

**Evidence of mitochondrial DNA in the chloroplast genome of
Convallaria keiskei and its subsequent evolution in the
Asparagales**

G. Raman, S. Park, E. M. Lee & S. J. Park*

Department of Life Sciences, Yeungnam University, Gyeongsan, Gyeongsan-buk, Republic of Korea

*Corresponding author: S. J. Park, Department of Life Sciences, Yeungnam University, Gyeongsan, Gyeongsan-buk, Republic of Korea 38541; Tel.: +82 (53) 810-2377; Fax: +82 (53) 810-4618; Email: sjpark01@ynu.ac.kr

Supplementary information

Supplementary Figures

Supplementary Fig. S1. Gene map of (A) *Convallaria keiskei* and (B) *Liriope spicata*. Genes lying outside of the outer layer circle are transcribed in a counterclockwise direction, whereas genes inside this circle are transcribed in a clockwise direction. The colored bars indicate known protein-coding genes, tRNA genes, and rRNA genes. The dashed darker gray area in the inner circle denotes GC content, while the lighter gray area indicates genome AT content. LSC, large-single-copy; SSC, small-single-copy; IR, inverted repeat.

Supplementary Fig. S2. Sequence alignment of 16 chloroplast genomes in the Asparagales family performed using the mVISTA program with *Convallaria keiskei* as a reference. The top gray arrow shows genes in order (Transcriptional direction) and the position of each gene. A 70% cut-off was used for the plots. The Y-scale represents the percent identity between 50±100%. Red and blue areas indicate intergenic and genic regions, respectively. 1. *Convallaria keiskei*, 2. *Liriope spicata*, 3. *Maianthemum bicolor*, 4. *Nolina atopocarpa*, 5. *Polygonatum cyrtoneura*, 6. *Asparagus officinalis*, 7. *Corydalis indivisa*, 8. *Esutrepheus latifolius*, 9. *Aphyllanthes monspeliensis*, 10. *Mila biflora*, 11. *Anthericum ramosum*, 12. *Yucca brevifolia*, 13. *Agapanthus coddii*, 14. *Allium sativum*, 15. *Xanthorrhoeae preissii* and 16. *Aloe vera*.

Supplementary Fig. S3. Percentages of variable sites in protein-coding regions across the five Nolinoideae sub-family chloroplast genomes.

Supplementary Fig. S4: The schematic diagram represents the position of mitochondrial DNA in the *C. keiskei* cp genome.

Supplementary Fig. S5: Blast hit shows the sequence homology between mitochondrial DNA insert of *Convallaria keiskei* cp genome and *rpl10* of *Cocos nucifera* mitochondrial genome.

