

1 Article

2 **Comparative genome analysis revealed gene**  
3 **inversions, boundary expansion and contraction,**  
4 **and gene loss in *Stemona sessilifolia* (Miq.) Miq.**  
5 **chloroplast genome**

6

7 Jingting Liu<sup>1</sup>, Mei Jiang<sup>1</sup>, Haimei Chen<sup>1</sup>, Yu Liu<sup>2</sup>, Chang Liu<sup>1\*</sup>, Wuwei Wu<sup>2\*</sup>

8

9 1 Key Laboratory of Bioactive Substances and Resource Utilization of Chinese Herbal Medicine from  
10 Ministry of Education, Engineering Research Center of Chinese Medicine Resources from Ministry of  
11 Education, Institute of Medicinal Plant Development, Chinese Academy of Medical Sciences, Peking  
12 Union Medical College, Beijing 100193, P. R. China.

13 2 Guangxi Botanical Garden of Medicinal Plants, Nanning 530023, P. R. China.

14 \*Corresponding Authors: Chang Liu; Tel: +86-10-57833111; Fax: +86-10-62899715; Email:  
15 [cliu6688@yahoo.com](mailto:cliu6688@yahoo.com); Wuwei Wu; Tel: +86-771-5602461; Email: [wuweiwu2013@163.com](mailto:wuweiwu2013@163.com).

16

17 Email Address:

18 JTL: [liujingtingy@163.com](mailto:liujingtingy@163.com)

19 MJ: [mjiang0502@163.com](mailto:mjiang0502@163.com)

20 HMC: [hmchen@implad.ac.cn](mailto:hmchen@implad.ac.cn)

21 YL: [52888147@qq.com](mailto:52888147@qq.com)

22 CL: [cliu6688@yahoo.com](mailto:cliu6688@yahoo.com)

23 WWW: [wuweiwu2013@163.com](mailto:wuweiwu2013@163.com)

24

25

26

27

28

29

30

31

32

33

34

## 35 Abstract

36 *Stemona sessilifolia* (Miq.) Miq., commonly known as Baibu, is one of the most popular herbal  
37 medicines in Asia. In Chinese Pharmacopoeia, Baibu has multiple authentic sources, and there are  
38 many homonym herbs sold as Baibu in the herbal medicine market. The existence of the counterfeits  
39 of Baibu brings challenges to its identification. To assist the accurate identification of Baibu, we  
40 sequenced and analyzed the complete chloroplast genome of *Stemona sessilifolia* using  
41 next-generation sequencing technology. The genome was 154,039 bp in length, possessing a typical  
42 quadripartite structure consisting of a pair of inverted repeats (IRs: 27,094 bp) separating by a large  
43 single copy (LSC: 81,950 bp) and a small single copy (SSC: 17,901 bp). A total of 112 unique genes  
44 were identified, including 80 protein-coding, 28 transfer RNA, and four ribosomal RNA genes.  
45 Besides, 45 tandem, 27 forward, 23 palindromic, and 72 simple sequence repeats were detected in  
46 the genome by repeat analysis. Compared with its counterfeits (*Asparagus officinalis* and *Carludovica*  
47 *palmata*), we found that IR expansion and SSC contraction events of *Stemona sessilifolia* resulted in  
48 two copies of the *rpl22* gene in the IR regions and partial duplication of the *ndhF* gene in the SSC  
49 region. Secondly, an approximately 3-kb-long inversion was identified in the LSC region, leading to  
50 the *petA* and *cemA* gene presented in the complementary strand of the chloroplast DNA molecule.  
51 Comparative analysis revealed some highly variable regions, including *trnF-GAA\_ndhJ*, *atpB\_rbcL*,  
52 *rps15\_ycf1*, *trnG-UCC\_trnR-UCU*, *ndhF\_rpl32*. Finally, gene loss events were investigated in the  
53 context of phylogenetic relationships. In summary, the complete plastome of *Stemona sessilifolia* will  
54 provide valuable information for the molecular identification of Baibu and assist in elucidating the  
55 evolution of *Stemona sessilifolia*.

## 56 Introduction

57 Radix Stemonae, also known as Baibu, is one of the most popular herbal medicines used in  
58 many Asian countries, including China, Korea, Japan, Thailand, and Vietnam. It has been used in  
59 treating various respiratory diseases such as bronchitis, pertussis, and tuberculosis [1, 2]. It was also  
60 well known for killing cattle parasites, agricultural pests, and domestic insects [3, 4]. Stenine B, one  
61 of the major chemical ingredients of Baibu, has been considered a potential drug candidate against  
62 Alzheimer's disease due to its significant acetylcholinesterase inhibitory activity [5]. Owing to the  
63 important medicinal values, extensive genetic, biochemical, and pharmacological studies on Baibu is  
64 needed.

65 According to Pharmacopoeia of the People's Republic of China (2015 edition), the root tubers of  
66 *Stemona tuberosa*, *Stemona japonica*, and *Stemona sessilifolia* were all considered as the authentic  
67 sources of Baibu. Although these three species were all employed as the raw materials of Baibu, we  
68 cannot ignore their inherent difference. For example, Stemona alkaloids are the major components  
69 responsible for Baibu's antitussive activities. However, their composition and contents vary among *S.*  
70 *tuberosa*, *S. japonica*, and *S. sessilifolia* [6, 7]. These three species differ in antitussive, anti-bacterial,  
71 and insecticidal activities [8]. Therefore, it is critical to determine the exact origin of plant materials  
72 used as Baibu.

73 On the other hand, multiple authentic sources and the homonym also increase the difficulty of  
74 identifying Baibu. In some area of China, another herbal medicine, *Aconitum kusnezoffii* Rchb., is  
75 also called Baibu. However, the therapeutic activity of *Aconitum kusnezoffii* is significantly different  
76 from the authentic sources of Baibu described in Chinese Pharmacopoeia. Researches even  
77 reported that it might result in toxicity when *Aconitum kusnezoffii* was taken in large quantities [9].  
78 Besides, counterfeits in the herbal market also brought challenges to the exact identification of Baibu.  
79 Due to their similar morphologic features to the authentic sources for Baibu, many counterfeits  
80 such as *Asparagus officinalis*, *Asparagus filicinus*, and *Asparagus acicularis* were sold as Baibu in  
81 the herbal market frequently [10]. Therefore, the exact identification of Baibu origin is critical for its  
82 usage as a medicinal herb.

83 DNA barcode was deemed a more efficient and effective method in identifying plant species  
84 compared to morphological characteristics. Typical barcodes such as *ITS*, *psbA-trnH*, *matK*, and  
85 *rbcL* have been used to distinguish different plant species [11-13]. However, these DNA barcodes  
86 were not always working effectively, especially when distinguishing closely related plant species.  
87 Such a phenomenon may attribute to single-locus DNA barcodes still lack adequate variations in  
88 closely related taxa. Compared with DNA barcodes, the chloroplast genome provides more  
89 abundant genetic information and higher resolution in identifying plant species. Some researchers  
90 have proposed using the chloroplast genome as a species-level DNA barcode [14, 15].

