

Implications of plastome evolution in the true lilies (monocot order Liliales)

Hoang Dang Khoa Do^{a,d,1}, Changkyun Kim^{a,1}, Mark W. Chase^{b,c}, Joo-Hwan Kim^{a,*}

^a Department of Life Science, Gachon University, Seongnam 13120, Republic of Korea

^b Royal Botanic Gardens, Kew, UK

^c Department of Environment and Agriculture, Curtin University, Bentley, Western Australia, Australia

^d Nguyen Tat Thanh Hi-Tech Institute, Nguyen Tat Thanh University, Ho Chi Minh City, Viet Nam

ARTICLE INFO

Keywords:

Comparative genomics
Mycoheterotrophic species
Liliales
Plastid genome

ABSTRACT

The families of the monocot order Liliales exhibit highly contrasting characteristic of photosynthetic and mycoheterotrophic life histories. Although previous phylogenetic and morphological studies of Liliales have been conducted, they have not examined molecular evolution associated with this contrasting phenomenon. Here, we conduct the first comparative plastome study of all ten families of Liliales using 29 newly sequenced plastid genomes analyzed together with previously published data. We also present a phylogenetic analysis for Liliales of 78 plastid genes combined with 22 genes from all three genomes (nuclear 18S rDNA and *phyC*; 17 plastid genes; and mitochondrial *matR*, *atpA*, and *cob*). Within the newly generated phylogenetic tree of Liliales, we evaluate the ancestral state changes of selected morphological traits in the order. There are no significant differences in plastid genome features among species that show divergent characteristics correlated with family circumscriptions. However, the results clearly differentiate between photosynthetic and mycoheterotrophic taxa of Liliales in terms of genome structure, and gene content and order. The newly sequenced plastid genomes and combined three-genome data revealed Smilacaceae as sister to Liliaceae instead of Philesiaceae and Ripogonaceae. Additionally, we propose a revised familial classification system of Liliales that consists of nine families, considering Ripogonaceae a synonym of Philesiaceae. The ancestral state reconstruction indicated synapomorphies for each family of Liliales, except Liliaceae, Melanthiaceae and Colchicaceae. A taxonomic key for all nine families of Liliales is also provided.

1. Introduction

Liliales sensu Angiosperm Phylogeny Group (2016) consist of approximately 1500 species in ten families which had been classified by various authors. For example, before the molecular era, this order was differently circumscribed by Cronquist (1981) and Dahlgren et al. (1985). As molecular phylogenetic studies progressed, Liliales underwent several revised family circumscriptions (Chase et al., 1995; Rudall et al., 2000; Vinnersten and Bremer, 2001; Fay et al., 2006; Petersen et al., 2012; Kim et al., 2013). Also, several studies to estimate the divergence time of Liliales and component families have been conducted (Vinnersten and Bremer, 2001; Chacón et al., 2012; Mennes et al., 2015; Givnish et al., 2016; Kim and Kim, 2018; Kim et al., 2019). These studies revealed that this order separated from its sister group and was widespread 123 million years ago (Mya) during the breakup of Gondwana. Recently, a broad phylogenomic study of Liliales was published by Givnish et al. (2016), which discussed familial

relationships, origins and ages. However, relationships of Smilacaceae to other families had low support.

Since the plastid genome (ptDNA) was discovered (Chun et al., 1963; Sager and Ishida, 1966; Tewari and Wildman, 1966), many studies have been conducted on angiosperm plastomes, characterizing structure, function and typically maternal inheritance (Bendich, 2004; Daniell et al., 2016; Cusimano and Wicke, 2016; Barnard-Kubow et al., 2017). Deletions, inversions, and duplications have been described and these changes have led to differences in plastid gene organization and sizes (Haberle et al., 2008; Cosner et al., 2004; Kim et al., 2016; Do et al., 2014; Knox, 2014; Do and Kim, 2017; Choi and Choi, 2017). For example, deletion of the inverted repeat (IR) region and inversions and duplication were found in Fabaceae (Fabales) and Melanthiaceae (Do et al., 2014; Kim et al., 2016; Do and Kim, 2017; Choi and Choi, 2017). These structural changes are also informative and reflect infrageneric relationships. For example, the inversion of the *ndhC-rbcL* region has been found in *Trillium* subgenus *Sessilium* (Melanthiaceae; Kim et al.,

* Corresponding author.

E-mail addresses: dhdhkhao@ntt.edu.vn (H.D.K. Do), ckim@gachon.ac.kr (C. Kim), m.chase@kew.org (M.W. Chase), kimjh2009@gachon.ac.kr (J. Kim).

¹ These authors contributed equally in this study.

2016) and also observed in two additional complete plastomes of the subgenus (including *T. cuneatum* (Accession number KR135077) and *T. decumbens* (Accession number KR534612); Schilling et al. unpublished).

Plastid genomes have been targets for studying evolution and phylogeny due to their distinctive features such as small size and highly conserved gene content and order. For example, plastid protein-coding genes were used for resolving relationships in magnoliids, monocotyledons and other angiosperms (Graham and Olmstead, 2000; Jansen et al., 2007; Ma et al., 2014; Goremykin et al., 2015; Barrett et al., 2016; Folk et al., 2017). Also, plastome data have been used for studies of biogeography, molecular evolution, and age estimation (Christin et al., 2014; Marcussen et al., 2015; Meseguer et al., 2015; Roalson and Roberts, 2016; Foster et al., 2017; Pouchon et al., 2018). Previously, several studies have been published on complete plastid genomes for Liliales (Liu et al., 2012; Kim and Kim 2013; Bodin et al., 2013; Do et al., 2013; Do et al., 2014; Nguyen et al., 2015; Huang et al., 2016; Kim et al., 2016; Song et al., 2016; Do and Kim, 2017; Li et al., 2017; Lu et al., 2017). These studies documented genomic events including gene duplications (*Paris*, Melanthiaceae), inversions (*Trillium*, Melanthiaceae) and deletions (Liliaceae, Alstroemeriaceae, and Colchicaceae; Kim and Kim, 2013; Do et al., 2014; Nguyen et al., 2015; Kim et al., 2016). Apart from these findings, gene content of mycoheterotrophic species in Corsiaceae has been described (Mennes et al., 2015); however, complete plastomes of these taxa have not been explored. Recently, studies on Liliales plastomes focused on certain species or groups but did not present an overview of plastome evolutionary history (Song et al., 2017; Lu et al., 2016; Du et al., 2017; Li et al., 2017). In addition to containing differently distributed families (i.e., Liliaceae in the Northern hemisphere, Corsiaceae in the Southern hemisphere, and Petermaniaceae only in Australia), Liliales also include autotrophic and mycoheterotrophic taxa. Different aspects of Liliales (i.e., classification, biogeography, and plastid genome evolution) have been addressed though the overall morphological evolution of all ten families of the Liliales has not been documented, except for pollen (Furness et al., 2015). Therefore, studying plastome and morphological character evolution will contribute to the exploration of diversification in Liliales.

In this study, we present plastid genome sequences of 29 representative species of Liliales. Based on these new data and previously published genomes on NCBI, we 1) explore plastome evolution in Liliales; 2) reconstruct familial relationship in Liliales based on data from all three genomes; and 3) investigate morphological differentiation and familial classification of Liliales with a key to these families.

2. Materials and methods

2.1. Plant sampling, total DNA extraction, and genome sequencing

Fresh leaves of 29 representative taxa covering nine families of Liliales were preserved in silica gel until DNA extraction (Supplementary Table S1). The aerial parts of the mycoheterotrophic *Corsia dispar* and *Arachnitis uniflora* (Corsiaceae) were used. The modified CTAB method was used to extract total DNA (Doyle and Doyle, 1987). The quality of DNA was tested using gel electrophoresis before library preparation and sequencing with HiSeq2000 and MiSeq Systems (Illumina, San Diego, California, USA) which generated the reads of 99 bp, 100 bp, and 301 bp. In some cases, the quality of DNA was insufficient for conducting next-generation sequencing (NGS). For these samples, we applied the long-range PCR method to amplify plastid DNA fragments, using previously published plastomes of Liliales as a guide (i.e., *Lilium longiflorum* - KC968977; *Veratrum patulum* - NC_022715; and *Smilax china* - HM536959). We used Primer3 (Untergasser et al., 2012) to design primer pairs, located in coding regions, that flank regions ranging from 6 to 10 kb. Also, an overlap of at least 100 bp was used to obtain the complete plastid DNA. After amplification, PCR products were purified and combined for NGS.

2.2. Genome assembly, annotation, and comparison

At least 1 GB of raw data was generated for each sample. The average length of raw reads was 301 bp. Before assembly, the raw reads were trimmed to sequences with less than 0.5% chance of error. Two methods were used to assemble plastomes based on the availability of reference sequences. In the families for which a plastome was already available, it was used as the reference for assembly. In the families without available sequences of plastid genome, the protein-coding regions extracted from published studies of Liliales were used as a template for isolating plastid reads from raw data. The isolated reads were then extracted and reassembled using the De novo Assembly option in Geneious ver. 7.1.7 (Kearse et al., 2012). The newly generated sequences were used as references for reassembling raw reads. These steps were repeated until complete plastomes were obtained. Gaps in plastid assemblies were checked and filled by Sanger sequencing using primers developed in the above method. Assembly information is shown in Supplementary Table S1. Gene content and order were annotated using published plastid DNA sequences of Liliales as references using a 90% similarity to identify genes in Geneious. All tRNAs were verified by tRNAscan-SE (Schattner et al., 2005) with default settings. Because the LSC and SSC regions are flanked by the two IR regions, the junctions among these regions were identified after calculating the length of IR regions using REPuter program (Kurtz et al., 2001). Complete plastomes of other species were downloaded from NCBI for comparative genomic analysis (<https://www.ncbi.nlm.nih.gov/>; Table 1). These plastid genome sequences were aligned using Mauve program embedded in Geneious to locate the indels, inversions, and point mutations in the interspecific and intraspecific comparison between mycoheterotrophic and photosynthetic taxa (Darling et al., 2004). The maps of plastomes were produced with OGDRAW (Lohse et al., 2013). After the protein-coding regions were aligned using MUSCLE embedded in Geneious, nonsynonymous and synonymous substitution rates were calculated by the DnaSP program (Edgar, 2004; Librado and Rozas, 2009).

2.3. Phylogenomic and phylogenetic analysis

From the complete plastomes of Liliales and sequence data in GenBank, two data sets were created. The first includes 78 protein-coding genes extracted from 82 plastome sequences of monocots and other angiosperm species (including 29 new plastid genome sequences of Liliales, Supplementary Table S2). A second set that includes mycoheterotrophic *Corsia* and *Arachnitis* was made up with 17 protein-coding genes from plastid DNA, nuclear 18S rDNA and *phyC*, and mitochondrial *matR*, *atpA*, and *cob* (Supplementary Table S3). Besides the available data on NCBI, some sequences of two nuclear and three mitochondrial genes were obtained by assembling the NGS data to available references from NCBI of Liliales with a minimum of 85% similarity using Geneious (Supplementary Table S3). For the second data set, a total of 26 genera that cover all ten families of Liliales were selected, and *Symplocarpus* (Araceae), *Acorus* (Acoraceae), and *Amborella* (Amborellaceae) were included as outgroups. For sequence alignment, we used MUSCLE embedded in Geneious (Edgar, 2004; Kearse et al., 2012) and then manually adjusted the alignment following the guidelines of Kelchner (2000).

Phylogenetic analyses were performed with maximum parsimony (MP), maximum likelihood (ML), and Bayesian inference (BI). The MP analyses of two data sets were carried out in PAUP* v4.0b10 (Swofford, 2003) with all characters equally weighted and unordered. Searches of 100 random taxon addition replicates used tree-bisection-reconnection (TBR) branch swapping and MulTrees. Bootstrap analyses (PBP, parsimony bootstrap percentages, 1000 pseudoreplicates) were conducted with the same parameters to examine the relative level of clade support (Felsenstein, 1985).

Before running the ML and BI analyses, jModelTest version 0.1.1

Table 1
Comparison of the features of ptDNAs from ten families of Liliales.