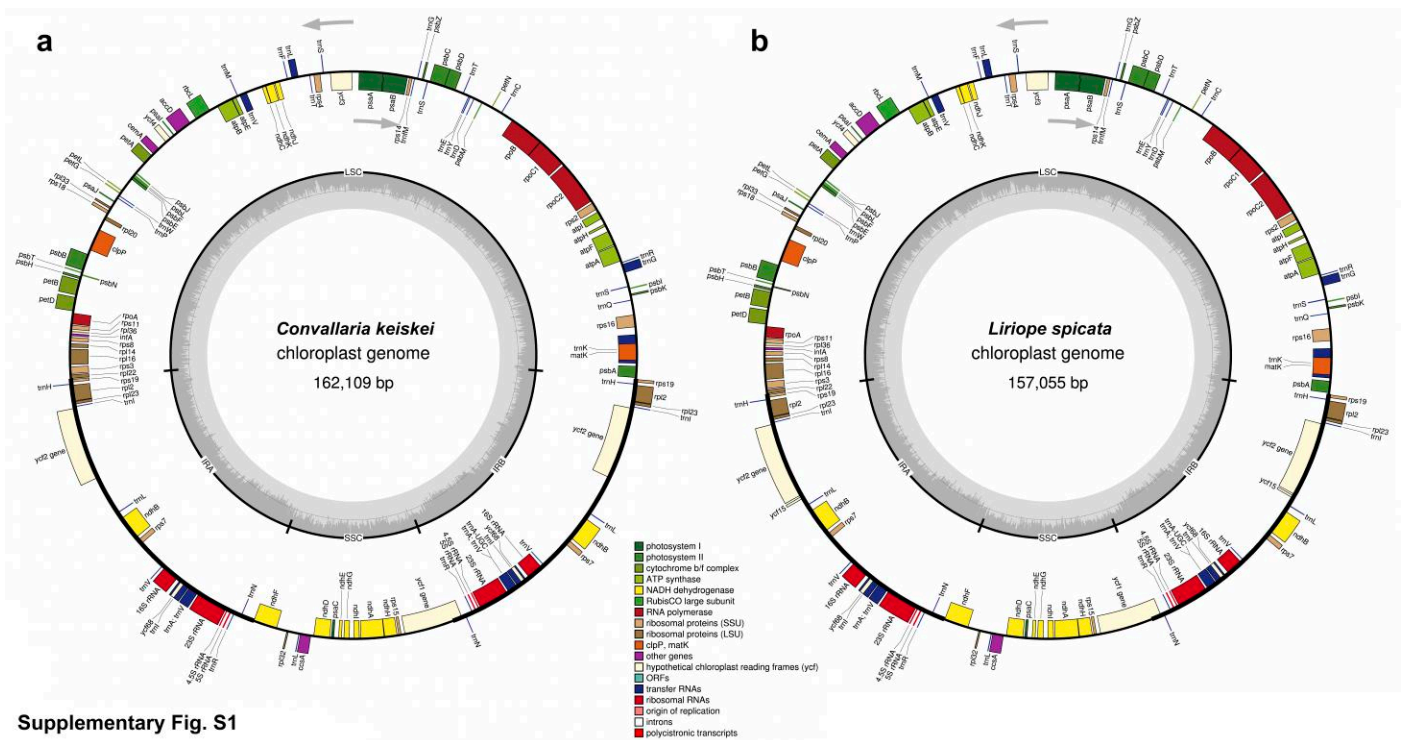
Supplementary Tables

Supplementary Table S1. Summary of chloroplast genome characteristics of *Convallaria keiskei* and *Liriope spicata*.

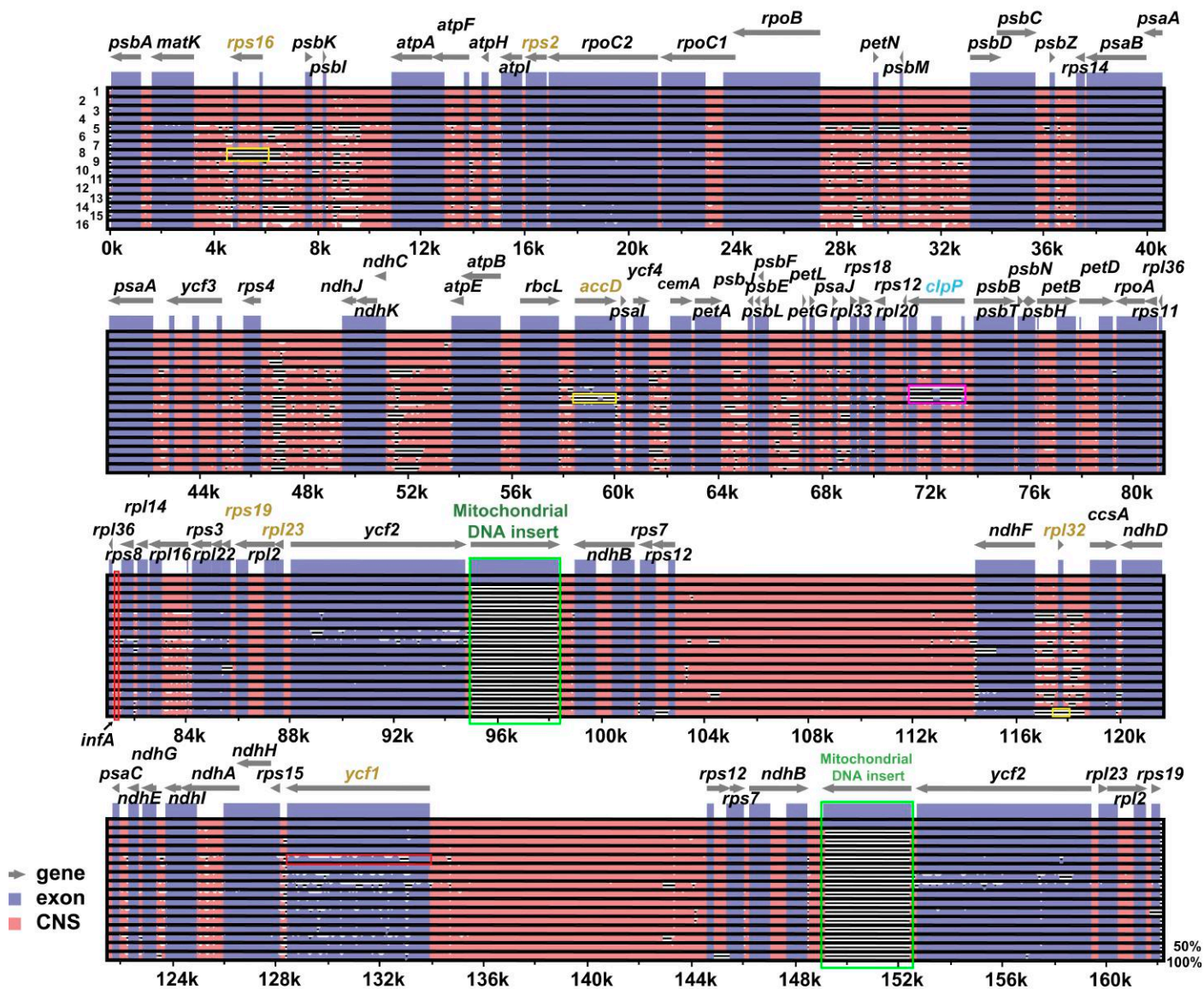
Supplementary Table S2. List of genes present in the cp genomes of *Convallaria keiskei* and *Liriope spicata*.