91 The chloroplast is an organelle presenting in almost all green plants. It is the center of  
92 photosynthesis and plays a vital role in sustaining life on earth by converting solar energy to  
93 carbohydrates. Besides photosynthesis, chloroplast also plays critical roles in other biological  
94 processes, including the synthesis of amino acids, nucleotides, fatty acids, and many secondary  
95 metabolites. Furthermore, metabolites synthesized in chloroplasts are often involved in plants'  
96 interactions with their environment, such as response to environmental stress and defense against  
97 invading pathogens [16-18]. Due to its essential roles in the cellular processes and relatively small  
98 genome size, the chloroplast genome is a good starting point for resolving phylogenetic ambiguity,  
99 discriminating closely related species, and revealing the plants' evolutionary process. To date, over  
100 5000 chloroplast genomes from a variety of land plants are available. Phylogenetic analyses have  
101 demonstrated chloroplast genomes' effectiveness in inferring phylogenetic and distinguishing closely  
102 related plant species [19, 20].

103 Unfortunately, the taxonomic coverage of the sequenced chloroplast genome is somewhat  
104 biased. For example, until now, the chloroplast genome of *Stemon sessilifolia* has not been  
105 reported. The lack of chloroplast genome information prohibited studies aiming to understand the  
106 evolutionary processes in the family Stemonaceae. Here, we reported the full plastid genome of  
107 *Stemon sessilifolia*. Based on the sequence data, we performed a multi-scale comparative genome  
108 analysis among *Stemon sessilifolia*, *Asparagus officinalis*, and *Carludovica palmata* (the major  
109 counterfeits of Baibu). We investigated the difference among these three species from three aspects,  
110 including general characteristics, repeat sequences, and sequence divergences. We also  
111 characterized the significant changes, including genome rearrangement, IR expansion, and SSC

112 contraction, in the plastid genome of *Stemona sessilifolia*, *Asparagus officinalis*, and *Carludovica*  
113 *palmata*.

114 Lastly, we investigated the gene loss events in Stemonaceae and its closely related families  
115 (Asparagoideae, Velloziaceae, Cyclanthaceae, Pandanaceae). The results obtained in this work will  
116 provide valuable information for species identification of herb materials that are used as Baibu.  
117 Furthermore, it lays the foundation for elucidating the evolutionary history of plant species in the  
118 family Stemonaceae.

## 119 Materials and Methods

### 120 Plant Material and DNA Extraction

121 We collected fresh young leaves of *Stemona sessilifolia* from multiple individuals in the Institute  
122 of Medicinal Plant Development (IMPLAD), Beijing, China, and stored them at -80°C for chloroplast  
123 DNA extraction. All samples were identified by Professor Zhao Zhang, from the Institute of Medicinal  
124 Plant Development, Chinese Academy of Medical Sciences & Peking Union Medical College. The  
125 voucher specimens were deposited in the herbarium of IMPLAD. *Stemona sessilifolia* is not an  
126 endangered or protected species. Therefore, specific permissions for the collection of *Stemona*  
127 *sessilifolia* were not required. Total DNA was acquired from 100mg fresh young leaves using a plant  
128 genomic DNA kit (Tiangen Biotech, Beijing, Co., Ltd.). Finally, 1.0% agarose gel and Nanodrop  
129 spectrophotometer 2000 (Thermo Fisher Scientific, United States) was used to evaluate the purity  
130 and concentration, respectively.

### 131 Genome Sequencing, Assembly, and Annotation

132 According to the standard protocol, the DNA of *Stemona sessilifolia* was sequenced using the  
133 Illumina Hiseq2000 platform, with insert sizes of 500 bases for the library. A total of 5,660,432  
134 paired-end reads (2 × 250bp) were obtained, and low-quality reads were trimmed with  
135 Trimmomatic software [21].

136 To extract reads belonging to the chloroplast genome, we downloaded 1,688 chloroplast  
137 genome sequences from GenBank and constructed a Basic Local Alignment Search Tool (BLASTn)  
138 database. All trimmed reads were mapped to this database using the BLASTN program [22], and  
139 reads with E-value > 1E-5 were extracted. The reads were assembled first using the SPAdes  
140 software with default parameters [23]. The contigs were then subjected to gap closure using the  
141 Seqman module of DNASTAR (V11.0) [24]. Finally, we evaluated the assembled genome's *quality*  
142 by mapping the reads to the genome using Bowtie2 (v2.0.1) with default settings [25]. For further  
143 evaluation, all the barcode sequences of *Stemona sessilifolia* available in GeneBank were download  
144 (S1 file), including *matK* (1), *petD*(1), *rbcL* (1), *rpoC1* (1), *rps16* (1), *rps19-rpl22-psbA* (1), *trnL* (3),  
145 and *trnL-trnF* (2), the number enclosed in parentheses represented the number of barcode  
146 sequence. The BLAST program was used to calculate the identity between the chloroplast genome  
147 sequence of *Stemona sessilifolia* and each barcode sequence. As a result, the barcode of  
148 *rps19-rpl22-psbA* is located at the boundary of LSC/IRb, with an identity value of 100%. All the other  
149 barcode sequences also gave identity values of 100%, indicating the high reliability of the chloroplast  
150 genome sequence.

151 Gene annotation of *Stemona sessilifolia* chloroplast genome was conducted using the  
152 CpGAVAS web service with the default parameters [26]. The tRNA genes were confirmed with  
153 tRNAscan-SE [27] and ARAGORN [28]. Then the gene/intron boundaries were inspected and  
154 corrected using the Apollo program [29]. The Cusp and Compseq programs from EMBOSS were  
155 used to calculate the codon usage and GC content [30]. Finally, OrganellarGenomeDRAW [31] was  
156 used to generate the circular chloroplast genome map of *Stemona sessilifolia*.

## 157 Repeat Sequence Analysis

158 Perl script MISA(<http://pgrc.ipk-gatersleben.de/misa/>) was used to identify simple sequence  
159 repeats (SSRs) with the following parameters: 8 repeat units for mononucleotide SSRs, 4 repeat  
160 units for di- and tri-nucleotide repeat SSRs, and 3 repeat units for tetra-, Penta-, and hexanucleotide  
161 repeat SSRs. Tandem Repeats Finder was used with parameters of 2 for matches and 7 for  
162 mismatches and indels [32]. For the minimum alignment score and the maximum period, the size  
163 was set to 50 and 500. Palindrome and forward repeats were identified by the REPuter web service  
164 [33]. The minimum repeat size and the similarity cutoff were set to 30 bp and 90%, respectively.

