Missouri Botanical Garden 85 (4), 531–553. https://doi.org/10.2307/2992015.

Tillich, M., Lehwark, P., Pellizzer, T., Ulbricht-Jones, E.S., Fischer, A., Bock, R., Greiner, S., 2017. GeSeq–versatile and accurate annotation of organelle genomes. Nucleic Acids Res. 45 (W1), W6–W11. https://doi.org/10.1093/nar/gkx391.

Yongbiao X., 2020. Snapdragon Genome Database. http://bioinfo.sibs.ac.cn/Am/. (accessed 25 July 2020).

Yousefi, N., Zarre, S., Heubl, G., 2016. Molecular phylogeny of the mainly Mediterranean genera Chaenorhinum, Kickxia and Nanorrhinum (Plantaginaceae, tribe Antirrhineae), with focus on taxa in the Flora Iranica region. Nordic J. Botany 34 (4), 455–463. https://doi.org/10.1111/njb.01000.