Family	Species	Accession number	Length (bp)			AT content (%)	GC content (%)	Gene content ^a	IR/LSC border	IR/SSC border
			Total	LSC region	SSC region					
Liliaceae	<i>Lilium hansonii</i>	MT261163	152 168	82 013	17 621	63	37	80/30/4	ψrps19 (137 bp)	ψycf1 (1230 bp)
	<i>Lilium cernuum</i>	MT261161	152 553	82 013	17 585	63	37	80/30/4	ψrps19 (137 bp)	(gap:5 bp)
	<i>Lilium formosanum</i>	MT261162	152 610	82 043	17 527	63	37	80/30/4	ψrps19 (137 bp)	ψycf1 (1230 bp)
	<i>Lilium lancifolium</i>	MH177880	152 479	81 888	17 607	63	37	80/30/4	ψrps19 (137 bp)	ψycf1 (1230 bp)
	<i>Lilium japonicum</i>	MT261164	152 613	82 038	17 487	62.9	37.1	80/30/4	ψrps19 (140 bp)	ψycf1 (1230 bp)
	<i>Lilium distichum</i>	NC029937	152 598	82 031	17 487	62.9	37.1	80/30/4	ψrps19 (137 bp)	ψycf1 (1230 bp)
	<i>Lilium anabile</i>	MT261159	152 614	82 048	17 584	63	37	80/30/4	ψrps19 (137 bp)	ψycf1 (1230 bp)
	<i>Lilium callosum</i>	MT261160	152 630	82 059	17 587	63	37	80/30/4	ψrps19 (137 bp)	ψycf1 (1230 bp)
	<i>Lilium tsingtauense</i>	KU230438	151 983	81 424	17 575	63	37	80/30/4	ψrps19 (137 bp)	ψycf1 (1230 bp)
	<i>Lilium longiflorum</i>	KC968977	152 793	82 230	17 523	62.98	37.02	80/30/4	ψrps19 (137 bp)	ψycf1 (1251 bp)
	<i>Lilium superbum</i>	NC026787	152 069	81 466	17 365	63	37	80/30/4	ψrps19 (137 bp)	ψycf1 (1290 bp)
	<i>Pritillaria ussuriensis</i>	MT261156	152 156	81 892	17 564	63	37	80/30/4	ψrps19 (28 bp)	(overlap: 45 bp)
	<i>Pritillaria taipaiensis</i>	KC543997	151 693	81 471	17 552	63	37	80/30/4	ψrps19 (11 bp)	ψycf1 (1230 bp)
	<i>Pritillaria unibracteata</i> var <i>wobuensis</i>	KF769142	151 009	81 290	17 541	63	37	80/30/4	ψrps19 (26 bp)	ψycf1 (1230 bp)
	<i>Pritillaria cirrhosa</i>	KF769143	151 991	81 769	17 535	63.1	36.9	80/30/4	ψrps19 (28 bp)	ψycf1 (1230 bp)
	<i>Pritillaria hupehensis</i>	NC024736	152 145	81 898	17 553	63	37	80/30/4	ψrps19 (26 bp)	ψycf1 (1231 bp)
	<i>Cardiocrinum cordatum</i>	MT261152	152 391	82 179	17 322	62.9	37.1	80/30/4	ψrps19 (78 bp)	ψycf1 (1225 bp)
	<i>Cardiocrinum cordatum</i>	KX575837	152 410	82 186	17 344	62.9	37.1	80/30/4	ψrps19 (73 bp)	(overlap: 1 bp)
	<i>Cardiocrinum cathayanum</i>	KX575836	152 415	82 368	17 319	62.9	37.1	80/30/4	ψrps19 (63 bp)	(gap: 20 bp)
	<i>Cardiocrinum giganteum</i>	KX528334	152 653	82 344	17 309	62.9	37.1	80/30/4	ψrps19 (140 bp)	(overlap: 1 bp)
	<i>Gagea triflora</i>	MT261157	150 345	81 856	17 447	63	37	80/30/4	ψrps19 (46 bp)	ψycf1 (1225 bp)
	<i>Amana kuocangshanica</i>	KY401423	151 058	81 916	17 424	63.3	36.7	79/30/4	ψrps19 (106 bp)	(overlap: 1 bp)
	<i>Amana edulis</i>	KY401425	151 136	82 029	17 429	63.3	36.7	79/30/4	ψrps19 (104 bp)	ψycf1 (1129 bp)
	<i>Amana anhuiensis</i>	KY101423	150 842	82 119	17 465	63.3	36.7	79/30/4	tmH_GUG/rps19 IGS	(gap:42 bp)
	<i>Amana wanzhensis</i>	KY401422	150 913	81 758	17 445	63.3	36.7	79/30/4	ψrps19 (106 bp)	ψycf1 (1145 bp)
	<i>Amana erythronioides</i>	KY401421	150 858	82 218	17 366	63.3	36.7	79/30/4	tmH_GUG/rps19 IGS	(gap:17 bp)
	<i>Amana latifolia</i>	KY401424	150 613	81 482	17 439	63.2	36.8	79/30/4	ψrps19 (105 bp)	ψycf1 (1154 bp)
									(overlap: 66 bp)	ψycf1 (1145 bp)
									(gap:5 bp)	ψycf1 (1121 bp)
									(gap:31 bp)	

(continued on next page)

Table 1 (continued)

Family	Species	Accession number	Length (bp)			AT content (%)	GC content (%)	Gene content [†]	IR/LSC border	IR/SSC border
			Total	LSC region	SSC region					
Smilacaceae	<i>Tulipa sylvestris</i>	MT261172	151 940	82 083	17 175	26 341	63.3	36.7	80/30/4	Ψ <i>rps19</i> (106 bp) Ψ <i>ycf1</i> (1589 bp)
	<i>Erythronium japonicum</i>	MT261155	151 416	82 053	17 361	26 001	63.4	36.6	80/30/4	(overlap: 40 bp) Ψ <i>ycf1</i> (1514 bp)
	<i>Clintonia edulis</i>	MT261153	156 214	84 693	17 795	26 863	63	37	80/30/4	Ψ <i>rps19</i> (123 bp) Ψ <i>rps19</i> (156 bp) Ψ <i>ycf1</i> (1650 bp)
	<i>Calochortus venustus</i>	MT261150	155 688	84 485	18 205	26 499	62.6	37.4	80/30/4	(overlap: 58 bp) Ψ <i>ycf1</i> (1301 bp)
	<i>Streptopus ovalis</i>	MT261171	157 359	85 753	18 006	26 800	62.9	37.1	80/30/4	(gap: 1 bp) Ψ <i>ycf1</i> (1338 bp)
	<i>Tricyrtis macropoda</i>	MT261173	155 453	84 960	17 786	26 353	62.6	37.4	80/30/4	(overlap: 30 bp) Ψ <i>ycf1</i> (1001 bp)
	<i>Smilax china</i>	HMS36959	157 878	84 608	18 536	27 376	62.75	37.25	80/30/4	(gap: 13 bp) Ψ <i>ycf1</i> (1308 bp)
	<i>Smilax nipponica</i>	MT261170	158 178	85 339	18 429	27 205	62.9	37.1	80/30/4	(gap: 25 bp) Ψ <i>ycf1</i> (1207 bp)
	<i>Smilax glycyphylla</i>	MT261169	158 922	85 925	18 483	27 257	63.1	36.9	80/30/4	(gap: 17 bp) Ψ <i>ycf1</i> (1189 bp)
	<i>Smilax chinensis</i>	MT261168	158 543	85 455	18 654	27 217	62.9	37.1	80/30/4	(gap: 15 bp) Ψ <i>rps19</i> (1221 bp)
Ripogonaceae	<i>Ripogonum scandens</i>	MT261167	160 287	87 803	17 784	27 350	62.3	37.6	80/30/4	(gap: 46 bp) Ψ <i>rps19</i> (233 bp)
Philesiaceae	<i>Phileta magellanica</i>	MT261166	158 786	87 035	17 531	27 110	62.4	37.6	80/30/4	<i>tmH_GUG/rps19</i> IGS Ψ <i>ycf1</i> (1636 bp)
Melanthiaceae	<i>Lapageria rosea</i>	MT261158	160 054	87 568	17 758	27 364	62.5	37.5	80/30/4	(overlap: 62 bp) Ψ <i>ycf1</i> (1621 bp)
	<i>Paris verticillata</i>	KJ433485	157 379	82 726	17 097	28 373	62.4	37.6	80/30/4	(overlap: 47 bp) Ψ <i>ycf1</i> (1236 bp)
	<i>Paris quadrifolia</i>	KX784051	157 097	83 772	18 287	27 924	62.3	37.7	80/30/4	Ψ <i>ycf1</i> (830 bp)
	<i>Paris polyphylla var chinensis</i>	KX784048	158 307	85 187	18 175	27 473	62.8	37.2	80/30/4	Ψ <i>rps3</i> (6 bp) <i>tmH_GUG/rps19</i> IGS
	<i>Paris dunniana</i>	KX784042	157 984	84 482	18 364	27 569	62.8	37.2	80/30/4	<i>tmH_GUG/rps19</i> IGS
	<i>Paris cronquistii</i>	KX784041	157 710	84 502	18 316	27 446	62.7	37.3	80/30/4	<i>tmH_GUG/rps19</i> IGS
	<i>Paris fargesii</i>	KX784043	157 518	84 549	18 311	27 329	62.7	37.3	80/30/4	Ψ <i>ycf1</i> (1211 bp)
	<i>Paris vietnamensis</i>	KX784050	158 224	84 794	18 360	27 535	62.8	37.2	80/30/4	Ψ <i>rps3</i> (6 bp)
	<i>Paris murei</i>	KX784046	157 891	84 420	18 361	27 555	62.7	37.3	80/30/4	<i>tmH_GUG/rps19</i> IGS
	<i>Paris luquanensis</i>	KX784045	158 451	84 408	18 403	27 820	62.7	37.3	80/30/4	Ψ <i>ycf1</i> (1331 bp)
Trilliaceae	<i>Paris luquanensis</i>	KX784044	158 345	84 396	18 671	27 639	62.7	37.3	80/30/4	Ψ <i>ycf1</i> (1361 bp)
	<i>Paris forrestii</i>	KX784044	158 345	84 396	18 671	27 639	62.7	37.3	80/30/4	Ψ <i>ycf1</i> (1446 bp)
	<i>Paris marmorata</i>	KX784047	157 566	84 221	18 301	27 522	62.7	37.3	80/30/4	Ψ <i>ycf1</i> (1395 bp)
	<i>Paris polyphylla var yunnanensis</i>	KX784049	157 547	84 224	18 319	27 502	62.7	37.3	80/30/4	Ψ <i>ycf1</i> (1283 bp)
	<i>Trillium maculatum</i>	KR780075	157 359	86 340	19 949	25 535	62.5	37.5	80/30/4	<i>tmH_GUG/rps19</i> IGS
	<i>Trillium ischonokii</i>	KR780076	156 852	83 981	19 869	26 501	62.5	37.5	80/30/4	Ψ <i>rps19</i> (350 bp)
	<i>Trillium decumbens</i>	NC027282	158 552	85 411	19 566	26 791	62.3	37.7	80/30/4	Ψ <i>rps3</i> (6 bp)
	<i>Trillium cuneatum</i>	NC027185	156 610	85 977	19 897	25 368	62.5	37.5	80/30/4	Ψ <i>rps19</i> (350 bp)
	<i>Xerophyllum tenax</i>	KM078035	156 746	83 910	18 096	27 370	62.2	37.8	80/30/4	Ψ <i>rps3</i> (6 bp)
										(continued on next page)

Table 1 (continued)

Family	Species	Accession number	Length (bp)			AT content (%)	GC content (%)	Gene content [†]	IR/LSC border	IR/SSC border
			Total	LSC region	SSC region					
Alstroemeriaceae	<i>Helontopsis tubiflora</i>	KM078036	157 940	84 840	18 018	27 541	62.5	37.5	80/30/4	ψycf1 (1028 bp) (gap: 1 bp) ψycf1 (1163 bp) rpl22/rps3 IGS
	<i>Chionographis japonica</i>	KF951065	154 646	81 653	18 195	27 399	62.3	37.7	79/30/4	(overlap: 88 bp) ψycf1 (1145 bp) (gap: 7 bp) ψycf1 (992 bp)
	<i>Veratrum patulum</i>	KF437397	153 699	83 372	17 607	26 360	62.3	37.7	80/30/4	tmH_GUG/rps19 IGS
	<i>Alstroemeria aurea</i>	KC968976	155 510	84 241	17 867	26 701	62.74	37.26	80/30/4	(gap: 10 bp) ψycf1 (1741 bp) ψrps19 (72 bp)
	<i>Bomarea edulis</i>	KM233641	154 925	84 049	17 699	26 566	61.8	38.2	80/30/4	tmH_GUG/rps19 IGS
Colchicaceae	<i>Luzuriaga radicans</i>	KM233640	157 885	86 912	18 175	26 399	61.9	38.1	80/30/4	(overlap: 45 bp) ψycf1 (1122 bp) (overlap: 19 bp) ψycf1 (2571 bp)
	<i>Androcymbium greuterocymbium</i>	MT261148	154 804	83,195	16,493	27,558	62.4	37.6	80/30/4	(overlap: 38 bp) ψycf1 (2779 bp) (overlap: 39 bp) ψycf1 (2555 bp)
	<i>Colchicum autumnale</i>	KP125337	156 462	84 246	16 734	27 741	62.4	37.6	80/30/4	(overlap: 49 bp) ψycf1 (1388 bp) ψycf1 (2535 bp)
	<i>Gloriosa superba</i>	KP125338	157 924	85 012	16 786	28 063	62.4	37.6	80/30/4	(overlap: 32 bp) ψycf1 (1546 bp) (gap: 22 bp) ψycf1 (1059 bp)
	<i>Tripladenia cunninghamii</i> <i>Wurmbea burttii</i>	MT261174 MT261176	155 652 155 297	84 314 82 848	17 708 16 581	26 815 27 934	62.4 62.3	37.6 37.7	80/30/4 80/30/4	tmH_GUG/rps19 IGS
Ptermanniaceae	<i>Uvalaria grandiflora</i>	MT261175	156 985	85 194	17 919	26 936	62.4	37.6	80/30/4	(gap: 8 bp) ψycf1 (1820 bp) (overlap: 41 bp) rps3 (53 bp)
	<i>Petermannia cirrhosa</i>	MT261165	156 852	85 223	18 359	26 635	62	38	80/30/4	rps19/rpl22 IGS
Campynemataceae	<i>Campynema lineare</i>	MT261151	156 305	84 547	16 892	27 433	63.1	36.9	80/30/4	rps19/rpl22 IGS
	<i>Corsia dispar</i> <i>Arachnitis uniflora</i>	MT261154 MT261149	63 172 24 846	– 8243	– 4255	– 6174	69.2 62.3	30.8 37.1	30/24/4 16/5/4	– rpl16 exon 2 (59 bp)

IGS: intergenic space; CDS: coding region; The hyphen (–) means no data.
[†] Gene content includes protein coding/tRNA/rRNA.

was used to identify a suitable model with Akaike's information criterion (Posada, 2008). The GTR + I + G model was found for both data sets. However, the TIM2 + I + G model was indicated for two nuclear genes data, whereas the GTR + I + G was calculated for three mitochondrial genes and 17 plastid genes data. Therefore, the incongruence length difference (ILD) test was used to assess data congruency among three data sets (nuclear, mitochondrial, and plastid genomes; Farris et al., 1995) by using PAUP v.4.0b10 and 1000 heuristic search replications. The ILD test indicated significant incongruence ($P = 0.001$) between the three data sets. However, although the individual topologies indicated some incongruences between three data sets, visual inspection of the separate analyses showed that none of the strongly supported clades (BP [parsimony bootstrap percentages] > 90) were mutually incompatible (data not shown). Furthermore, when all molecular data sets were combined, the strict consensus MP tree was better resolved than either individual analysis. Therefore, we decided to combine nuclear, mitochondrial, and plastid genome data sets with their weakly-incongruence trees. The data sets were unpartitioned and set up with GTR + I + G model. We used the IQ-TREE web server (<http://iqtree.cibiv.univie.ac.at/>) to make the ML searches (Trifinopoulos et al., 2016). Support value was calculated with 1000 replicates of rapid bootstrap (Stamatakis et al., 2008). BI used MrBayes v3.1.2 (Ronquist and Huelsenbeck, 2003). For the analyses, two simultaneous runs were performed starting from random trees for at least

1×10^6 generations. One tree was sampled every 1000 generations. In total, 25% of trees were discarded as burn-in samples. The running was only terminated when the standard deviation of split frequencies was lower than 0.01. Then the effective sample size values (ESS) were checked for model parameters (at least 200). The phylogenetic trees were illustrated using the FigTree v1.4.0 program (<http://tree.bio.ed.ac.uk/software/figtree/>).