## 165 Comparative Genomic Analysis

166 A total of four species, including *Stemona sessilifolia*, *Asparagus officinalis* (NC\_034777),  
167 *Carludovica palmata* (NC\_026786), and *Sciaphila densiflora* (NC\_027659), were subjected to  
168 multiple sequence alignment using mVISTA with default parameters [34]. Subsequently, 20 introns  
169 and 108 intergenic regions shared by *Stemona sessilifolia*, *Asparagus officinalis*, and *Carludovica*  
170 *palmata* were extracted using custom MatLab scripts to perform sequence divergence analysis.  
171 Firstly, sequences of each intergenic-region/intron were aligned individually using the CLUSTALW2  
172 (v2.0.12) [35] program with options "-type = DNA -gapopen = 10 -gapext = 2". Secondly, Pairwise  
173 distances were calculated with the Distmat program in EMBOSS (v6.3.1) using the Kimura  
174 2-parameters (K2p) evolution model [36]. We attempted to discover highly divergent regions for the  
175 development of novel molecular markers. To identify the occurrence of genome rearrangement  
176 events in the chloroplast genome of *Stemona sessilifolia*, synteny analysis among the three species  
177 mentioned above were performed using Mauve Alignment [37].

## 178 Phylogenetic Analysis

179 A total of 11 chloroplast genomes were distributed into Stemonaceae (3), Cyclanthaceae (1),  
180 Pandanaceae (1), Velloziaceae (1), and Asparagoideae (5) were retrieved from the RefSeq  
181 database. The protein sequences shared by these chloroplast genomes were used to construct a  
182 phylogenetic tree with *Veratrum patulum* and *Paris dunniana* as outgroup taxa (S1 Table). Fifty-eight  
183 proteins were involved, including ACCD, ATPA, ATPB, ATPE, ATPF, ATPH, ATPI, CLPP, MATK,  
184 NDHB, NDHC, NDHJ, NDHK, PETA, PETB, PETD, PETG, PETL, PETN, PSAA, PSAB, PSAJ,  
185 PSBA, PSBB, PSBC, PSBD, PSBE, PSBF, PSBH, PSBI, PSBJ, PSBK, PSBL, PSBM, PSBN, PSBT,  
186 RBCL, RPL2, RPL14, RPL16, RPL22, RPL23, RPL33, RPL36, RPOA, RPOB, RPOC1, RPS2, RPS3,  
187 RPS4, RPS7, RPS8, RPS11, RPS14, RPS18, RPS19, YCF3, AND YCF4. All these protein  
188 sequences were aligned using the CLUSTALW2 (v2.0.12) program with options "-gap open = 10  
189 -gapext = 2 -output = phylip". We used Maximum Likelihood (ML) method to infer the evolutionary

190 history of *Stemona sessilifolia* and species closely related to it. The detailed parameters were  
191 "raxmlHPC-PTHREADS-SSE3 -f a -N 1000 -m PROTGAMMACPREV -x 551314260 -p 551314260  
192 -o Nicotiana\_tabacum, Solanum\_lycopersicum -T 20".

## 193 Results and discussion

### 194 General characteristics of chloroplast genomes

195 The gene map of *Stemona sessilifolia* is shown in Fig 1. The sequence is provided in S2 File  
196 along with those of the major counterfeit of Baibu, *Asparagus officinalis* (NC\_034777), and  
197 *Carludovica palmate* (NC\_026786). The chloroplast genomes of *Stemona sessilifolia* and two other  
198 species share the standard features of possessing a typical quadripartite structure consisting of a  
199 pair of inverted repeats (IRs) separating a large single copy (LSC) and a small single copy (SSC),  
200 similar to other angiosperm chloroplast genomes [38].

201 We then carried out a multi-scale comparative genome analysis of these three chloroplast  
202 genomes from four aspects, including the size, the guanine-cytosine (GC) content, the count of  
203 genes, and the gene organization (Table 1). The complete circle chloroplast genomes of *S.*  
204 *sessilifolia*, *A. officinalis*, and *C. palmate* were 154,039 bp, 156,699 bp, and 158545 bp, respectively.  
205 Compared to *A. officinalis* and *C. palmate*, *S. sessilifolia* showed a relatively short SSC region and a  
206 relatively long IR region. We speculated that the chloroplast genome of *S. sessilifolia* might  
207 undertake IR expansion and SSC contraction simultaneously. There has no significant difference  
208 among *S. sessilifolia*, *A. officinalis*, and *C. palmate*. Such a result may attribute to the high  
209 conservation of tRNAs and rRNAs. The length of CDS regions of *A. officinalis* and *C. palmate* is  
210 shorter than *S. sessilifolia*, indicating gene loss events may occur in the chloroplast genome of *A.*  
211 *officinalis* and *C. palmate*.

212 **Table 1.** Chloroplast genome characteristics of *Stemona sessilifolia* *Asparagus officinalis*  
213 and *Carludovica palmate*.

Plastome	Characteristics	Species			State
		<i>Stemona sessilifolia</i>	<i>Asparagus officinalis</i>	<i>Carludovica palmate</i>	
<b>Size (bp)</b>	Genome	154039	156699	158545	>
	LSC	81950	84999	71426	>
	IR	27094	26531	26529	<
	SSC	17901	18638	18364	>
	tRNA genes	2874	2863	2816	<
	rRNA genes	9056	9052	8866	<
<b>GC content</b>	CDS	79641	77436	77802	<
	Overall	38.00	37.59	37.74	<
	LSC	36.18	35.60	35.79	<

(%)	IR	42.70	42.92	42.81	>
	SSC	32.13	31.50	31.51	<
	tRNA genes	53.42	53.57	53.40	>
	rRNA genes	55.22	55.38	55.38	-
	CDS	38.31	38.1	38.41	>
	1st position	45.7	45.64	45.93	>
	2nd position	38.46	38.56	38.39	>
	3rd position	30.78	30.09	30.91	<
<b>NO. of genes</b>	Total	112	110	112	<
	protein-coding genes	80	78	80	<
	tRNAs	28	28	28	-
	rRNAs	4	4	4	-
	Genes with introns	18	18	18	-
	Genes in IR	21	22	18	>

214 LSC: Large single-copy, IR: Inverted repeat, SSC: Small single-copy, CDS: Coding sequence. ">" and "<"  
215 indicated the characteristic parameters of *Asparagus officinalis* greater than and less than *Stemona*  
216 *sessilifolia*, respectively. "-" represented the characteristic parameters of *Asparagus officinalis* and  
217 *Stemona sessilifolia* were equal to each other.

218 **Figure 1.** Gene maps of chloroplast genomes of *Stemona sessilifolia*, *Asparagus officinalis*, and  
219 *Carludovica palmate*. Genes inside and outside the circle were transcribed clockwise and  
220 counterclockwise, respectively. The darker gray in the inner circle indicated GC content. Genes  
221 with different functions were characterized with varying bars of color

222 For GC content, *S. sessilifolia* showed a higher value in LSC, SSC, and CDS regions than *A. officinalis* and *C. palmate*, even in the complete chloroplast genome. However, in the IR regions, *A. officinalis* and *C. palmate* showed a GC content value larger than *S. sessilifolia*. The GC content  
223 decreased remarkably from the first position to the third position in the codon position scale. Such a  
224 result was in line with the phenomenon observed in most land plant plastomes.