Supplementary Table S3. The list of pseudogenes, intron deletion gene, and gene inversion in the Asparagales order of plants.

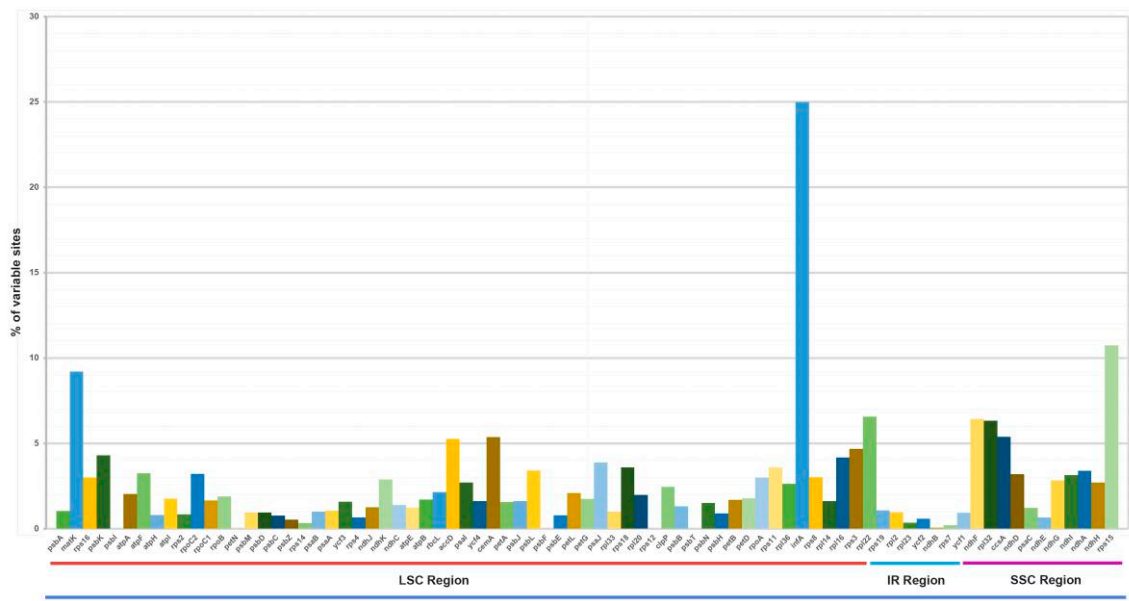
Supplementary Table S4: List of primers used in this study.