2.4. Morphological data analysis

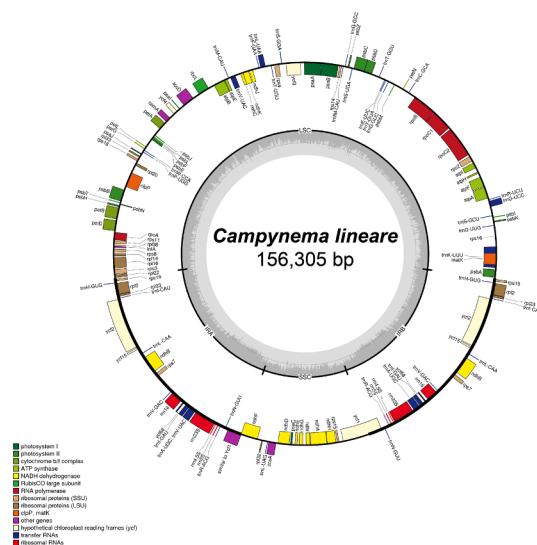
Morphological data (12 characters) for 60 genera that represent all ten families of Liliales were used (Supplementary Table S4) with outgroups including *Dioscorea* (Dioscoreaceae, Dioscoreales), *Narthecium* (Nartheciaceae, Dioscoreales) and *Japonolirion* (Petrosaviaceae, Petrosaviales). We used parsimony in Mesquite (Maddison and Maddison, 2017) to trace character evolution in Liliales on the phylogeny inferred from two data sets as the backbone to illustrate patterns of morphological change.

3. Results

3.1. Differentiation of plastid genomes in Liliales

The 29 newly completed plastid genomes in Liliales exhibited wide

A



B

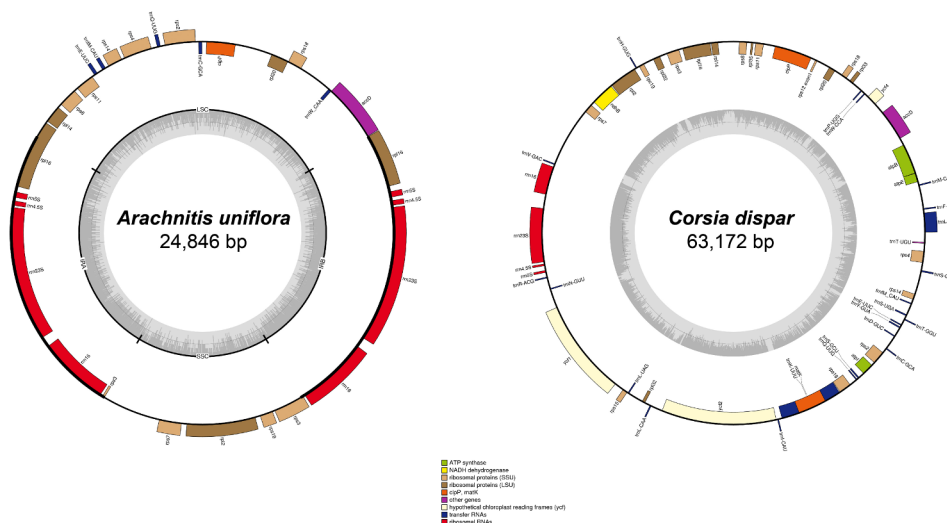


Fig. 1. Map of photosynthetic and mycoheterotrophic plastid genomes and their gene contents. A. Plastid genome of *Campynema lineare* (representative of the photosynthetic taxa). B. Plastid genomes of *Arachnitis uniflora* and *Corsia dispar* (representative of the mycoheterotrophic taxa). Genes shown outside of the outer circle are transcribed counterclockwise, whereas those shown inside are transcribed clockwise. The thick lines in the small circles indicate the inverted repeat regions. The dark gray area in the inner circle indicates the CG content of the plastome. The colors represent different groups of genes. LSC: large single copy; SSC: small single copy; IRA: inverted repeat region A; IRB: inverted repeat region B.

variation in size and gene content. The smallest plastid genome, 24,846 bp, occurred in *Arachnitis uniflora*. In contrast, *Ripogonum scandens* possessed the largest plastid genome, 160,287 bp (Table 1). Liliales exhibited two groups of plastid genomes due to photosynthetic and mycoheterotrophic life histories (Fig. 1). In the first group of photosynthetic taxa, the plastid DNA sequences had the typical quadripartite structure and genes related to photosynthesis (Supplementary Table S5). However, the length of plastid DNA varied among and within photosynthetic families, ranging from 150 to 160 kb. The largest range was found in Liliaceae (15 genera), whereas Smilacaceae (monotypic) showed the least length variation. Gene content was highly conserved and comprised 80 protein-coding genes, 30 tRNA genes and four rRNA genes. Gene deletions were also observed in Liliales (Supplementary Table S6). In particular, the *rps16* gene was exhibited a partial deletion in some species in Melanthiaceae and Campynemataceae. In *Colchicum autumnale* (Colchicaceae), the *ycf15* gene was absent. A partial deletion of *ycf15* was found in *Gagea trifolia*, *Tulipa sylvestris*, and *Erythronium japonicum* though in other species of Liliaceae *ycf15* was complete. Similarly, in Alstroemeriaceae, the conserved reading frame of *ycf15* was absent in *Alstroemeria aurea* and *Bomarea edulis* whereas an exon of *ycf15* with internal stop codon was found in *Luzuriaga radicans*. In Liliales, the partial deletion of *infA* was detected in some species of Smilacaceae, Melanthiaceae (*Veratrum*), Alstroemeriaceae (*Alstroemeria*), and Liliaceae (*Tricyrtis*, *Streptopus*, *Erythronium*, and *Tulipa*) whereas internal stop codons in *infA* were found in species from other families (including Colchicaceae and Campynemataceae). However, putatively functional copies of *infA* were recorded in species of Petermanniaceae, Philesiaceae, and Ripogonaceae. Notably, Li et al. (2017) found the absence of *infA* in the plastid genome of six *Amana* species (Liliaceae). Unlike other plastid genes, the *cemA* start codon has a polyA sequence (Supplementary Table S6), which ranged from 5 to 17 bp and initially was annotated as a pseudogene in species of Liliaceae, Ripogonaceae, Philesiaceae, Melanthiaceae, and Petermanniaceae. However, in Melanthiaceae pseudogenization of *cemA* was only found in *Paris* and *Trillium* (tribe Parideae).

Unlike the high conservation of plastid genes in the photosynthetic taxa, there were large rearrangements and gene losses in the mycoheterotrophic species, *Corsia dispar* and *Arachnitis uniflora* (Corsiaceae; Fig. 1B), in which there were 58 and 25 exons, respectively (Supplementary Table S5). In these taxa, photosynthesis related genes were deleted or pseudogenized, whereas genes for ribosomal proteins, rRNAs, and some tRNAs remained intact (Supplementary Table S5), notably, *matK* in *Corsia dispar*. The number of retained genes was variable between *C. dispar* and *C. boriensis*. Besides 28 shared coding regions, the former possessed partial sequences of *ycf4*, *rps16*, *ndhB*, *atpI*, *atpE*, *atpB*, and *psaA*, whereas only fragments of *psaB*, *psbC*, and *ccsA* were recorded in the latter. Although the plastid genome of *A. uniflora* is the smallest among Liliales species, it still had the quadripartite structure in contrast to the species of *Corsia*.

3.2. Intraspecific and interspecific comparative genomic analysis

Due to the availability of many complete plastid genomes and a full set of protein-coding genes, we conducted detailed comparisons (Supplementary Table S7, Supplementary Table S8). The results showed that the total length of these genomes varied, although they were highly similar (> 99%, Supplementary Table S7). The circular structure was routinely preserved, but the borders of inverted repeat and large and small single copy regions (IR-LSC-SSC) varied slightly. Specifically, an overlap of 1 bp and a gap of 20 bp were found in *Cardiocrinum cordatum*, albeit with the border located at an identical position in *ycf1* (1225 bp, Supplementary Table S7). A similar pattern was recorded in *Lilium hansonii* and *Campynema lineare*. In contrast, *Lilium tsingtauense* has an identical IR-LSC-SSC borders. However, this species exhibited the largest number of point mutations in both intergenic spacers (IGS) and exons and contained the longest indel, 593 bp. Similar to *L.*

tsingtauense, there were several point mutations and indels in two plastid genome sequences of *Campynema lineare*. Additionally, both *L. tsingtauense* and *C. lineare* have small inversions in the IGS regions, ranging from 5 to 14 bp (Supplementary Table S7). There was also a small inversion of 10 bp in *ycf1* of *Campynema lineare*. Comparisons within the monotypic genera were not conducted but the availability of protein-coding sequences from these taxa revealed several point mutations and indels (Supplementary Table S7). Mycoheterotrophic *Arachnitis uniflora* had more point mutations in exons than in those of photosynthetic species. We also conducted detailed interspecific comparison between *Smilax nipponica* (East Asia)/ *S. glycyphilla* (eastern Australia), and *Corsia dispar* (Queensland, Australia)/ *C. boriensis* (Papua New Guinea, Supplementary Table 8), which showed significant differences between photosynthetic and mycoheterotrophic taxa. In contrast to the *Corsia* species, which exhibited 1610 point mutations and 101 indels (1–144 bp), only 141 point mutations and nine indels were detected in two *Smilax* species. For Ka/Ks ratios (Fig. 2; Supplementary Table S7, Supplementary Table S8), the mycoheterotrophic species were not consistently higher. For example, the Ka/Ks ratio for *ycf2* of *Campynema lineare* exceeded those of *Corsia dispar* and *C. boriensis*. Higher rates in *C. lineare* were also found in *rpl2*, *rpl16*, *rps18*, *rps19*, *matK*, *rpl22*, and *rps15*. Among mycoheterotrophic taxa, Ka/Ks ratios fluctuated across conserved plastid exons (Fig. 2).

3.3. Characterization of LSC-IR junction and SSC-IR boundary

In comparison with IR-LSC junction, IR-SSC boundary was similar among species of Liliales (Table 1). In fact, the location of IR-SSC border was found within the coding region of *ycf1* except for *Arachnitis uniflora* (in which it was found in *rps3*) and three *Trillium* species (*T. maculatum*, *T. decumbens*, and *T. cuneatum*) for which the boundary was located in the *ndhF* exon. Additionally, the length of *ycf1* varied from 830 to 2779 bp. However, an identical length for *ycf1* was found in the species of *Lilium*, *Fritillaria*, *Amana*, and *Paris* studied (Table 1). The boundary between IR and LSC varied among species of Liliales, ranging from *trnH-GUG/rps19* IGS to a part of *rps3* (Table 1). However, in the mycoheterotrophic species, the IR-LSC junction was located in *rpl16* exon 2 in *Arachnitis uniflora*, whereas in *Corsia dispar* the IR region was absent.

3.4. Familial relationship in liliales

The alignment matrix of 78 plastid genes for 53 ingroup (Liliales) and 29 outgroups species contained 68,270 characters, of which 51,841 (75.9%) were variable and 26,717 (39.1%) were potentially parsimony-informative. The MP analysis produced two equally most-parsimonious trees (tree length = 152,476 steps; CI = 0.51; RI = 0.64). The MP and ML trees (not shown) were identical in topology to the BI tree (Fig. 3). Liliales and all families as currently circumscribed are well supported except for the nodes of Smilacaceae/Liliaceae and within the genera of Liliaceae and Colchicaceae which are fairly low for one of the analyses (Fig. 3). A similar pattern was also found in the three genomes data of which the nodes of Liliaceae/Smilacaceae, genera of Liliaceae, and Melanthiaceae/(Liliaceae/Smilacaceae/Philesiaceae/Ripogonaceae) were not highly supported (Fig. 4). Melanthiaceae/ Corsiaceae (*Arachnitis* and *Corsia*) and Campynemataceae were well supported as sister to the rest Liliales families. Petermanniaceae were sister to the clade of Alstroemeriaceae/Colchicaceae. For the remaining families, Melanthiaceae are sister to Liliaceae/Smilacaceae/Philesiaceae/Ripogonaceae, within which Smilacaceae are sister to Liliaceae and Philesiaceae are sister to Ripogonaceae. Within Liliaceae, *Tricyrtis*/*Prosartes*/*Streptopus* are sister to all other genera, after which *Calochortus* is sister to the rest (Fig. 3).