227 We identified 112, 110, and 112 genes in the chloroplast genomes of *S. sessilifolia*, *A. officinalis*,  
228 and *C. palmate*, respectively. All of these three chloroplast genomes have 28 tRNAs and four rRNAs.  
229 The number of genes with introns in each species is 18, similar to reports in prior works [39].  
230 Therefore, we may conclude that there have no intron loss events occurred in the chloroplast  
231 genomes of these three species. All the genes with introns were described in Table S2. Besides, 21,  
232 22, and 18 genes were predicted for *S. sessilifolia*, *A. officinalis*, and *C. palmate* in IR regions.

233 The gene organizations were compared in Table 2. In the upstream region and the downstream  
234 region of the *C. palmate* chloroplast genome, premature stop codons were discovered in the *ycf1*  
235 gene, resulting in the loss of this gene. Compared to *S. sessilifolia*, we found the shorter CDS  
236 regions of *C. palmate* is directly related to the loss of this gene. We also found a full-length and a  
237 pseudogene of *ndhF* gene coexist in the chloroplast genome of *S. sessilifolia*, which further indicated  
238 SSC contraction events.

239

240      **Table 2.** Genes presented in chloroplast genomes of *Stemona sessilifolia*, *Asparagus*  
241      *officinalis* and *Carludovica palmate*.

Category for genes	Group of genes	Name of genes
Ribosome RNA genes	rRNA genes	<i>rrn16S</i> <sup>a</sup> , <i>rrn23S</i> <sup>a</sup> , <i>rrn5S</i> <sup>a</sup> , <i>rrn4.5S</i> <sup>a</sup>
Transfer RNA genes	tRNA genes	<i>trnT-UGU</i> , <i>trnR-ACG</i> <sup>a</sup> , <i>trnT-GGU</i> , <i>trnS-UGA</i> , <i>trnfM-CAU</i> , <i>trnF-GAA</i> , <i>trnL-UAG</i> , <i>trnV-UAC</i> <sup>*</sup> , <i>trnL-CAA</i> <sup>a</sup> , <i>trnM-CAU</i> <sup>a</sup> , <i>trnG-GCC</i> , <i>trnQ-UUG</i> , <i>trnA-UGC</i> <sup>a</sup> , <sup>**</sup> , <i>trnD-GUC</i> , <i>trnP-UGG</i> , <i>trnI-CAU</i> <sup>a</sup> , <i>trnE-UUC</i> <sup>**</sup> , <i>trnL-UAA</i> <sup>**</sup> , <i>trnK-UUU</i> <sup>**</sup> , <i>trnW-CCA</i> , <i>trnY-GUA</i> , <i>trnI-GAU</i> <sup>a</sup> , <sup>*</sup> , <i>trnG-UCC</i> <sup>*</sup> , <i>trnS-GGA</i> , <i>trnR-UCU</i> , <i>trnH-GUG</i> <sup>a</sup> , <i>trnS-GCU</i> , <i>trnN-GUU</i> <sup>a</sup> , <i>trnV-GAC</i> <sup>a</sup> , <i>trnC-GCA</i>
Others	Large subunit of ribosome	<i>rpl14</i> , <i>rpl16</i> <sup>*</sup> , <i>rpl2</i> <sup>a,*</sup> , <i>rpl20</i> , <i>rpl22</i> <sup>a</sup> , <i>rpl23</i> <sup>a</sup> , <i>rpl32</i> , <i>rpl33</i> , <i>rpl36</i>
	Small subunit of ribosome	<i>rps11</i> , <i>rps12</i> <sup>a,b,*</sup> , <i>rps14</i> , <i>rps15</i> , <i>rps16</i> <sup>*</sup> , <i>rps18</i> , <i>rps19</i> <sup>a</sup> , <i>rps2</i> , <i>rps3</i> , <i>rps4</i> , <i>rps7</i> <sup>a</sup> , <i>rps8</i>
	DNA dependent RNA polymerase	<i>rpoA</i> , <i>rpoB</i> , <i>rpoC1</i> <sup>*</sup> , <i>rpoC2</i>
	Subunits of NADH dehydrogenase	<i>ndhA</i> <sup>*</sup> , <i>ndhB</i> <sup>a,*</sup> , <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>
	Subunits of cytochrome b/f complex	<i>petA</i> , <i>petB</i> <sup>*</sup> , <i>petD</i> <sup>*</sup> , <i>petG</i> , <i>petL</i> , <i>petN</i>
	Subunits of photosystem I	<i>psaA</i> , <i>psaB</i> , <i>psaC</i> , <i>psaI</i> , <i>psaJ</i>
	Subunits of photosystem II	<i>psbA</i> , <i>psbB</i> , <i>psbC</i> , <i>psbD</i> , <i>psbE</i> , <i>psbF</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> , <i>ycf3</i>
	Large subunit of rubisco	<i>rbcL</i>
	Subunits of ATP synthase	<i>atpA</i> , <i>atpB</i> , <i>atpE</i> , <i>atpF</i> <sup>*</sup> , <i>atpH</i> , <i>atpI</i>
	Subunit of Acetyl-CoA-carboxylase	<i>accD</i>
	C-type cytochrome synthesis gene	<i>ccsA</i>
	Envelope membrane protein	<i>cemA</i>
	Protease	<i>clpP</i> <sup>**</sup>
	Translational initiation factor	<i>infA</i>
	Maturase	<i>matK</i>

---

Conserved open reading frames	<i>ycf1</i> , <i>ycf2<sup>a</sup></i> , <i>ycf15<sup>**</sup></i> , <i>ycf4</i>
Pseudo genes	<i>ycf1<sup>ψ</sup></i> , <i>ndhF<sup>ψ</sup></i> , <i>infA<sup>ψ</sup></i> , <i>ycf15<sup>a,ψ</sup></i> , <i>ycf68<sup>a,ψ</sup></i>

---

242 \* Gene with one intron, \*\* Gene with two introns, a Gene with two copies, b Trans-splicing gene,  
243 ψ Pseudogene. Black, green, red, and blue indicated genes identified in all species, both in  
244 *Stemona sessilifolia* and *Asparagus officinalis*, only in *Stemona sessilifolia*, and only in  
245 *Asparagus officinalis*, respectively.