Supplementary Fig. S1

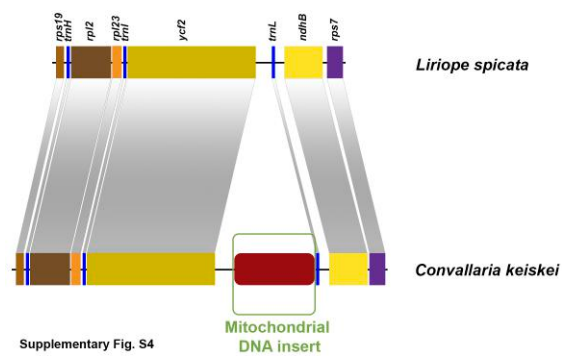


Supplementary Fig. S2



Supplementary Fig. S3

Protein-coding genes



Supplementary Fig. S4



Supplementary Table S1. Summary of chloroplast genome characteristics of *Convallaria keiskei* and *Liriope spatia*

Genome features	<i>Convallaria keiskei</i>	<i>Liriope spatia</i>
Size (bp)	162,109	157,055
LSC length (bp)	85,344	85,374
SSC length (bp)	18,487	18,727
IR length (bp)	29,139	26,477
Total number of genes	131	131
Number of genes	112	112
Protein-coding genes	78+7	78+7
tRNA genes	30+8	30+8
rRNA genes	4+4	4+4
Number of genes duplicated in IR	19	19
GC content (overall %)	37.9	37.6
GC content - LSC (%)	35.7	35.6
GC content - SSC (%)	31.4	31.2
GC content - IR (%)	43.2	43.0

1 **Supplementary Table S2.** List of genes present in the cp genomes of *Convallaria keiskei* and *Liriope spatia*.

Category	Group of genes	Name of genes				
RNA genes	Ribosomal RNA genes	<i>rrn4.5^a</i>	<i>rrn5^a</i>	<i>rrn16^a</i>	<i>rrn23^a</i>	
	Transfer RNA genes	<i>trnA-UGC^{a, b}</i>	<i>trnC-GCA</i>	<i>trnD-GUC</i>	<i>trnE-UUC</i>	<i>trnF-GAA</i>
		<i>trnM-CAU</i>	<i>trnG-GCC^b</i>	<i>trnG-UCC</i>	<i>trnH-GUG^a</i>	<i>trnI-CAU^a</i>
		<i>trnI-GAU^{a, b}</i>	<i>trnK-UUU</i>	<i>trnL-CAA^a</i>	<i>trnL-UAA^b</i>	<i>trnL-UAG</i>
		<i>trnM-CAU</i>	<i>trnN-GUU^a</i>	<i>trnP-UGG</i>	<i>trnQ-UUG</i>	<i>trnR-ACG^a</i>
		<i>trnR-UCU</i>	<i>trnS-GCU</i>	<i>trnS-GGA</i>	<i>trnS-UGA</i>	<i>trnT-GGU</i>
		<i>trnT-UGU</i>	<i>trnV-GAC^a</i>	<i>trnV-UAC^b</i>	<i>trnW-CCA</i>	<i>trnY-GUA</i>
Photosynthetic genes	Subunits of photosystem I	<i>psaA</i>	<i>psaB</i>	<i>psaC</i>	<i>psaI</i>	<i>psaJ</i>
		<i>ycf3^c</i>	<i>ycf4</i>			
	Subunits of photosystem II	<i>psbA</i>	<i>psbB</i>	<i>psbC</i>	<i>psbD</i>	<i>psbE</i>
		<i>psbF</i>	<i>psbH</i>	<i>psbI</i>	<i>psbJ</i>	<i>psbK</i>
		<i>psbL</i>	<i>psbM</i>	<i>psbN</i>	<i>psbT</i>	<i>psbZ</i>
	Subunits of cytochrome	<i>petA</i>	<i>petB^b</i>	<i>petD^b</i>	<i>petG</i>	<i>petL</i>
		<i>petN</i>				
	Subunits of ATP synthase	<i>atpA</i>	<i>atpB</i>	<i>atpE</i>	<i>atpF^b</i>	<i>atpH</i>
		<i>atpI</i>				
	The large subunit of Rubisco	<i>rbcL</i>				
	Subunits of NADH dehydrogenase	<i>ndhA^b</i>	<i>ndhB^{a, b}</i>	<i>ndhC</i>	<i>ndhD</i>	<i>ndhE</i>
		<i>ndhF</i>	<i>ndhG</i>	<i>ndhH</i>	<i>ndhI</i>	<i>ndhJ</i>
		<i>ndhK</i>				
ATP-dependent protease subunit P	<i>clpP^c</i>					
Chloroplast envelope membrane protein	<i>cemA</i>					
Transcription and translation-related genes	Small subunit of the ribosome	<i>rps2</i>	<i>rps3</i>	<i>rps4</i>	<i>rps7^a</i>	<i>rps8</i>
		<i>rps11</i>	<i>rps12^{a, c, d}</i>	<i>rps14</i>	<i>rps15</i>	<i>rps16^e</i>
		<i>rps18</i>	<i>rps19</i>			
	Large subunit of ribosome	<i>rpl2^a</i>	<i>rpl14</i>	<i>rpl16^b</i>	<i>rpl20</i>	<i>rpl22</i>
		<i>rpl23</i>	<i>rpl32</i>	<i>rpl33</i>	<i>rpl36</i>	
	DNA-dependent RNA polymerase	<i>rpoA</i>	<i>rpoB</i>	<i>rpoC1^b</i>	<i>rpoC2</i>	
	Translational initiation factor	<i>infA^e</i>				