Among the 23,335 characters of the combined three-genome matrix for 26 ingroup (Liliales) and three outgroups accessions, 3656 (15.7%) were potentially parsimony-informative. The MP analysis of the

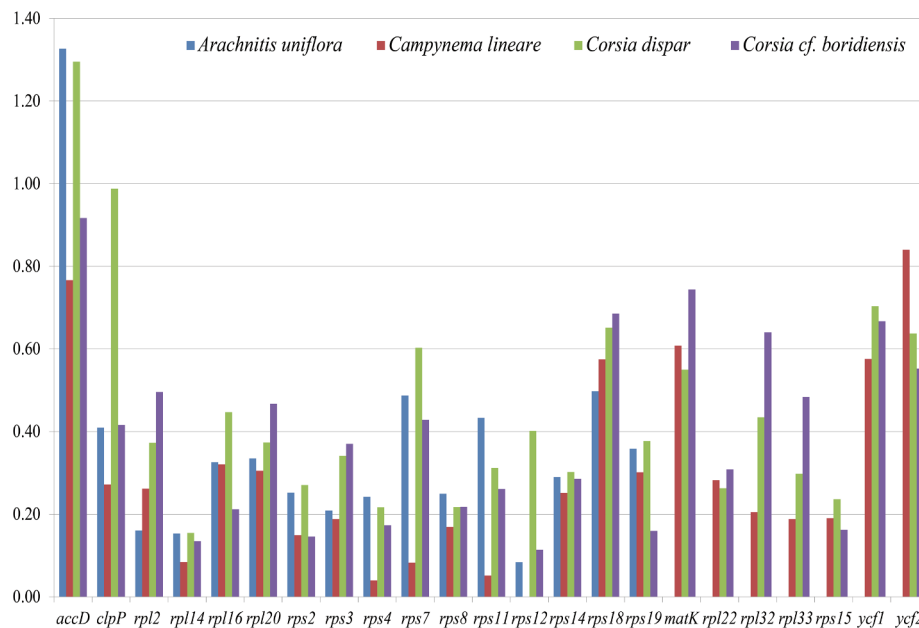


Fig. 2. The comparison of synonymous and nonsynonymous substitutions among genes of photosynthetic and mycoheterotrophic taxa.

combined data set resulted in single most-parsimonious tree (tree length = 15,680; CI = 0.69; RI = 0.58). The MP and ML trees (not shown) were identical to the BI tree (Fig. 4). Consistent with results from plastome data set, Liliales and all families were well supported. Familial relationships were also the same. However, in Liliaceae, *Calochortus* and *Streptopus* formed a clade separate from *Tricyrtis*, which was sister to other genera. Although the three-genome data revealed different relationships in Liliaceae, support for these differing relationships was low (Fig. 4).

3.5. Reconstruction of ancestral characters in Liliales

Patterns of morphological change indicate that exstrose anthers and nectaries are synapomorphies for Liliales (Fig. 5) although introrse anthers and no nectaries are present in some families (i.e., Liliaceae, Smilacaceae, Ripogonaceae, Philesiaceae, Colchicaceae, Alstroemeriaceae, Melanthiaceae, and Campynemataceae). Rhizomes, capsules, and parallel venation are plesiomorphic features in Liliales. In particular, inflorescences are diverse in Liliales: solitary flowers, racemes, thyrses, umbels, corymbs, panicles, spikes, and cymes. Colchicaceae have greater inflorescence variation than the other families. We also identified synapomorphic characteristics for specific families. For example, resupinate (inverted) leaves are found in all Alstroemeriaceae. Meanwhile, stipules were only recorded in Smilacaceae. Also, petiolate leaves with reticulate veins are apomorphies of Petermanniaceae. Corsiaceae have unique features such as leaflessness and tuber-like roots, presumably due to their association with fungi (Claroideoglomeraceae, Glomeraceae, and Acaulosporaceae; Renny et al., 2017). Dioecious flowers are found in members of Smilacaceae, Ripogonaceae, Philesiaceae, Melanthiaceae, and Colchicaceae. The ancestral ovary position was not clearly identified for Liliales. However, a superior ovary is a synapomorphy of Liliaceae, Smilacaceae, Ripogonaceae, and Philesiaceae. In contrast, inferior or half inferior is a synapomorphy of Petermanniaceae, Campynemataceae, and Corsiaceae. Alternate leaves are ancestral but whorled in Liliaceae and Melanthiaceae and opposite in Ripogonaceae.

4. Discussion

4.1. Evolution of plastid genomes in Liliales

Plastid genomes of Liliales exhibit two distinct types with respect to gene content and order linked to being an autotroph or mycoheterotroph (Fig. 1, Table 1, and Supplementary Table S5). Specifically, Corsiaceae had a greater number of point mutations and indels in comparison with the other families (Supplementary Table S7, Supplementary Table S8), indicating higher rate of genomic divergence in mycoheterotrophic taxa. Previously, this pattern was detected between autotrophic and mycoheterotrophic species in Orchidaceae (Schelkunov et al., 2015). Within the photosynthetic group of Liliales, plastomes were highly conserved in structure, and gene content and order. In comparison with other Liliales families, Melanthiaceae displayed the greatest number of inversions, duplications, pseudogenes and deletions (Kim et al., 2016; Do and Kim, 2017). Apart from these events, *infA*, *cemA*, *rps16*, and *ycf15* exhibited different stages of pseudogenization/deletion (Supplementary Table S6, Nguyen et al., 2015; Do and Kim 2017). The *infA* gene, a translation initiation factor, has been deleted several times, mostly in eudicot species due to translocation to the nucleus (Millen et al., 2001). This event was also recorded in monocot species such as *Allium* (Asparagaceae, Asparagales), *Cyperus* (Cyperaceae, Poales), *Lemna minor*, *Colocasia esculenta* (both Araceae, Alismatales) and others (Millen et al., 2001; Mardanov et al., 2008; Ahmed et al., 2012). The *cemA* exon, previously called *ycf10*, codes for a protein localized in the inner chloroplast envelope membrane (Sasaki et al., 1993). This gene was recognized to be involved in the uptake of inorganic carbon into chloroplasts of algae (Rolland et al., 1997). Furthermore, the homologous gene *cotA* in cyanobacteria affected CO₂ transport and extrusion of protons (Katoh et al., 1996a, b). Although these studies focused on algae and bacteria, these results suggest the potential essential role of *cemA*, but no research on the function of *cemA* in angiosperms has been previously conducted except for sequence comparisons of this gene in plastomes (Yang et al., 2010; Do et al., 2014). It was also reported that *cemA* was lost from the plastid genome of the mycotrophic monocot *Petrosavia stellaris* (Petrosaviales)

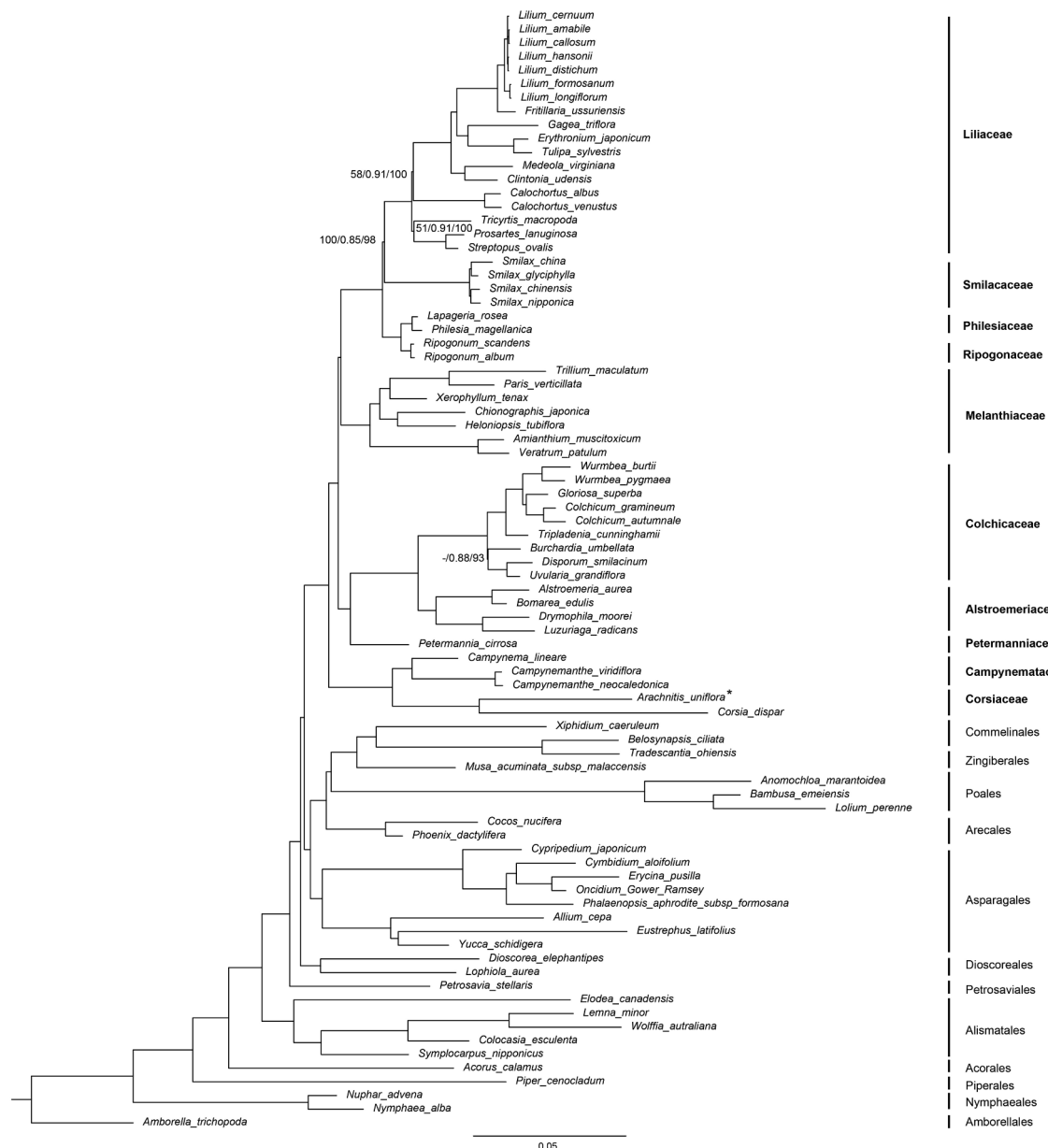


Fig. 3. The Bayesian tree of Liliales inferred from 78 plastid protein-coding genes. Numbers indicate support (maximum parsimony bootstrap (PBP)/posterior probability (PP)/ maximum likelihood bootstrap (MBP)). The asterisk indicates a long branch of *Arachnitis uniflora* compared with other species in this tree. This branch was reduced by a factor of ten. Only support under PBP = 90/PP = 1.00/ MBP = 100) is shown. The dashes “-” indicate incongruence between MP and ML/BI trees.

and the unicellular green alga *Ostreococcus tauri* (Robbens et al., 2007; Logacheva et al., 2014), although in the latter it was present in the nuclear genome. According to this study, *cemA* is intact in plastid DNA with different lengths of homopolymers downstream from its start codon (Supplementary Table S6). The *rps16* gene, encoding ribosomal protein S16, is commonly detected in plastomes. However, loss of this gene was also recorded in some taxa such as *Connarus*, *Epifagus*, *Pinus*, *Viola*, *Fagus*, legume species, and the moss *Physcomitrella patens* subsp. *patens* (Downie and Palmer, 1992; Doyle et al., 1995; Sugiura et al., 2003). To define the underlying cause for loss of this essential gene, it was proposed that the *rps16* gene was transferred to the nucleus and its protein product was able to target both chloroplast and mitochondria in the case of *Medicago truncatula* and *Populus alba* (Ueda et al., 2008). These results suggest that the transfer of *rps16* in multiple lineages has apparently occurred independently several times in plants. It was lost or partly deleted in Campynemataceae and tribes Chionographideae

and Melanthieae of Melanthiaceae; therefore, ribosomal protein S16 was predicted to be untranscribed and untranslated. However, this deficiency could be compensated by nuclear *rps16* products as described previously (Ueda et al., 2008). Previously, Kim et al. (2016) suggested that an inversion of *ndhC-rbcL* region in ptDNA occurred in the two subgenera *Sessilium* and *Trillium*. Nguyen et al. (2015) characterized the loss of *ycf15* only in tribe Colchiceae of Colchicaceae species. Collectively, the genomic events in Liliales are useful source of molecular markers for further studies of evolutionary patterns of plastomes, phylogeny, and population genetics. For example, the process of termination of *infA* in Liliales was reflected through pseudogenization followed by partial and then full deletion. In fact, partial deletion of *infA* was found in all examined species of Smilacaceae in this study, suggesting that this event is a synapomorphy for this monogeneric family. Previously, Yang et al. (2010) reported the polyA sequences in *cemA* among monocots. In this study, it also varied among Liliales.