## 246 Repeat Sequence Analysis

247 Simple sequence repeats (SSRs), the tandem repeat sequences consisting of 1-6 repeat  
248 nucleotide units, are widely distributed in prokaryotic and eukaryotic genomes. High polymorphism  
249 makes the SSRs effective molecular markers in species identification, population genetics, and  
250 phylogenetic research [40, 41]. In the current study, we investigated the distribution of SSRs in the  
251 genomes and their count and type (Fig 2). As a result, a total of 81, 59, and 72 SSRs were detected  
252 in *S. sessilifolia*, *A. officinalis*, and *C. palmata*, respectively. Mononucleotide motifs showed the  
253 highest frequency of SSRs in these species, followed by di-nucleotides and tri-nucleotides.  
254 Compared to *A. officinalis* and *C. palmata*, *S. sessilifolia* contained more SSRs. However, three  
255 tri-nucleotide repeats were detected in *C. palmata* and one in *A. officinalis*, but none were identified  
256 in *S. sessilifolia*. As expected, most mono-nucleotide and di-nucleotide repeats consisted of A/T and  
257 AT/AT repeats, respectively. The results suggest that these chloroplast genomes are rich in short  
258 poly-A and poly-T motifs, while poly-C and poly-G are relatively rare. We then use Tandem Repeats  
259 Finder [32] and REPuter [33] to detect long repeats and found 95, 70, and 95 long repeat sequences  
260 in *S. sessilifolia*, *A. officinalis*, and *C. palmata*, respectively. For *S. sessilifolia*, the number of  
261 Tandem repeats, Forward repeats, and Palindromic repeats was 45, 27, and 23, respectively. The  
262 number of the corresponding repeat sequences for *A. officinalis* was 45, 11, and 14, respectively.  
263 The number of the repeat sequences for *C. palmata* was 45, 33, and 17, respectively.

264 There have significant differences in the types of repeat sequence among *S. sessilifolia*, *A.*  
265 *officinalis*, and *C. palmata*. The repeat occurrence in *S. sessilifolia* was similar to that of *A. officinalis*  
266 but significantly higher than that of *C. palmata*. It should be noted that the size of the *A. officinalis*  
267 and *C. palmata* chloroplast genome is larger than the chloroplast genomes of *S. sessilifolia*.  
268 Therefore, the relatively larger size of the chloroplast genome of *A. officinalis* and *C. palmata* does  
269 not result from the repeat sequence.

270 **Figure 2.** Simple sequence repeats (SSRs) and long repeat sequences are identified in the  
271 chloroplast genomes. (A) Distribution of different types of SSRs in the chloroplast genomes. (B)  
272 Distribution of long repeat sequences in the chloroplast genomes. (C) Frequency of SSR motifs  
273 in different repeat class types.

## 274 Sequence divergence analysis

275 To evaluate the genome sequence divergence, we aligned sequences from four species using  
276 mVISTA [34] (Fig 3). The chloroplast genome of *S. sessilifolia* was significantly different from *A.*  
277 *officinalis* and *C. palmata*. Severe gene loss events always lead to highly reduced plastomes [20, 42].  
278 As expected, the non-coding regions were more divergent than coding regions among these species.

279 The two most divergent regions were *ycf4-psbJ* regions (red square A) and *rpl22* coding regions (red  
280 square B). We suspected that such a phenomenon might result from gene loss events or genome  
281 rearrangement events, and the detailed reasons will be discussed later. *Ycf1* gene is also highly  
282 divergent, which may occur due to the occurrence of pseudogenization. In summary, the LSC region  
283 showed the highest divergence, followed by the SSC region, and the IR regions were less divergent  
284 than the LSC and SSC region. Compared to the coding areas, the intergenic spacers displayed  
285 higher divergence.

286 Highly divergent regions always assist in the development of molecular markers. Because  
287 non-coding regions are evolved more rapidly than coding regions, the intergenic regions and intron  
288 regions were always considered ideal candidate regions of molecular markers with high resolution.  
289 Therefore, we calculated the Kimura 2-parameter (K2p) distances for each set of the intergenic  
290 regions and intron regions. A relatively higher K2p value between any two species is necessary to  
291 distinguish each species from the other two species. Therefore, we calculated the minimal K2P  
292 (MK2P) value for each set of intergenic regions and intron regions. The non-coding regions with  
293 higher MK2P values are likely to be the candidate regions of high-resolution molecular markers.  
294 Consequently, for introns (S3 Table), the MK2p value ranges from 0.0055 to 0.1096. *ClpP\_intron2*  
295 with the highest MK2p value followed by *rpl16\_intron1*, the third, fourth, and fifth were *rps16\_intron1*,  
296 *ndhA\_intron1*, and *trnL-UAA\_intron1*, respectively. For intergenic spacers (S4 Table), five highly  
297 conserved intergenic spacers were observed, including *ndhA\_ndhH*, *psaB\_psaA*, *psbL\_psbF*,  
298 *rpl2\_rpl23*, and *trnL-GAU\_trnA-UGC*. The MK2p value of intergenic spacers ranges from 0 to 0.3301,  
299 and the top-10 intergenic spacers with higher MK2p values were listed as follows: *trnF-GAA\_ndhJ*,  
300 *atpB\_rbcL*, *rps15\_ycf1*, *trnG-UCC\_trnR-UCU*, *ndhF\_rpl32*, *accD\_psaI*, *rps2\_rpoC2*,  
301 *trnS-GCU\_trnG-UCC*, *trnT-UGU\_trnL-UAA*, and *rps16\_trnQ-UUG*. In conclusion, compared to  
302 introns, we observed higher divergence in intergenic spacers. The intergenic spacers with large K2p  
303 values represent good candidate molecular markers to distinguish these three species.

304 **Figure 3.** Comparison of four chloroplast genomes using mVISTA program. Gray arrows  
305 indicated the orientations and positions of genes. Untranslated regions, conserved non-coding  
306 regions, and coding regions were characterized by sky-blue block, red block, and blue block. We  
307 adopted a cutoff value of 70% in the process of alignment.

## 308 **Rearrangement of the chloroplast genome**

309 To investigate whether there are significant differences in *ycf3-psbJ* regions (red square A in Fig  
310 3) and *rpl22* coding regions (red square B in Fig 3) between *S. sessilifolia* and its closely related  
311 species, we conducted synteny analysis. As plotted in Fig 4, we detected a large inversion of 3 kb  
312 long in the LSC region. Interestingly, such an approximately 3-kb long inversion was confirmed  
313 located in *ycf3-psbJ* regions. Therefore, we can conclude that the occurrence of genome  
314 rearrangement events leads to a significant difference in the *ycf3-psbJ* areas between *S. sessilifolia*  
315 and the other two species. To investigate whether such an inversion that exists in *S. sessilifolia* is  
316 unique, we conducted synteny analysis between the chloroplast genome of *S. sessilifolia* and  
317 species in Lioscoreales and Liliales, which belong to the two closely related orders of Pandanales.  
318 Compared to any species in Dioscoreales and Liliales, inversion in *ycf3-psbJ* regions in *S. sessilifolia*

319 was always visible (data are not shown). Therefore, inversion in the *ycf3-psbj* areas may be unique  
320 to *S. sessilifolia*.

321 **Figure 4.** Comparison of three chloroplast genomes using MAUVE algorithm. Local collinear  
322 blocks were colored to indicate syntenic regions, and histograms within each block indicated the  
323 degree of sequence similarity.