Other genes	Maturase	<i>matK</i>				
	Subunit of acetyl-CoA	<i>accD</i>				
	C-type cytochrome synthesis gene	<i>ccsA</i>				
	Component of TIC complex	<i>ycf1^{a,f}</i>				
Genes of unknown function	Hypothetical proteins	<i>ycf2^a</i> ,				

2

3 ^a - Two gene copies in IRs; ^b - Gene containing a single intron; ^c - Gene containing two introns; ^d - Gene divided into two independent
4 transcription units; ^e - Pseudogene; ^f - Two copies - One intact gene and another is a pseudogene.

Supplementary Table S3. List of deleted genes in Asparagales

Sl. No.	Gene Name	Species Name	Sub-Family Name	GenBank Accession number
1.	<i>rpl32</i>	<i>Aloe vera</i>	Asphodeloideae	NC_035506
2.	<i>rpl23</i>	<i>Eustrephus latifolius</i>	Lomandroideae	NC_025305
3.	<i>rps19</i> - IRa	<i>Polygonatum cyrtoneuma</i>	Nolinoideae	NC_028429
4.	<i>rps19</i> - Both	<i>Hesperaloe parviflora</i>	Agavoideae	NC_032703
5.	<i>rps19</i> - Both	<i>Schoenolirion croceum</i>	Agavoideae	NC_032710
6.	<i>rps19</i> - Both	<i>Hosta ventricosa</i>	Agavoideae	NC_032706
7.	<i>rps19</i> - Both	<i>Hesperocallis undulata</i>	Agavoideae	NC_032704
8.	<i>rps19</i> - Both	<i>Hesperoyucca whipplei</i>	Agavoideae	NC_032705
9.	<i>rps19</i> - Both	<i>Chlorogalum pomeridianum</i>	Agavoideae	NC_032701
10.	<i>rps19</i> - Both	<i>Camassia scilloides</i>	Agavoideae	NC_032700
11.	<i>rps19</i> - Both	<i>Anthericum ramosum</i>	Agavoideae	NC_035972
12.	<i>rps16</i>	<i>Eustrephus latifolius</i>	Lomandroideae	NC_025305
13.	<i>rps16</i>	<i>Cordyline indivisa</i>	Lomandroideae	NC_035998
14.	<i>rps16</i>	<i>Allium sativum</i>	Allioideae	NC_031829
15.	<i>rps16</i>	<i>Hosta ventricosa</i>	Agavoideae	NC_032706
16.	<i>rps16</i>	<i>Beschorneria septentrionalis</i>	Agavoideae	NC_032699
17.	<i>rps2</i>	<i>Allium sativum</i>	Allioideae	NC_031829
18.	<i>rbcL</i> , <i>atpB</i> , <i>atpE</i> inversion	<i>Cordyline indivisa</i>	Lomandroideae	NC_035998
19.	<i>accD</i>	<i>Eustrephus latifolius</i>	Lomandroideae	NC_025305
20.	<i>clpP</i>	<i>Eustrephus latifolius</i>	Lomandroideae	NC_025305
21.	<i>clpP</i>	<i>Cordyline indivisa</i>	Lomandroideae	NC_035998
22.	<i>infA</i>	<i>Eustrephus latifolius</i>	Lomandroideae	NC_025305
23.	<i>infA</i>	<i>Allium sativum</i>	Allioideae	NC_031829
24.	<i>infA</i>	<i>Anthericum ramosum</i>	Agavoideae	NC_035972
25.	<i>infA</i>	<i>Aphyllanthes monspeliensis</i>	Aphyllanthoideae	NC_035968
26.	<i>infA</i>	<i>Aloe vera</i>	Asphodeloideae	NC_035506
27.	<i>infA</i>	<i>Milla biflora</i>	Brodiaeoideae	NC_036000
28.	<i>infA</i>	<i>Camassia scilloides</i>	Agavoideae	NC_032700
29.	<i>infA</i>	<i>Yucca brevifolia</i>	Agavoideae	NC_032711
30.	<i>infA</i>	<i>Hesperocallis undulata</i>	Agavoideae	NC_032704
31.	<i>infA</i>	<i>Hosta ventricosa</i>	Agavoideae	NC_032706
32.	<i>infA</i>	<i>Nolina atopocarpa</i>	Nolinoideae	NC_032708
33.	<i>infA</i>	<i>Convallaria majalis</i>	Nolinoideae	In this study
34.	<i>infA</i>	<i>Liriope spatia</i>	Nolinoideae	In this study
35.	<i>ycf1</i>	<i>Polygonatum cyrtoneuma</i>	Nolinoideae	NC_028429