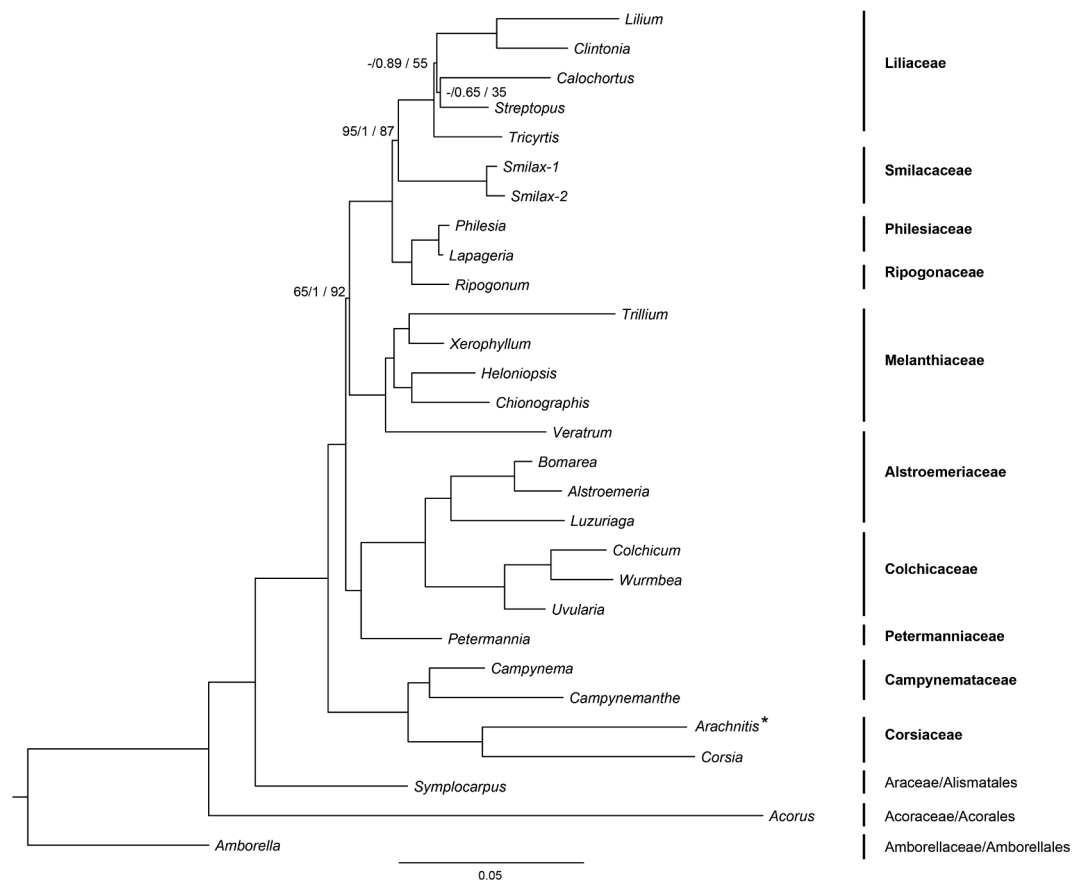


Fig. 4. The Bayesian tree of Liliales inferred from combined data of plastid, mitochondrial and nuclear sequences. Numbers indicate support (maximum parsimony bootstrap (BPB)/posterior probability (PP)/ maximum likelihood bootstrap (MBP)). The asterisk indicates a long branch of *Arachnitis uniflora* compared with other species in this tree. This branch was reduced by a factor of five. Only support under BPB = 90/PP = 1.00/ MBP = 100) is shown. The dashes “-” indicate incongruence between MP and ML/BI trees.

Therefore, *cemA* could be a potential region for further population studies not only in Liliales but also in other angiosperms.

In contrast to the highly conserved of plastid genomes of the photosynthetic taxa, the mycoheterotrophic species of Liliales exhibited significant changes in structure and gene content with different patterns of loss and the order of the retained genes (Fig. 1, Supplementary Table S5). Previously, Graham et al. (2017) surveyed and summarized the loss of genes across mycoheterotrophic taxa, but did not include the two samples Corsiaceae included here. Within the two *Corsia* species gene content was different (Supplementary Table S5). Further studies on plastid DNA should be conducted to further clarify the evolution of this mycoheterotrophic genus, which was estimated to have diverged approximately from *Arachnitis* 53 million years ago (Mya; Mennes et al., 2015).

Our analyses of the plastome evolution in the Liliales confirmed the quadripartite plastome structure, the IR expansion and contractions driving the plastid genome size, changes in the position of the IR-LSC boundaries, and the extreme reduction of the plastome in non-photosynthetic taxa. The quadripartite plastome structure has been hypothesized to be essential for stability (Daniell et al., 2016) and was highly conserved in the samples studied. The IR expansion and contraction affect the total size of plastid genome (Ma et al., 2013). In some eudicots (i.e., Fabaceae), the IR is lost (Choi and Choi, 2017), which has been thoroughly characterized across photosynthetic and non-photosynthetic species (Downie and Jansen, 2015; Wang et al., 2015; Raman and Park, 2016; Cauz-santos et al., 2017). In this study, various IR-LSC junctions were recorded; however, some identical boundaries were also found among the examined species (Table 1). Previously, Wang et al. (2008) surveyed 123 species of basal angiosperms,

magnoliids, monocots and eudicot groups and suggested three types of IR-LSC junctions, concluding that the duplication of *trnH-rps19* could be a synapomorphy of extant monocot orders. However, the limited number of monocots sampled prevented them from concluding this definitively. For example, duplication of *trnH* can be found in basal angiosperms (Nymphaeales) and eudicots (Actinidiaceae and Ericaceae, both Ericales; Yao et al., 2015). Additionally, the IR region expanded to include *psbB* in *Berberis bealei* (Berberidaceae, Ranunculales; Ma et al., 2013). Based on our observations here, we propose five types of IR-LSC junctions (Supplementary Fig. S1A). In type I, the IR-LSC border based that proposed in the literature (Wang et al., 2008) is located downstream of *trnH-GUG*. We found junctions upstream of *rps19*, *rpl22*, and *rps3* in types II, III, and IV, respectively. In type V species, the IR-LSC border is located in or downstream of the *rps3* exon. In Liliales, type III, the most common, is found in all ten families. Notably, Melanthiaceae possess four types (I, II, III, and IV), whereas type I was not found in any Liliales (Supplementary Fig. S1B). The presence of various IR-LSC junctions in each family of Liliales revealed a high rate of differentiation. For phylogenetic and barcoding studies, *trnH-psbA* is not a suitable region in Liliales at familial level (Table 1, Supplementary Fig. S1B). Further observations of IR-LSC junctions in other species of monocots and basal angiosperms revealed that the five junction types are present (Supplementary Table S9). However, some orders have conserved junctions (i.e., Arecales, and Laurales), whereas others have variable IR-LSC borders (i.e., Asparagales, Alismatales, and Zingiberales). The IR-LSC junctions among basal angiosperms and monocot orders reveal changes from type I to type V (Supplementary Table S9). However, IR-LSC junctions appear to be randomly distributed in the surveyed taxa, especially Liliales.

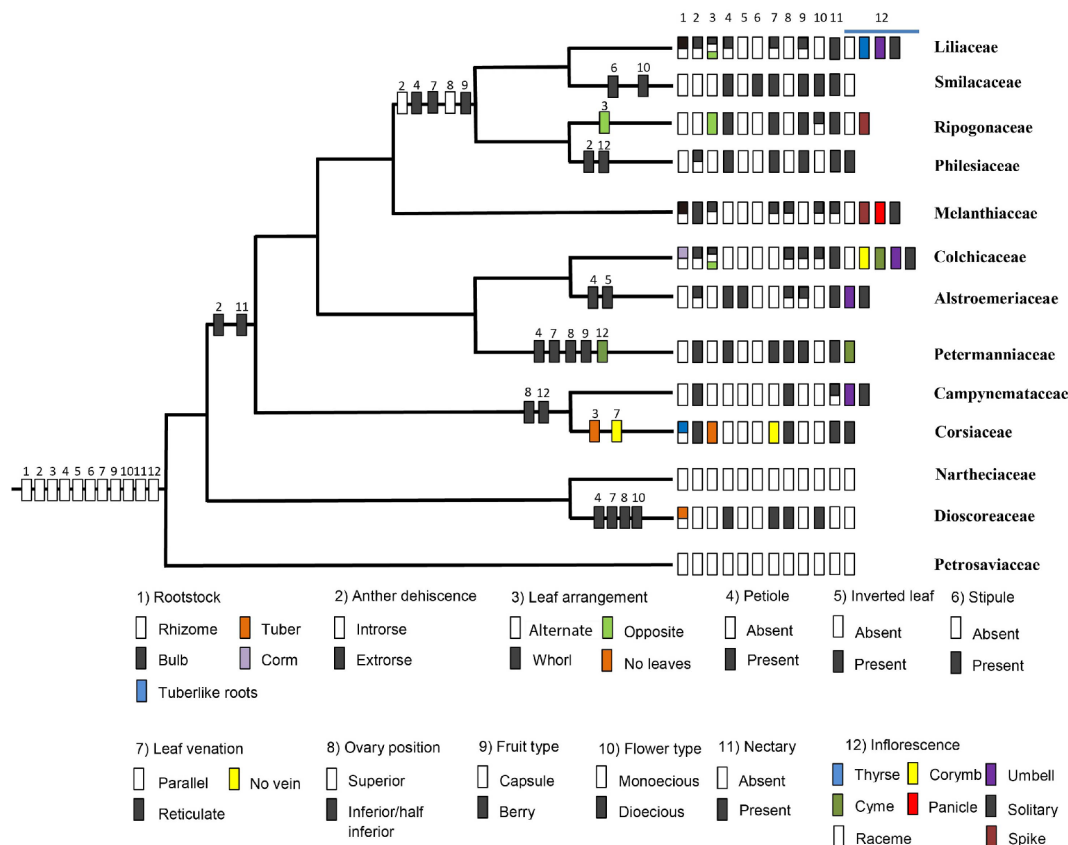


Fig. 5. Diagram of morphological characteristics in Liliales. The tree was drawn from familial relationship generated from plastid genome and three-genome data.

4.2. The junction between IR and SSC regions: A new pool of molecular markers

The observation of IR-SSC borders among Liliales and other monocots revealed that this region could be potential for mining molecular markers (Table 1, Supplementary Table S9). Most IR-SSC borders in Liliales were located in the *ycf1* exon (> 5000 bp; Table 1). Uniquely, three *Trillium* species exhibited a junction in *ndhF* with *ycf1* in the SSC. Previously, Drescher et al. (2000) attempted to determine the function of *ycf1* and *ycf2* and concluded these exons were indispensable for cell survival. However, plastid genome data revealed that *ycf1* and *ycf2* were absent in many species of Poales (Katayama and Ogihara, 1996; Guisinger et al., 2010), implying therefore that they could be replaced by copies in the nucleus or mitochondria. Losses of *rps16* and *infA* were recorded in several clades with nuclear replacements hypothesized (Millen et al., 2001; Choi and Choi, 2017). However, there have no records of *ycf1* and *ycf2* copies outside of plastid genomes, which need further studies to clarify the reason for the loss of these two genes. The *ycf1* exon is not only one of the largest in the plastome but also that with the highest nucleotide diversity (Morris and Duvall, 2010), making it a target for mining angiosperm barcodes (Dong et al., 2014; Dong et al., 2015). Among species with *ycf1*-located IR-SSC borders, there are three types across angiosperms (Supplementary Fig. S2, Supplementary Table S9), making this region promising for development of DNA barcodes.

4.3. Familial relationships and a revised classification of Liliales

The plastome and combined three-genome data resulted in identical estimates of familial relationships in Liliales (Figs. 3, 4), justifying a combined analysis. Neyland and Hennigan (2003) first revealed a relationship of *Corsia* with *Campynema* (Campynemataceae) based on large subunit ribosome rDNA sequences. Kim et al. (2013) showed that

Corsiaceae was not a member of Liliales, but this outcome might have been affected by the small number of samples and too little sequence data (only *rbcl* of *Arachnitis uniflora* was used). Recently, Mennes et al. (2015) and Givnish et al. (2016) analysed all remaining protein-coding regions in the plastid DNA of *A. uniflora* and *Corsia* sp. to produce results congruent with ours: Corsiaceae (*Arachnitis* and *Corsia*) are sister to Campynemataceae, which together are sister to the rest of Liliales (Figs. 3, 4).

Although, Givnish et al. (2016) concluded that their study resolved all uncertainty in relationship within Liliales, the relationships among Smilacaceae, Philesiaceae and Ripogonaceae were not highly supported in their study. Additionally, Givnish et al. (2016) found that *Calochortus* forms a group with *Prosartes*, *Scoliopus* and *Streptopus*, whereas *Tricyrtis* was sister to the rest of Liliaceae. However, Kim and Kim (2018) used four plastid genes with a greater number of samples from *Calochortus*, *Tricyrtis*, and *Streptopus* and found *Tricyrtis* as a sister with strong support to *Streptopus*, *Prosartes*, and *Scoliopus* rather than *Calochortus*. Our study strongly supported the relationships (Fig. 3) found by Kim and Kim (2018), but the relationships of *Calochortus*, *Tricyrtis*, and *Streptopus* are not the same in the 78-gene and three-genome analyses (Figs. 3, 4). The different relationship among genera of Liliaceae could be caused by taxonomic gap in this study. In the 78-gene and three-genome data, there are 11 and 6 genera out of 15 genera of Liliaceae in comparison to those of Kim and Kim (2018), respectively. Additionally, the small number of sequence data from nuclear and mitochondrial genomes might result in the different topologies of two data sets. In this study, the topology of phylogenetic trees inferred from nuclear and mitochondrial genomes are different (Data not shown). Previously, there was incongruence reported among nuclear, mitochondrial and plastid data in reconstructing of phylogenetic relationship (Petersen et al., 2006; Seberg et al., 2012; Wang et al., 2014), but although we observed incongruence within Liliaceae, both our analyses showed a highly supported relationship of Liliaceae to Smilacaceae (Figs. 3, 4).

The placement of the families of Liliales has been intensively debated. In our results, Smilacaceae are strongly supported as sister to Liliaceae (Figs. 3, 4), a result also found in some previous studies (Patterson and Givnish, 2002; Chase et al., 2006; Fay et al., 2006). In some previous studies, Smilacaceae, Ripogonaceae, and Philesiaceae were found to form a clade, but support was weak (Vinnersten and Bremer, 2001; Kim et al., 2013). In contrast, Givnish et al. (2016) placed Smilacaceae sister to Liliaceae, but support was also low. Kim et al. (2013) found that Melanthiaceae-Petermanniaceae formed a clade that is sister to other families, whereas Petersen et al. (2013) showed a polytomy of Melanthiaceae, Petermanniaceae, Alstroemeriaceae, Colchicaceae, Liliaceae, and Smilacaceae/Ripogonaceae/Philesiaceae. In the results of this study, Melanthiaceae are sister to Philesiaceae/Ripogonaceae plus Liliaceae/Smilacaceae. Additionally, Petermanniaceae are sister to Alstroemeriaceae/Colchicaceae, a result also found by Fay et al. (2006) and Mennes et al. (2015).