## 324 **IR expansion and SSC contraction**

325 IR contraction and expansion are common evolutionary events contributing to chloroplast  
326 genomes size variation [43]. Here, boundary comparison analysis was performed by which we  
327 attempt to identify IR contraction and expansion events (Fig 5). Compared to *A. officinalis* and *C.*  
328 *palmate*, the relatively larger IR regions indicated IR expansion events in *S. sessilifolia*.  
329 Simultaneously, the SSC region was shorter than *A. officinalis* and *C. palmate* by 465-737bp,  
330 suggesting the occurrence of SSC contraction events in *S. sessilifolia*. For *A. officinalis* and *C.*  
331 *palmate*, the *rpl22* gene is located at the LSC region with one copy. However, the IR regions of *S.*  
332 *sessilifolia* spanned to the intergenic spacers between the *rpl22* gene and *rps3* gene, resulting in two  
333 copies of the *rpl22* gene. Therefore, we can claim that the significant difference in *rpl22* coding  
334 regions between *S. sessilifolia* and its closely related species was attributed to IR expansion events.  
335 The IRb/SSC boundary extended into the *ycf1* genes by 1146-1260bp, creating *ycf1* pseudogene in  
336 *S. sessilifolia* and *C. palmate*. Considering premature stop codons were discovered in the *ycf1* gene,  
337 only one *ycf1* pseudogene was annotated in the SSC region in *A. officinalis*. The *ndhF* gene located  
338 at SSC regions in *A. officinalis* and *C. palmate*, and it ranges from 10-40bp away from the SSC/IRa  
339 boundary. However, in *S. sessilifolia*, the SSC region's shortening leads to the *ndhF* gene extended  
340 into the IRa region by 186bp. The *ndhF* gene located at the SSC/IRa junction resulted in partial  
341 duplication of this gene at the corresponding region. An overlap of 186bp between the *ndhF* gene  
342 and *ycf1* pseudogene was also observed in *S. sessilifolia*. Summarily, compared to *A. officinalis* and  
343 *C. palmate*, significant boundary expansion and contraction events were observed in *S. sessilifolia*  
344 simultaneously.

345 **Figure 5.** Comparison of IR, LSC, and SSC regions among *Stemona sessilifolia*, *Carludovica*  
346 *palmata*, and *Asparagus officinalis*. Numbers around the genes represented the gene lengths  
347 and the distances between the gene ends and boundary sites. Please note that the figure  
348 features were not to scale.  $\psi$  indicates pseudogene.

## 349 **Phylogenetic Analysis**

350 The chloroplast genome has been successfully used to determine plant categories and reveal  
351 plant phylogenetic relationships [44, 45]. To determine the phylogenetic position of *S. sessilifolia*, we  
352 constructed a phylogenetic tree with species in Stemonaceae and its closely related families  
353 (Asparagoideae, Velloziaceae, Cyclanthaceae, Pandanaceae). A total of 13 chloroplast genomes  
354 were retrieved from the RefSeq database, and 58 protein sequences shared by these species were  
355 used to construct a phylogenetic tree with *Veratrum patulum*, and *Paris dunniana* served as an  
356 outgroup (Fig 6). As a result, species in Stemonaceae, Asparagoideae, and Velloziaceae formed a  
357 cluster, respectively. Besides, *S. sessilifolia* and *S. japonica* formed a cluster within Stemonaceae  
358 with a bootstrap value of 100%, indicating the sister relationship between these two species.

359 As showed in Fig 6, a series of gene loss events were observed throughout Stemonaceae and  
360 its closely related families (Asparagoideae, Velloziaceae, Cyclanthaceae, Pandanaceae). A total of  
361 21 genes are lost in these species, including *ycf68* (11), *lhbA* (9), *infA* (4), *psbZ* (4), *ycf1* (3), *ccsA* (1),  
362 *ndhA* (1), *ndhD* (1), *ndhE* (1), *ndhF* (1), *ndhG* (1), *ndhH* (1), *ndhI* (1), *psaC* (1), *psaI* (1), *ycf2* (1),  
363 *rps16* (1), *rpl20* (1), *rpoC2* (1), *rps12* (1), and *rps15* (1), the number enclosed in parentheses  
364 represented the frequency of gene loss events. As expected, closely related species always tend to  
365 undertake the same gene loss events. A series of clusters formed by species undertaken the same  
366 gene deletion events further confirmed such a phenomenon. *C. palmata* and *P. tectorius* formed a  
367 cluster that lacked *psbZ* gene. The species from Pandanales (Stemnaceae, Cyclanthaceae,  
368 Pandanaceae, and Velloziaceae) formed a cluster without *ycf68* gene. The species from  
369 Asparagoideae formed a cluster without *lhbA* gene.

370 *Ycf68* gene has the highest frequency of gene deletion, and the second was *lhbA* gene. The  
371 following three were the *infA* gene, *psbZ*, and *ycf1* gene, respectively. The *ycf68* gene was only  
372 found in two species (*Asparagus racemosus*, *Asparagus setaceus*), and the *lhbA* gene was only  
373 found in four species (*C. palmata*, *C. heterosepala*, *P. tectorius* and *S. japonica*). The function of the  
374 *ycf68*, *lhbA*, and *ycf1* gene remained unknown. The occurrence of premature stop codons may  
375 account for these three genes' rare existence in chloroplast genomes [38, 46, 47]. As one of the  
376 most active genes in the chloroplast genome, the *infA* gene plays an essential role in protein  
377 synthesis. The frequent absence of the *infA* gene may contribute to the transfer of this gene between  
378 cytoplasm and nucleus [38, 48]. The lack of the subunits of the photosystem II gene *psbZ* was  
379 frequently observed in Pandanales (Stemnaceae, Cyclanthaceae, Pandanaceae). For each of the  
380 remaining 16 genes, only one gene loss event was observed, respectively. There was gene absence  
381 in each species' chloroplast genome, indicating the variation in chloroplast genomes' contents.  
382 However, for 16 out of 21 genes, the frequency of gene loss events was only one, suggesting the  
383 chloroplast genome is highly conserved on the scale of gene contents. Such a phenomenon is  
384 consistent with the highly conserved nature of the chloroplast genome and its feature of rich in  
385 variation.

386 **Figure 6.** Molecular phylogenetic Analyses of Pandanales and its closely related orders. We  
387 constructed the tree with the sequences of 58 proteins presented in 116 species using the  
388 Maximum Likelihood method implemented in RAxML with *Nicotiana tabacum* and *Solanum*  
389 *Lycopersicum* served as an outgroup. The numbers associated with the nodes indicate  
390 bootstrap values tested with 1000 replicates. We marked the orders and families for each  
391 species besides the branches and the occurrence of gene loss events.

## 392 Discussion

393 In this study, we sequenced and analyzed the chloroplast genome of *Stemona sessilifolia* and  
394 performed multi-scale comparative genomics of *Stemona sessilifolia*, *Asparagus officinalis*, and  
395 *Carludovica palmata* (the major counterfeit of Baibu). We also characterized the major changes in  
396 the chloroplast genome of *Stemona sessilifolia* compared with those of Dioscoreales, Liliales, and  
397 Pandanales, including genome rearrangement, IR expansion, and SSC contraction, and investigate  
398 the occurrence of gene loss events in Dioscoreales, Liliales, Pandanales, and Asparagaceae.