Supplementary Table S3. The list of pseudogenes, intron deletion gene, and gene inversion in the Asparagales order of plants.

Sub-Family Name	Species Name	<i>accD</i>	<i>clpP</i> [#]	<i>infA</i>	<i>rbcL-atpE</i> *	<i>rpl23</i>	<i>rpl32</i>	<i>rps2</i>	<i>rps16</i>	<i>rps19</i> [@]	<i>ycf1</i> ^{\$}
Agavoideae	<i>Anthericum ramosum</i>			■						■	
	<i>Beschorneria septentrionalis</i>								■		
	<i>Camassia scilloides</i>			■						■	
	<i>Chlorogalum pomeridianum</i>									■	
	<i>Hesperaloe parviflora</i>									■	
	<i>Hesperocallis undulata</i>			■						■	
	<i>Hesperoyucca whipplei</i>									■	
	<i>Hosta ventricosa</i>			■					■	■	
	<i>Schoenolirion croceum</i>									■	
	<i>Yucca brevifolia</i>			■							
Allioideae	<i>Allium sativum</i>			■				■	■		
Aphyllanthoideae	<i>Aphyllanthes monspeliensis</i>			■							
Asparagoideae	<i>Asparagus officinalis</i>										
Asphodeloideae	<i>Aloe vera</i>			■			■				
Brodiaeoideae	<i>Milla biflora</i>			■							
Lomandroideae	<i>Cordyline indivisa</i>		■		■				■		
	<i>Eustrephus latifolius</i>	■	■	■		■			■		
Nolinoideae	<i>Convallaria keiskei</i> ¹			■							
	<i>Liriope spatia</i> ²			■							
	<i>Mianthemum bicolor</i>										
	<i>Nolina atopocarpa</i>			■							
	<i>Polygonatum cyrtonema</i>			■						■	■

[#] Two introns of *clpP* gene were lost in the cp genome.

* The protein-coding genes *rbcL*, *atpB* and *atpE* inverted in the LSC region.

[@] *rps19* – Two copies of *rps19* genes were deleted in IR regions of above mentioned species except for *Polygonatum crytonema*.

^{\$} *ycf1* – Two copies of *ycf1* were pseudogenes in *Polygonatum crytonema*.

^{1, 2} These plants were used in the present study.

Supplementary Table S4. List of primers used in this study

Primer name	Primer sequence position (bp)	Primer location region	Primer sequence (5' - 3' direction)	Product size (bp)
1F	94,245 – 94,269	<i>ycf2</i>	GGAACCATGCAGTACCAGACACGAG	4,910
1R	99,131 – 99,154	<i>ndhB</i>	ACCAAGAAATAACCCCTCACGTGCG	
2F	87,259 – 87,283	<i>rpl2</i>	TGCCTTGCGGTAATGATTCCTCTGG	8,943
2R	96,178 – 96,201	Mitochondrial DNA region	AGAACCCCATTTTCGACCAAGTTCCG	
3F	96,238 – 96,262	Mitochondrial DNA region	GTTACTGGTCGAGCCACTGGAATGG	8,867
3R	105,080 – 105,104	<i>rrn16</i>	AAGGGCAGGTTCTTACGCGTACTC	
4F	87,259 – 87,283	<i>rpl2</i>	TGCCTTGCGGTAATGATTCCTCTGG	17,846
4R	105,080 – 105,104	<i>rrn16</i>	AAGGGCAGGTTCTTACGCGTACTC	