Previously, different classifications of Liliales have been suggested due to the combination of families. For example, Uvulariaceae were embedded in Colchicaceae and treated as a subfamily Uvularioideae by Nordenstam (1998), whereas Trilliaceae was considered a tribe (Paridae) of Melanthiaceae (Angiosperm Phylogeny Group, 1998). Additionally, Scoliopaceae, Tricyrtidaceae, and Calochortaceae were considered synonyms of Liliaceae (Patterson and Givnish, 2002). The most recent change within Liliales was the merger of Luzuriagaceae with Alstroemeriaceae (Angiosperm Phylogeny Group, 2009). *Luzuriaga* species have inverted leaves, which are also found in other Alstroemeriaceae. Additionally, previous molecular studies revealed close relationship of *Luzuriaga* to *Alstroemeria*, *Drymophila* and *Bomarea* species (Vinnersten and Bremer, 2001; Patterson and Givnish, 2002; Chase et al., 2006; Fay et al., 2006; Chacón et al., 2012; Kim et al., 2013; Givnish et al., 2016). The only monotypic family, Petermanniaceae, has a close relationship with Alstroemeriaceae and Colchicaceae (Figs. 3, 4). Also, it has unique morphological features (i.e., woody climbers with fibrous roots, leaf-opposed inflorescences, and reticulate leaf veins) in comparison to Alstroemeriaceae and Colchicaceae. This family split from the common ancestor with Colchicaceae and Alstroemeriaceae approximately 98.1 Ma (Givnish et al., 2016, Supplementary Table S10). In this study, Ripogonaceae and Philesiaceae are found to be a highly supported clade in both the 78-plastid gene and multi-genome phylogenies. The close relationship of Ripogonaceae and Philesiaceae was also highly supported in previous studies inferred from molecular data (Kim et al., 2013; Mennes et al., 2015; Givnish et al., 2016). These two families also share morphology such as short woody rhizomes, cuticular wax with parallel platelets, petiolate leaves, crassinucellate ovules, berries, and a disintegrating testa (Kim et al. 2013). Based on the combined evidence, we propose merging Ripogonaceae with Philesiaceae.

Key to the genera of Philesiaceae

1. Opposite leaves; inflorescence raceme or spike; small bisexual/unisexual flower (8 mm); pollen prolate, surface reticulate.....

Ripogonum

1. Alternate leaves; solitary flower; bisexual flowers; pollen inaperturate, surface echinate; 2

2. Erect shrubs; axillary flower; linear leaves; perianth whorls unequal (inner much longer than outer); anther dehiscing by pores.....

Philesia

2. Vines, terminal flowers, ovate to lanceolate leaves, perianth whorls equal; anther dehiscing intorsely.....

Lapageria

Key to the families of Liliales

1. Achlorophyllous, herbs without leaves**Corsiaceae**

1. Chlorophyllous, herbs with green leaves,

2. Leaves petiolate

3. Leaves resupinate (twisted/ upside-down), parallel veins.....**Alstroemeriaceae**
3. Leaves not twisted or upside-down, reticulate veins
4. Leaves with tendril-like stipules.....**Smilacaceae**
4. Leaves without tendril-like stipules
5. Leaves opposed pedunculate inflorescences, tendrils.....**Petermanniaceae**
5. Terminal solitary flower/axillary racemes, no tendrils.....**Philesiaceae**
2. Leaves without petiole
6. Greenish tepals; small flowers with inferior ovary and two distinct phases (male/female phase; filaments recurve and twist counter clockwise after the abscission of anthers) within the same inflorescence.....**Campynemataceae**
6. Colorful tepals (white, red, yellow, etc.) ovary superior to inferior, flowers without recurve and twist filaments
7. Perennial herbs with leaves alternate, opposite or verticillate, cuticular wax with transversally ridged rodlets.....**Liliaceae**
7. Perennial herbs often with basal sheathing leaves, cuticular wax with parallel platelets
8. Calcium oxalate crystals lacking, raphides absent, androecial nectaries, stigma wet.....**Colchicaceae**
8. Calcium oxalate crystals present (raphides, cuboidal, and styloid types), raphides present, stigma dry/wet, tepal nectaries.....**Melanthiaceae**

4.4. Morphological evolution in Liliales

The evolution of morphology revealed both conserved and derived features among Liliales which may have enabled dispersal into novel environments (Fig. 5, Supplementary Table S4). Although only two (i.e., nectaries on tepals and extrorse anthers) out of twelve features appear to be ancestral characters, the remaining ten characters revealed various putatively adaptive histories among members of Liliales. For example, different rootstock types evolved from a rhizome (i.e., bulbs in Liliaceae, Melanthiaceae; corms in Colchicaceae; and tuber-like roots in Corsiaceae, Colchicaceae), which aid plants to tolerance the xeric condition. Also, the fruit type changed from capsules to berries (i.e., in Smilacaceae, Petermanniaceae, and Colchicaceae). Therefore, new seed dispersers for fleshy fruits might be attracted, such as birds, mammals, and lizards (Schupp, 1993; Schupp et al., 2010; Viana et al., 2016; Wotton et al., 2016). Although having similar morphological features such as climbing vines and berry fruits that might help a species to survive in dense forests, the monotypic family Petermanniaceae contrasts with the situation in Smilacaceae. Smilacaceae are one of the biggest families in Liliales (approximately 200 species) with a cosmopolitan distribution (Angiosperm Phylogeny Group, 2016). Also correlated is the shift to dioecy in Smilacaceae, which prevents self-pollination and maintains/increases genetic variation, putatively assisting plants to adapt to new environments. The colorful fruits (orange-red to blue-black berry) of Smilacaceae that are probably bird-dispersed, might have contributed to its cosmopolitan. Phylogenetically, Smilacaceae are sister to Liliaceae (Figs. 3, 4); however, there are not many shared morphologies among these two families (i.e., trimerous flowers with nectaries on tepals and extrorse/introrse anthers). Within Liliaceae, various features were defined to support the division into four subfamilies (Kim and Kim, 2018). There findings revealed a high rate of morphological evolution of Liliaceae in comparison with Smilacaceae and a question on the close relationship between Smilacaceae and Liliaceae instead of Philesiaceae. Further studies with new data sources should be conducted to confirm the relationship between Smilacaceae and Liliaceae because the support values for these two families were not always 100% in the present study.

The early split of the mycoheterotrophic Corsiaceae in the Liliales

resulted in significant gene content and order differentiation in their plastid genomes and distinct morphologies compared to other families (Fig. 5, Table 1, Supplementary Table S5). The family includes two genera and comparative few species which have symbiosis with different fungi (Domínguez and Sérsic, 2004; Domínguez et al., 2009; Renny et al., 2017). Within the two species of *Corsia* sampled here, gene content is divergent, suggesting high plastid genome diversity in 24 species of this genus. Further studies on *Corsia* should be performed to enhance understanding of mycoheterotrophy.

5. Conclusions

Based on our comprehensive comparative plastome analysis of Liliales, we discovered that mycoheterotrophic Corsiaceae have significantly different plastid DNA sequences and organization, differing from each other as well as from their photosynthetic relatives. Plastome evolution is ongoing in this family, making it an ideal system to study this phenomenon. Otherwise, the plastid genomes of Liliales are all highly similar in structure and content among the families. However, some variations were detected (i.e., different type of LSC-IS-SSC junctions, deletion of coding regions, and inversion), which are useful for studies of molecular markers, phylogeny, and population genetics. The mycoheterotrophic Corsiaceae exhibits strong variation in plastome structure due to the loss or reduction of most photosynthesis related genes and this is associated with morphology in which plants are reduced in size and leafless.

CRedit authorship contribution statement

Hoang Dang Khoa Do: Data curation, Formal analysis, Investigation, Visualization, Writing - original draft. **Changkyun Kim:** Formal analysis, Investigation, Writing - original draft. **Mark W. Chase:** Resources, Validation, Writing - review & editing. **Joo-Hwan Kim:** Conceptualization, Supervision, Project administration, Writing - review & editing.

Acknowledgments

We would like to thank Sang-Chul Kim of Gachon University (South Korea), Anett Krämer at the Botanische Gärten der Universität Bonn (Germany), Peter Brownless at Royal Botanic Garden Edinburgh (United Kingdom), Botanical Garden Maise (Belgium), Garden of Auckland (New Zealand), Bruce Gray at the Australian Tropical Herbarium (Australia) and Ian Christie at the Scottish Rock Garden Club (United Kingdom), Jérôme Munzinger and Vanessa Hequet at the Herbarium of IRD Nouméa (NOU; New Caledonia, France) for collecting and providing the plant material for this study. This work was supported by the National Research Foundation of Korea (NRF) Grant Fund (NRF-2017R1D1A1B06029326; MEST 2010-0029131) and Scientific Research of Korea National Arboretum (KNA) Grant Fund (KNA 1-2-13, 14-2).

Appendix A. Supplementary material

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.ympev.2020.106818>.

References

- Ahmed, I., Biggs, P.J., Matthews, P.J., Collins, L.J., Hendy, M.D., Lockhart, P.J., 2012. Mutational dynamics of aroid chloroplast genomes. *Genome Biol. Evol.* 4 (12), 1316–1323. <https://doi.org/10.1093/gbe/evs110>.
- Angiosperm Phylogeny Group, 2009. An update of the angiosperm phylogeny group classification for the orders and families of flowering plants: APG III. *Bot. J. Linn. Soc.* 161, 105–121. <https://doi.org/10.1111/j.1095-8339.2009.00996.x>.
- 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), 1–20. <https://doi.org/10.1111/boj.12385>.
- Angiosperm Phylogeny Group, 1998. An ordinal classification for the families of flowering plants. *Ann. Mo. Bot. Gard.* 85, 531–553. <https://doi.org/10.2307/2992015>.
- Barnard-Kubow, K.B., McCoy, M.A., Galloway, L.F., 2017. Biparental chloroplast inheritance leads to rescue from cytonuclear incompatibility. *New Phytol.* 213, 1466–1476. <https://doi.org/10.1111/nph.14222>.
- Barrett, C.F., Baker, W.J., Comer, J.R., Conran, J.G., Lahmeyer, S.C., Leebens-Mack, J.H., Li, J., Lim, G.S., Mayfield-Jones, D.R., Perez, L., Medina, J., Pires, J.C., Santos, C., Stevenson, D., Zomlefer, W.B., Davis, J.I., 2016. Plastid genomes reveal support for deep phylogenetic relationships and extensive rate variation among palms and other commelinid monocots. *New Phytol.* 209, 855–870. <https://doi.org/10.1111/nph.13617>.
- Bendich, A.J., 2004. Circular chloroplast chromosomes: the grand illusion. *Plant Cell* 16, 1661–1666. <https://doi.org/10.1105/tpc.160771>.
- Bodin, S.S., Kim, J.S., Kim, J.H., 2013. Complete Chloroplast Genome of *Chionographis japonica* (Willd.) Maxim. (Melanthiaceae): Comparative Genomics and Evaluation of Universal Primers for Liliales. *Plant Mol. Biol. Rep.* 31 (6), 1407–1421. <https://doi.org/10.1007/s11105-013-0616-x> (2013).
- Cauz-Santos, L.A., Munhoz, C.F., Rodde, N., Cauet, S., Santos, A.A., Penha, H.A., Dornelas, M.C., Varani, A.M., Oliveira, G.C., Bergès, H., Vieira, M.L., 2017. The chloroplast genome of *Passiflora edulis* (passifloraceae) assembled from long sequence reads: structural organization and phylogenomic studies in malpighiales. *Front. Plant Sci.* 8, 334. <https://doi.org/10.3389/fpls.2017.00334>.
- Chacón, J., de Assis, M.C., Meerow, A.W., Renner, S.S., 2012. From east gondwana to central america: historical biogeography of the alstroemeriaceae. *J. Biogeogr.* 39, 1806–1818. <https://doi.org/10.1111/j.1365-2699.2012.02749.x>.
- Chase, M.W., Duvall, M.R., Hills, H.G., Conran, J.G., Cox, A.V., Eguarte, L.E., Hartwell, J., Fay, M.F., Chaddick, L.R., Cameron, K.M., Hoot, S., 1995. Molecular phylogenetics of Liliales in Rudall, P.J., Cribb, P.J., Cutler, D.F., Humphries, C.J. (Eds) *Monocotyledons: systematics and evolution*, pp. 109–137. Kew Royal Botanic Gardens, UK.
- Chase, M.W., Fay, M.F., Devey, D.S., Maurin, O., Rønsted, N.J., Pillon, Y., Petersen, G., Seberg, O., Tamura, M.N., Asmussen, C.B., Hilu, K., Borsch, T., Davis, J.I., Stevenson, D.W., Pires, J.C., Givnish, T.J., Sytsma, K.J., McPherson, M.M., Graham, S.W., Rai, H.S., 2006. Multi-gene analyses of monocot relationships: a summary. In: Columbus, J.T., Friar, E.A., Porter, J.M., Prince, L.M., Simpson, M.G. (Eds.), *Monocots: Comparative Biology and Evolution* (excluding Poales). Rancho Santa Ana Botanic Garden, Claremont, CA, pp. 63–75.
- Choi, I.S., Choi, B.H., 2017. The distinct plastid genome structure of *Maackia fauriei* (Fabaceae: Papilionoideae) and its systematic implications for genistoids and tribe Sophoreae. *PLoS ONE* 12 (4), e0173766. <https://doi.org/10.1371/journal.pone.0173766>.
- Christin, P.A., Spriggs, E., Osborne, C.P., Strömberg, C.A., Salamin, N., Edwards, E.J., 2014. Molecular dating, evolutionary rates, and the age of the grasses. *Syst. Biol.* 63 (2), 153–165. <https://doi.org/10.1093/sysbio/syt072>.
- Chun, E.H.L., Vaughan, M.H., Rich, A., 1963. The isolation and characterization of DNA associated with chloroplast preparations. *J. Mol. Biol.* 7, 130–141.
- Cosner, M.E., Raubeson, L.A., Jansen, R.K., 2004. Chloroplast DNA rearrangements in Campanulaceae: phylogenetic utility of highly rearranged genomes. *BMC Evol. Biol.* 4, 27. <https://doi.org/10.1186/1471-2148-4-27>.
- Cronquist, A., 1981. *An integrated system of classification of flowering plants*. Columbia University Press.
- Cusimano, N., Wicke, S., 2016. Massive intracellular gene transfer during plastid genome reduction in nongreen Orobanchaceae. *New Phytol.* 210, 680–693. <https://doi.org/10.1111/nph.13784>.
- Dahlgren, R.M.T., Clifford, H.T., Yeo, P.F., 1985. *The families of the monocotyledons*. Springer Verlag, Berlin.
- Daniell, H., Lin, C.S., Yu, M., Chang, W.J., 2016. Chloroplast genomes: diversity, evolution, and applications in genetic engineering. *Genome Biol.* 17, 134. <https://doi.org/10.1186/s13059-016-1004-2>.
- Darling, A.C.E., 2004. Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res.* 14 (7), 1394–1403. <https://doi.org/10.1101/gr.2289704>.
- Do H. D. K., Kim J.-H., 2017. A Dynamic Tandem Repeat in Monocotyledons Inferred from a Comparative Analysis of Chloroplast Genomes in Melanthiaceae. *Front. Plant Sci.* 8, 693. <https://doi.org/10.3389/fpls.2017.00693>.
- Do, H.D.K., Kim, J.S., Kim, J.H., 2013. Comparative genomics of four Liliales families inferred from the complete chloroplast genome sequence of *Veratrum patulum*. O. Loes. (Melanthiaceae). *Gene* 530 (2), 229–235. <https://doi.org/10.1016/j.gene.2013.07.100>.
- Do, H.D.K., Kim, J.S., Kim, J.H., 2014. A *trnL* CAU triplication event in the complete chloroplast genome of *Paris verticillata* M.Bieb. (Melanthiaceae, Liliales). *Genome Biol. Evol.* 6 (7), 1699–1706. <https://doi.org/10.1093/gbe/evu138>.
- Domínguez, L.S., Melville, L., Sérsic, A., Faccio, A., Peterson, R.L., 2009. The mycoheterotroph, *Arachnitis uniflora*, has a unique association with arbuscular mycorrhizal fungi. *Botany* 87, 1198–1208. <https://doi.org/10.1139/B09-081>.
- Domínguez, L.S., Sérsic, A., 2004. The southernmost myco-heterotrophic plant, *Arachnitis uniflora*: root morphology and anatomy. *Mycologia* 96, 1143–1151. <https://doi.org/10.1080/15572536.2005.11832912>.
- Dong, W., Liu, H., Xu, C., Zuo, Y., Chen, Z., Zhou, S., 2014. A chloroplast genomic strategy for designing taxon specific DNA mini- barcodes: a case study on ginsengs. *BMC Genet.* 15, 138. <https://doi.org/10.1186/s12863-014-0138-z>.
- Dong, W., Xu, C., Li, C., Sun, J., Zuo, Y., Shi, S., Cheng, T., Guo, J., Zhou, S., 2015. ycf1, the most promising plastid DNA barcode of land plants. *Sci. Rep.* 5, 8348. <https://doi.org/10.1038/srep08348>.
- Downie, S.R., Palmer, J.D., 1992. *Use of chloroplast DNA rearrangements in*