399 Our results show that the genome of *Stemona sessilifolia* is very similar to that of *Stemona*  
400 *japonica* previously reported. In both chloroplast genomes of *S. sessilifolia* and *S. japonica*, the rps12  
401 gene contained two introns. It is a trans-spliced gene with a 5' end exon located in the LSC region,  
402 and the 3' end exon and intron located in the IR regions [49]. Also, we detected a large inversion in  
403 both species. The SSC region was found to have a reverse orientation in *S. japonica*. The SSC  
404 region's reverse direction has been interpreted as a major inversion existing within the species  
405 [50-52].

406 Interestingly, a 3-kb long inversion was detected in the chloroplast genome of *S. sessilifolia*. It  
407 might result from a genome rearrangement event. This unique inversion phenomenon led to  
408 significant differences in the ycf3-psbJ region between *Stemona sessilifolia* and its related species,  
409 which can be used as a candidate region to identify *Stemona sessilifolia* from counterfeits.

410 SSRs have been widely used as molecular markers in the studies of species identification,  
411 population genetics, and phylogenetic investigations based on their high-degree variations [53]. The  
412 SSR consisting of A/T is the *most abundant type* in *S. sessilifolia* and *S. japonica*. These SSRs loci  
413 were mainly located in intergenic regions and would help develop new phylogenetic markers for  
414 species identification and discrimination [49]. Only forward and palindrome repeats were found in the  
415 *S. sessilifolia* cp genome regarding the long repeat sequences. The biological implication of these  
416 repeats remains to be elucidated.

417 Also, there were significant differences in IR contraction and expansion between *Stemona*  
418 *sessilifolia* and other species. At the IRa/LSC border, the spacer from rpl22 coding regions to the  
419 border is longer in *Stemona sessilifolia* (309 bp) than that of *Stemona japonica* (65 bp). The IRb/SSC  
420 boundary extended into the ycf1 genes by only 18bp and created a ycf1 pseudogene in *Stemona*  
421 *japonica* [49]. However, that region is 1146-1260bp long in *Stemona sessilifolia*. The function of ycf1  
422 genes is mostly unknown, but it evolves rapidly [54]. The larger contraction and expansion of the IR  
423 region in *Stemona sessilifolia* may lead to evolutionary differences between *Stemona sessilifolia*  
424 and its closely related species. This may need further verification.

425 *Stemona sessilifolia* and *Stemona japonica* are the authentic sources of Baibu, according to  
426 Pharmacopoeia of the People's Republic of China (2015 edition). Phylogenetic analyses showed  
427 that they were placed close to each other with a bootstrap value of 100%. *Asparagus officinalis* and  
428 *Carludovica palmata* (the major counterfeit of Baibu) were on the other branches. When we  
429 investigated the gene loss events in the phylogenetic relationship context, we also see the cp  
430 genomes of *Stemona sessilifolia* and *Stemona japonica* have similar gene loss patterns. These  
431 findings support the pharmaceutical use of *Stemona sessilifolia* and *Stemona japonica* as genuine  
432 Baibu. Also, they suggest the urgent need for new molecular markers for the identification of genuine  
433 Baibu. This study will be of value in determining genome evolution and understanding phylogenetic  
434 relationships within Pandanales and other species closed to Pandanales.

435 **Conclusions**

436 In summary, the complete plastome of *Stemona sessilifolia* (Miq.) Miq. was provided in the  
437 current study. We believe it will benefit as a reference for further complete chloroplast genome  
438 sequencing within the family. A multi-scale comparative genome analysis among *Stemona*  
439 *sessilifolia*, *Asparagus officinalis*, and *Carludovica palmate* (the major counterfeit of Baibu) was  
440 based on sequence data provided performed. Comparative Analysis of these three species revealed  
441 the existence of a unique inversion in the ycf3-psbJ regions. Interestingly, IR expansion and SSC  
442 contraction were observed simultaneously in *Stemona sessilifolia*, resulting in a rare boundary  
443 pattern. Some highly variable regions were screened as potential DNA barcodes for identification of  
444 these three species, including *trnF-GAA\_ndhJ*, *atpB\_rbcL*, *rps15\_ycf1*, *trnG-UCC\_trnR-UCU*,  
445 *ndhF\_rpl32*. Phylogenetic analyses showed that the two *Stemona* species were placed close to each  
446 other with a bootstrap value of 100%. Finally, we investigated the gene loss events in the context of  
447 the phylogenetic relationship. Closely related species always share similar gene loss patterns,  
448 consistent with those observed previously. This study will be of value in determining genome  
449 evolution and understanding phylogenetic relationships within Stemonaceae and families closed to  
450 Stemonaceae.

## 451 **Acknowledgments**

452 This work was supported by the CAMS Innovation Fund for Medical Sciences (CIFMS)  
453 (2016-I2M-3-016, 2017-I2M-1-013) from the Chinese Academy of Medical Science,  
454 National Natural Science Foundation of China (31460237),Guangxi science and technology  
455 program (AD17292003), Guangxi R&D project of medical and health technology (S201529). The  
456 funders were not involved in the study design, data collection, and Analysis, decision to publish, or  
457 manuscript preparation.

## 458 **Author Contributions**

459 CL and WW conceived the research; JTL and MJ carried out the bioinformatics studies and prepared  
460 the manuscript; HMC and YL collected samples of *Stemona sessilifolia*, extracted DNA for  
461 next-generation sequencing. All authors have read and approved the manuscript.

## 462 **Conflicts of Interest**

463 The authors declare no conflict of interest.

## 464 **References**

- 465 1. Pilli RA, Rosso GB, Oliveira MdCFd. The chemistry of *Stemona* alkaloids: An update. Natural  
466 Product Reports. 2010;27.
- 467 2. Wang Z, Yang W, Yang P, Gao B, Luo L. Effect of Radix Stemonae concentrated decoction on  
468 the lung tissue pathology and inflammatory mediators in COPD rats. BMC Complementary and  
469 Alternative Medicine. 2016;16(1):1-7.
- 470 3. Brem B, Seger C, Pacher T, Hofer O, Greger H. Feeding deterrence and contact toxicity of  
471 *Stemona* alkaloids-a source of potent natural insecticides. Journal of Agricultural & Food Chemistry.  
472 2002;50(22):6383-8.

473 4. Liu ZL, Goh SH, Ho SH. Screening of Chinese medicinal herbs for bioactivity against *Sitophilus*  
474 *zeamais* Motschulsky and *Tribolium castaneum* (Herbst). *Journal of Stored Products Research*.  
475 2007;43(3):290-6.

476 5. Lai DH, Yang ZD, Xue WW, Sheng J, Shi Y, Yao XJ. Isolation, characterization, and  
477 acetylcholinesterase inhibitory activity of alkaloids from roots of *Stemona sessilifolia*. *Fitoterapia*.  
478 2013;89(Complete):257-64.