- reconstructing plant phylogeny in. In: Soltis, E.S., Soltis, D.E., Doyle, J.J. (Eds.), *Molecular systematics of plants*. Chapman and Hall, pp. 14–35.
- Downie, S.R., Jansen, R.K., 2015. A comparative analysis of whole plastid genomes from the apiales: expansion and contraction of the inverted repeat, mitochondrial to plastid transfer of DNA, and identification of highly divergent noncoding regions. *Syst. Bot.* 40 (1), 336–351. <https://doi.org/10.1600/036364415X686620>.
- Doyle, J.J., Doyle, J.L., 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem. Bull.* 19, 11–15.
- Doyle, J.J., Doyle, J.L., Palmer, J.D., 1995. Multiple independent losses of two genes and one intron from legume chloroplast genomes. *Syst. Bot.* 20, 272–294. <https://doi.org/10.2307/2419496>.
- Du, Y., Bi, Y., Yang, F., Zhang, M., Chen, X., Xue, J., Zhang, X., 2017. Complete chloroplast genome sequences of *Lilium*: insights into evolutionary dynamics and phylogenetic analyses. *Sci. Rep.* 7, 5751. <https://doi.org/10.1038/s41598-017-06210-2>.
- Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucl. Acids Res.* 32 (5), 1792–1797. <https://doi.org/10.1093/nar/gkh240>.
- Fay, M.F., Chase, M.W., Ronsted, N., Devey, D.S., Pillon, Y., Pires, J.C., Petersen, G., Seberg, O., Davis, J.L., 2006. Phylogenetics of Liliales: summarized evidence from combined analyses of five plastid and one mitochondrial loci in Columbus, J.T., Friar, E.A., Porter, J.M., Prince, L.M., Simpson, M.G. (Eds) *Monocots: comparative biology and evolution*, pp. 559–565. Rancho Santa Ana Botanic Garden.
- Farris, J. D., Kallersjö, M., Kluge, A.G., Bult, C. 1995. Constructing a significance test for incongruence. *Syst. Biol.* 44:570–572.
- Felsenstein, J., 1985. Confidence limits on phylogenies: an approach using bootstrap. *Evolution* 39, 783–791. <https://doi.org/10.2307/2408678>.
- Folk, R.A., Mandel, J.R., Freudenstein, J.V., 2017. Ancestral gene flow and parallel organellar genome capture result in extreme phylogenomic discord in a lineage of angiosperms. *Syst. Biol.* 66 (3), 320–337. <https://doi.org/10.1093/sysbio/syw083>.
- Foster, C.S.P., Sauquet, H., van der Merwe, M., McPherson, H., Rossetto, M., Ho, S.Y.W., 2017. Evaluating the impact of genomic data and priors on bayesian estimates of the angiosperm evolutionary timescale. *Syst. Biol.* 66 (3), 338–351. <https://doi.org/10.1093/sysbio/syw086>.
- Furness, C.A., Thomas, G., Rudall, P.J., 2015. Pollen structure and diversity in Liliales. *Int. J. Plant Sci.* 176, 697–723. <https://doi.org/10.1086/682211>.
- Givnish, T.J., Zuluaga, A., Marques, I., Lam, V.K., Gomez, M.S., Iles, W.J., Ames, M., Spalink, D., Moeller, J.R., Briggs, B.G., Lyon, S.P., Stevenson, D.W., Zomlefer, W., Graham, S.W., 2016. Phylogenomics and historical biogeography of the monocot order Liliales: out of Australia and through Antarctica. *Cladistics* 32, 581–605. <https://doi.org/10.1111/clc.12153>.
- Goremykin, V.V., Nikiforova, S.V., Cavaliere, D., Pindo, M., Lockhart, P., 2015. The root of flowering plants and total evidence. *Syst. Biol.* 64 (5), 879–891. <https://doi.org/10.1093/sysbio/syv028>.
- Graham, S.W., Lam, V.K.Y., Merckx, V.S.F.T., 2017. Plastomes on the edge: the evolutionary breakdown of mycoheterotroph plastid genomes. *New Phytol.* 214, 48–55. <https://doi.org/10.1111/nph.14398>.
- Graham, S.W., Olmstead, R.G., 2000. Utility of 17 chloroplast genes for inferring the phylogeny of the basal angiosperms. *Am. J. Bot.* 87, 1712–1730. <https://doi.org/10.2307/2656749>.
- Guisinger, M.M., Chumley, T.W., Kuehl, J.V., Boore, J.L., Jansen, R.K., 2010. Implications of the plastid genome sequence of *Typha* (typhaceae, poales) for understanding genome evolution in poaceae. *J. Mol. Evol.* 70, 149. <https://doi.org/10.1007/s00239-009-9317-3>.
- Haberle, R.C., Fourcade, H.M., Boore, J.L., Jansen, R.K., 2008. Extensive rearrangements in the chloroplast genome of *Trachelium caeruleum* are associated with repeats and tRNA genes. *J. Mol. Evol.* 66, 350. <https://doi.org/10.1007/s00239-008-9086-4>.
- Huang, Y., Li, X., Yang, Z., Yang, C., Yang, J., Ji, Y., 2016. Analysis of complete chloroplast genome sequences improves phylogenetic resolution in paris (melanthiaceae). *Front. Plant Sci.* 7, 1797. <https://doi.org/10.3389/fpls.2016.01797>.
- Jansen, R.K., Cai, Z., Raubeson, L.A., Daniell, H., dePamphilis, C.W., Leebens-Mack, J., Muller, K.F., Guisinger-Bellian, M., Haberle, R.C., Hansen, A.K., Chumley, T.W., Lee, S.-B., Peery, R., McNeal, J.R., Kuehl, J.V., Boore, J.L., 2007. Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns. *Proc. Natl. Acad. Sci.* 104 (49), 19369–19374. <https://doi.org/10.1073/pnas.0709121104>.
- Katayama, H., Ogihara, Y., 1996. Phylogenetic affinities of the grasses to other monocots as revealed by molecular analysis of chloroplast DNA. *Curr. Genet.* 29, 572–581. <https://doi.org/10.1007/BF02426962>.
- Katoh, A., Lee, K.S., Fukuzawa, H., Ohshima, K., Ogawa, T., 1996a. *cemA* homologue essential to CO₂ transport in the cyanobacterium in *Synechocystis* PCC 6803. *PNAS* 93, 4006–4010. <https://doi.org/10.1073/pnas.93.9.4006>.
- Katoh, A., Sonoda, M., Katoh, H., Ogawa, T., 1996b. Absence of light-induced proton extrusion in a *cotA*-less mutant of *Synechocystis* sp. strain PCC 6803. *J. Bacteriol.* 178, 5452–5455. <https://doi.org/10.1128/jb.178.18.5452-5455.1996>.
- Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P., Drummond, A., 2012. Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28 (12), 1647–1649. <https://doi.org/10.1093/bioinformatics/bts199>.
- Kelchner, S.A., 2000. The evolution of non-coding chloroplast DNA and its application in plant systematics. *Ann. Mo. Bot. Gard.* 87, 482–498. <https://doi.org/10.2307/2666142>.
- Kim, C., Kim, S.-C., Kim, J.-H., 2019. Historical biogeography of melanthiaceae: a case of out-of-north america through the bering land bridge. *Front. Plant Sci.* <https://doi.org/10.3389/fpls.2019.00396>.
- Kim, J.S., Kim, J.-H., 2018. Updated molecular phylogenetic analysis, dating and biogeographical history of the lily family (Liliaceae: Liliales). *Bot. J. Linn. Soc.* 187 (4), 579–593. <https://doi.org/10.1093/botlinnean/boy031>.
- Kim, Jung Sung, Hong, Jeong-Ki, Chase, Mark W., Fay, Michael F., Kim, Joo-Hwan, 2013. Familial relationships of the monocot order Liliales based on a molecular phylogenetic analysis using four plastid loci: matK, rbcL, atpB and atpF-H. *Bot. Bot. J. Linn. Soc.* 172 (1), 5–21. <https://doi.org/10.1111/boj.2013.172.issue-110.1111/boj.12039>.
- Kim, J.S., Kim, J.H., 2013. Comparative Genome Analysis and Phylogenetic Relationship of Order Liliales Insight from the Complete Plastid Genome Sequences of Two Lilies (*Lilium longiflorum* and *Alstroemeria aurea*). *PLoS ONE*, 8(6), e68180. <https://doi.org/10.1371/journal.pone.0068180>.
- Kim, S.C., Kim, J.S., Kim, J.H., 2016. Insight into infrageneric circumscription through complete chloroplast genome sequences of two *Trillium* species. *AoB PLANTS*. 8 (plw015), 2016. <https://doi.org/10.1093/aobpla/plw015>.
- Knox, B.E., 2014. The dynamic history of plastid genomes in the Campanulaceae *sensu lato* is unique among angiosperms. *PNAS* 111 (30), 11097–11102. <https://doi.org/10.1073/pnas.1403363111>.
- Kurtz, S., 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, P., Lu, R.S., Xu, W.Q., Ohi-Toma, T., Cai, M.Q., Qiu, Y.X., Cameron, K.M., Fu, C.X., 2017. Comparative Genomics and phylogenomics of east asian tulips (*amara*, liliaceae). *Front. Plant Sci.* 8, 451. <https://doi.org/10.3389/fpls.2017.00451>.
- Librado, P., Rozas, J., 2009. DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25, 1451–1452. <https://doi.org/10.1093/bioinformatics/btp187>.
- Liu, J., Qi, Z., Zhao, Y.P., Fu, C.X., Xiang, Q.Y., 2012. Complete cpDNA genome sequence of *Smilax china* and phylogenetic placement of Liliales – Influences of gene partitions and taxon sampling. *Mol. Phyl. Evol.* 64 (3), 545–562. <https://doi.org/10.1016/j.ympev.2012.05.010>.
- Logacheva, M.D., Schelkunov, M.I., Nuraliev, M.S., Samigullin, T.H., Penin, A.A., 2014. The plastid genome of mycoheterotrophic monocot *Petrosavia stellaris* exhibits both gene losses and multiple rearrangements. *Genome Biol. Evol.* 6 (1), 238–246. <https://doi.org/10.1093/gbe/evu001>.
- 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. *Nucl. Acids Res.* 41, W575–81. <https://doi.org/10.1093/nar/gkt289>.
- Lu, R.S., Li, P., Qiu, Y.X., 2017. The complete chloroplast genomes of three *cardiocrinum* (liliaceae) species: comparative genomic and phylogenetic analyses. *Front. Plant Sci.* 7, 2054. <https://doi.org/10.3389/fpls.2016.02054>.
- Lu, R.-S., Li, P., Qiu, Y.-X., 2016. The complete chloroplast genomes of three *cardiocrinum* (liliaceae) species: comparative genomic and phylogenetic analyses. *Front. Plant Sci.* 7, 2054. <https://doi.org/10.3389/fpls.2016.02054>.
- Ma, P.-F., Zhang, Y.-X., Zeng, C.-X., Guo, Z.-H., Li, D.-H., 2014. Chloroplast phylogenomic analyses resolve deep-level relationships of an intractable bamboo tribe arundinarieae (poaceae). *Syst. Biol.* 63 (6), 933–950. <https://doi.org/10.1093/sysbio/syu054>.
- Ma, J., Yang, B., Zhu, W., Sun, L., Tian, J., Wang, X., 2013. The complete chloroplast genome sequence of *Mahonia bealei* (Berberidaceae) reveals a significant expansion of the inverted repeat and phylogenetic relationship with other angiosperms. *Gene* 528 (2), 120–131. <https://doi.org/10.1016/j.gene.2013.07.037>.
- Maddison, W.P., Maddison, D.R., 2017. Mesquite: a modular system for evolutionary analysis. Version 3, 31. <http://mesquiteproject.org>.
- Marcussen, T., Heier, L., Brysting, A.K., Oxelman, B., Jakobsen, K.S., 2015. From gene trees to a dated allopolyploid network: insights from the angiosperm genus *viola* (violaceae). *Syst. Biol.* 64 (1), 84–101. <https://doi.org/10.1093/sysbio/syu071>.
- Mardanov, A.V., Ravin, N.V., Kuznetsov, B.B., Samigullin, T.H., Antonov, A.S., Kolganova, T.V., Skyabin, K.G., 2008. Complete sequence of the duckweed (*Lemna minor*) chloroplast genome: structural organization and phylogenetic relationships to other angiosperms. *J. Mol. Evol.* 66 (6), 555–564. <https://doi.org/10.1007/s00239-008-9091-7>.
- Mennes, C.B., Lam, V.K., Rudall, P.J., Lyon, S.P., Graham, S.W., Smets, E.F., Merckx, V.S., 2015. Ancient gondwana break-up explains the distribution of the mycoheterotrophic family coriariaceae (Liliales). *J. Biogeogr.* 42, 1123–1136. <https://doi.org/10.1111/jbi.12486>.
- Meseguer, A.S., Lobo, J.M., Ree, R., Beerling, D.J., Sanmartín, I., 2015. Integrating fossils, phylogenies, and niche models into biogeography to reveal ancient evolutionary history: the case of *hypericum* (hypericaceae). *Syst. Biol.* 64 (2), 215–232. <https://doi.org/10.1093/sysbio/syu088>.
- Millen, R.S., Olmstead, R.G., Adams, K.L., Palmer, J.D., Lao, N.T., Heggie, L., Kavanagh, T.A., Hibberd, J.M., Gray, J.C., Morden, C.W., Calie, P.J., Jermin, L.S., Wolfe, K.H., 2001. Many parallel losses of *infA* from chloroplast DNA during angiosperm evolution with multiple independent transfers to the nucleus. *Plant Cell* 13 (3), 645–658. <https://doi.org/10.1105/tpc.13.3.645>.
- Morris, L.M., Duvall, M.R., 2010. The chloroplast genome of *Anomochloa marantoides* (Anomochloideae; Poaceae) comprises a mixture of grass-like and unique features. *Am. J. Bot.* 97, 620–627. <https://doi.org/10.3732/ajb.0900026>.
- Neyland, R., Hennigan, M., 2003. A phylogenetic analysis of large subunit (26S) ribosome DNA sequences suggests that the Corsiaceae are polyphyletic. *N. Z. J. Bot.* 41 (1), 1–11. <https://doi.org/10.1080/0028825X.2003.9512828>.
- Nguyen, P.A.T., Kim, J.S., Kim, J.H., 2015. The complete chloroplast genome of colchicine plants (*Colchicum autumnale* L. and *Gloriosa superba* L.) and its application for identifying the genus. *J. Planta.* 242, 223. <https://doi.org/10.1007/s00425-015-2303-7>.
- Nordenstam, B. 1998. Colchicaceae. In K. Kubitzki [ed.], *The families and genera of vascular plants*, vol. 3, pp.175–185. Springer-Verlag, Berlin, Germany.
- Patterson, T.B., Givnish, T.J., 2002. Phylogeny, concerted convergence, and phylogenetic

- niche conservatism in the core Liliales: insights from *rbcL* and *ndhF* sequence data. *Evolution* 56, 233–252. <https://doi.org/10.1111/j.0014-3820.2002.tb01334.x>.
- Petersen, G., Seberg, O., Davis, J.I., 2012. Phylogeny of the Liliales (Monocotyledons) with special emphasis on data partition congruence and RNA editing. *Cladistics* 29, 274–295. <https://doi.org/10.1111/j.1096-0031.2012.00427.x>.
- Petersen, G., Seberg, O., Davis, J.I., Goldman, D.H., Stevenson, S.W., Campbell, L.M., Michelangeli, F.A., Specht, C.D., Chase, M.W., Fay, M.F., Pires, J.C., Freudenstein, J.V., Hardy, C.R., Simmons, M.P., 2006. Mitochondrial data in monocot phylogenetics. *Aliso* 22, 52–62.
- Posada, D., 2008. jModelTest: Phylogenetic model averaging. *Mol. Biol. Evol.* 25, 1253–1256. <https://doi.org/10.1093/molbev/msn083>.
- Pouchon, C., Fernández, A., Nassar, J.M., Boyer, F., Aubert, S., Lavergne, S., Mavárez, J., 2018. Phylogenomic analysis of the explosive adaptive radiation of the Espeletia complex (asteraceae) in the tropical andes. *Syst. Biol.* 67 (6), 1041–1060. <https://doi.org/10.1093/sysbio/syy022>.
- Raman, G., Park, S.J., 2016. The complete chloroplast genome sequence of *ampelopsis*: gene organization, comparative analysis, and phylogenetic relationships to other angiosperms. *Front. Plant Sci.* 7, 341. <https://doi.org/10.3389/fpls.2016.00341>.
- Renny, M., Acosta, M.C., Cofré, N., Domínguez, L.S., Bidartondo, M.I., Sérsic, A.N., 2017. Genetic diversity patterns of arbuscular mycorrhizal fungi associated with the mycoheterotrophic *Arachnitis uniflora*. *Phil. (Corsiaceae). Ann Bot.* 119 (8), 1279–1294. <https://doi.org/10.1093/aob/mcx023>.
- Roalson, E.H., Roberts, W.R., 2016. Distinct processes drive diversification in different clades of gesneriaceae. *Syst. Biol.* 65 (4), 662–684. <https://doi.org/10.1093/sysbio/syw012>.
- Robbens, S., Derelle, E., Ferraz, C., Wuyts, J., Moreau, H., Van de Peer, Y., 2007. The complete chloroplast and mitochondrial DNA sequence of *ostreococcus tauri*: organelle genomes of the smallest eukaryote are examples of compaction. *Mol. Biol. Evol.* 24 (4), 956–968. <https://doi.org/10.1093/molbev/msm012>.
- Rolland, N., Dorne, A.J., Amoroso, G., Sültemeyer, D.F., Joyard, J., Rochaix, J.D., 1997. Disruption of the plastid *ycf10* open reading frame affects uptake of inorganic carbon in the chloroplast of *Chlamydomonas*. *EMBO J.* 16 (22), 6713–6726. <https://doi.org/10.1093/emboj/16.22.6713>.
- Ronquist, F., Huelsenbeck, J.P., 2003. MRBAYES 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19, 1572–1574.
- Rudall, P.J., Stobart, K.L., Hong, W.P., Conran, J.G., Furness, C.A., Kite, G., Chase, M.W., 2000. Consider the lilies: systematics of Liliales in Wilson, K.L., & Morrison, D. (Eds) *Monocots: systematics and evolution*, pp. 347–359. CSIRO Publishing.
- Sager, R., Ishida, M.R., 1966. Chloroplast DNA in *Chlamydomonas*. *PNAS* 50, 725–730.
- Sasaki, Y., Sekiguchi, K., Nagano, Y., Matsumo, R., 1993. Chloroplast envelope protein encoded by the chloroplast genome. *FEBS Lett.* 316, 93–98. [https://doi.org/10.1016/0014-5793\(93\)81743-J](https://doi.org/10.1016/0014-5793(93)81743-J).
- Schattner, P., Brooks, A.N., Lowe, T.M., 2005. The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. *Nucl. Acids Res.* 33, 686–689. <https://doi.org/10.1093/nar/gki366>.
- Schellkunov, M.I., Shtratnikova, V.Y., Nuraliev, M.S., Selosse, M.A., Penin, A.A., Logacheva, M.D., 2015. Exploring the limits for reduction of plastid genomes: a case study of the mycoheterotrophic orchids *epipogium aphyllum* and *epipogium roseum*. *Genome Biol. Evol.* 7 (4), 1179–1191. <https://doi.org/10.1093/gbe/evv019>.
- Schupp, E.W., Jordano, P., Gómez, J.M., 2010. Seed dispersal effectiveness revisited: a conceptual review. *New Phytol.* 188, 333–353. <https://doi.org/10.1111/j.1469-8137.2010.03402.x>.
- Schupp, E.W., 1993. Quantity, quality and the effectiveness of seed dispersal by animals. *Vegetatio* 107, 15. <https://doi.org/10.1007/BF00052209>.
- Seberg, O., Petersen, G., Davis, J.I., Pires, C., Stevenson, D.W., Chase, M.W., Fay, M.F., Devey, D.S., Jørgensen, T., Kenneth J. Sytsma, K.J., Pillon, Y., 2012. Phylogeny of the Asparagales based on three plastid and two mitochondrial genes. *American Journal of Botany*, 99(5), 875–889.
- Song, J.H., Yoon, C.Y., Do, H.D.K., Lee, W.B., Kim, J.H., 2016. The complete chloroplast genome sequence of *Lilium tsingtauense* Gilg (sect *Martagon*, Liliaceae). *Mitochondrial DNA Part B* 1 (1), 318–320. <https://doi.org/10.1080/23802359.2016.1172043>.
- Song, Y., Wang, S., Ding, Y., Xu, J., Li, M.F., Zhu, S., Chen, N., 2017. Chloroplast genomic resource of *Paris* for species discrimination. *Sci. Rep.* 7, 3427. <https://doi.org/10.1038/s41598-017-02083-7>.
- Stamatakis, A., Hoover, P., Rougemont, J., 2008. A rapid bootstrap algorithm for the RAxML Web Servers. *System Biol.* 57 (5), 758–771. <https://doi.org/10.1080/10635150802429642>.
- Sugiura, C., Kobayashi, Y., Aoki, S., Sugita, C., Sugita, M., 2003. Complete chloroplast DNA sequence of the moss *Physcomitrella patens*: evidence for the loss and relocation of *rpoA* from the chloroplast to the nucleus. *Nucl. Acids Res.* 31 (18), 5324–5331. <https://doi.org/10.1093/nar/gkg726>.
- Swofford, D.L., 2003. PAUP*. Phylogenetic analysis using parsimony (* and other methods). Version 4. Sinauer Associates, Sunderland.
- Tewari, K.K., Wildman, S.G., 1966. Chloroplast DNA from tobacco leaves. *Science* 153, 1269–1271.
- Trifinopoulos, J., von Nguyen, L.-T., Haeseler, A., Minh, B.Q., 2016. W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. *Nucl. Acids Res.* 44, W232–W235. <https://doi.org/10.1093/nar/gkw256>.
- Ueda, M., Nishikawa, T., Fujimoto, M., Takanashi, H., Arimura, S., Tsutsumi, N., Kadowaki, K., 2008. Substitution of the gene for chloroplast RPS16 was assisted by generation of a dual targeting signal. *Mol. Biol. Evol.* 25 (8), 1566–1575. <https://doi.org/10.1093/molbev/msn102>.
- Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B.C., Remm, M., Rozen, S.G., 2012. Primer3 – new capabilities and interfaces. *Nucl. Acids Res.* 40 (15), e115. <https://doi.org/10.1093/nar/gks598>.
- Viana, D.S., Gangoso, L., Bouten, W., Figuerola, J., 2016. Overseas seed dispersal by migratory birds. In: *Proc. R. Soc. B* 283, pp. 20152406. <https://doi.org/10.1098/rspb.2015.2406>.
- Vinnersten, A., Bremer, K., 2001. Age and biogeography of major clades in Liliales. *Am. J. Bot.* 88, 1695–1703. <https://doi.org/10.2307/3558415>.
- Wang, M., Cui, L., Feng, K., Deng, P., Du, X., Wan, F., Weining, S., Nie, X., 2015. Comparative analysis of asteraceae chloroplast genomes: structural organization, RNA editing and evolution. *Plant Mol. Biol. Rep.* 33, 1526. <https://doi.org/10.1007/s11105-015-0853-2>.
- Wang, R.J., Cheng, C.L., Chang, C.C., Wu, C.L., Su, T.M., Chaw, S.M., 2008. Dynamics and evolution of the inverted repeat-large single copy junctions in the chloroplast genomes of monocots. *BMC Evol. Biol.* 8, 36. <https://doi.org/10.1186/1471-2148-8-36>.
- Wang, Wei, Li, HongLei, Chen, ZhiDuan, 2014. Analysis of plastid and nuclear DNA data in plant phylogenetics—evaluation and improvement. *Sci. China Life Sci.* 57 (3), 280–286. <https://doi.org/10.1007/s11427-014-4620-7>.
- Wotton, Debra M, Drake, Donald R, Powlesland, Ralph G, Ladley, Jenny J, 2016. The role of lizards as seed dispersers in New Zealand. *J. R. Soc. New Zealand* 46 (1), 40–65. <https://doi.org/10.1080/03036758.2015.1108924>.
- Yang, M., Zhang, X., Liu, G., Yin, Y., Chen, K., Yun, Q., Zhao, D., Al-Mssallem, I.S., Yu, J., 2010. The complete chloroplast genome sequence of date palm (*Phoenix dactylifera* L.). *PLoS ONE* 5 (9), e12762. <https://doi.org/10.1371/journal.pone.0012762>.
- Yao, X., Tang, P., Li, Z., Li, D., Liu, Y., Huang, H., 2015. The first complete chloroplast genome sequences in actinidiaceae: genome structure and comparative analysis. *PLoS One* 10 (6), e0129347. <https://doi.org/10.1371/journal.pone.0129347>.