479 6. Fan LL, Xu F, Hu JP, Yang DH, Chen HB, Komatsu K, et al. Binary chromatographic fingerprint  
480 analysis of *Stemonae Radix* from three *Stemona* plants and its applications. *Journal of Natural*  
481 *Medicines*. 2015.

482 7. Li SL, Jiang RW, Hon PM, Cheng L, Shaw PC. Quality evaluation of *Radix Stemonae* through  
483 simultaneous quantification of bioactive alkaloids by high-performance liquid chromatography  
484 coupled with diode array and evaporative light scattering detectors. *Biomedical Chromatography*  
485 *Bmc*. 2010;21(10):1088-94.

486 8. Xu YT, Hon PM, Jiang RW, Cheng L, Li SH, Chan YP, et al. Antitussive effects of *Stemona*  
487 *tuberosa* with different chemical profiles. *Journal of Ethnopharmacology*. 2006;108(1):46-53.

488 9. Yan Y, Zhang A, Dong H, Yan G, Wang X. Toxicity and Detoxification Effects of Herbal Caowu  
489 via Ultra Performance Liquid Chromatography/Mass Spectrometry Metabolomics Analyzed using  
490 Pattern Recognition Method. *Pharmacognosy Magazine*. 2017;13(52):683-92.

491 10. Fan LL, Zhu S, Chen HB, Yang DH, Cai SQ, Komatsu K. Molecular analysis of *Stemona* plants  
492 in China based on sequences of four chloroplast DNA regions. *Biological & Pharmaceutical Bulletin*.  
493 2009;32(8):1439.

494 11. Vere ND, Rich TCG, Trinder SA, Long C. DNA Barcoding for Plants. *Methods in Molecular*  
495 *Biology*. 2015;1245:101-18.

496 12. Penikar ZF, Buzan EV. 20 years since the introduction of DNA barcoding: From theory to  
497 application. *Journal of Applied Genetics*. 2013;55(1):43-52.

498 13. W., John, Kress, Carlos, García-Robledo, María, et al. DNA barcodes for ecology, evolution,  
499 and conservation. *Trends in Ecology & Evolution*. 2015.

500 14. Li X, Yang Y, Henry RJ, Rossetto M, Wang Y, Chen S. Plant DNA barcoding: from gene to  
501 genome. *Biological Reviews*. 2015.

502 15. Hollingsworth PM, Li DZ, Michelle VDB, Twyford AD. Telling plant species apart with DNA: from  
503 barcodes to genomes. *Philosophical Transactions of the Royal Society B Biological Sciences*.  
504 2016;371(1702):20150338.

505 16. H., E., Neuhaus, and, M., J., et al. NONPHOTOSYNTHETIC METABOLISM IN PLASTIDS.  
506 *Annual Review of Plant Physiology & Plant Molecular Biology*. 2000.

507 17. Krzysztof B, Burch-Smith TM. Chloroplast signaling within, between and beyond cells. *Frontiers*  
508 *in Plant Science*. 2015;6(781):781.

509 18. Bhattacharyya D, Chakraborty S. Chloroplast: the Trojan horse in plant–virus interaction.  
510 *Molecular Plant Pathology*. 2018.

511 19. Ubuka T, Tsutsui K. Comparative and Evolutionary Aspects of Gonadotropin-Inhibitory  
512 Hormone and FMRFamide-Like Peptide Systems. *Frontiers in Neuroscience*. 2018;12.

513 20. Lam VKY, Marybel SG, Graham SW. The Highly Reduced Plastome of Mycoheterotrophic  
514 *Sciaphila* (Triuridaceae) Is Colinear with Its Green Relatives and Is under Strong Purifying Selection.  
515 *Genome Biology and Evolution*. 2015;(8):8.

516 21. Bolger AM, Marc L, Bjoern U. Trimmomatic: a flexible trimmer for Illumina sequence data.  
517 *Bioinformatics*. 2014;(15):2114-20.

518 22. Camacho C, Coulouris G, Avagyan V, Ning M, Madden TL. BLAST+: architecture and  
519 applications. *BMC Bioinformatics* 10:421. *Bmc Bioinformatics*. 2009;10(1):421.

520 23. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new  
521 genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol*.  
522 2012;19(5):455-77. Epub 2012/04/18. DOI: 10.1089/cmb.2012.0021. PubMed PMID: 22506599;  
523 PubMed Central PMCID: PMCPMC3342519.

524 24. Burland TG. DNASTAR's Lasergene sequence analysis software. *Methods Mol Biol*.  
525 2000;132:71-91.

526 25. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short  
527 DNA sequences to the human genome. *Genome Biol*. 2009;10(3):R25. Epub 2009/03/06. DOI:  
528 10.1186/GB-2009-10-3-r25. PubMed PMID: 19261174; PubMed Central PMCID:  
529 PMCPMC2690996.

530 26. Liu C, Shi L, Zhu Y, Chen H, Zhang J, Lin X, et al. CpGAVAS, an integrated web server for the  
531 annotation, visualization, Analysis, and GenBank submission of completely sequenced chloroplast  
532 genome sequences. *BMC Genomics*. 2012;13:715. Epub 2012/12/22. DOI:  
533 10.1186/1471-2164-13-715. PubMed PMID: 23256920; PubMed Central PMCID:  
534 PMCPMC3543216.

535 27. Schattner P, Brooks AN, Lowe TM. The tRNAscan-SE, snoScan and snoGPS web servers for  
536 the detection of tRNAs and snoRNAs. *Nucleic Acids Res*. 2005;33(Web Server issue):W686-9. Epub  
537 2005/06/28. DOI: 10.1093/nar/gki366. PubMed PMID: 15980563; PubMed Central PMCID:  
538 PMCPMC1160127.

539 28. Laslett D, Canback B. ARAGORN, a program to detect tRNA genes and tmRNA genes in  
540 nucleotide sequences. *Nucleic Acids Res*. 2004;32(1):11-6. Epub 2004/01/06. DOI:  
541 10.1093/nar/gkh152. PubMed PMID: 14704338; PubMed Central PMCID: PMCPMC373265.

542 29. Misra S, Harris N. Using Apollo to browse and edit genome annotations. 2006.

543 30. Rice P, Longden I, Bleasby A. EMBOSS: the European Molecular Biology Open Software Suite.  
544 *Trends Genet*. 2000;16(6):276-7. Epub 2000/05/29. DOI: 10.1016/s0168-9525(00)02024-2. PubMed  
545 PMID: 10827456.

546 31. Lohse M, Drechsel O, Bock R. OrganellarGenomeDRAW (OGDRAW): a tool for the easy  
547 generation of high-quality custom graphical maps of plastid and mitochondrial genomes. *Current  
548 Genetics*. 2007;52(5-6):267-74.

549 32. Benson G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res*.  
550 1999;27(2):573-80. Epub 1998/12/24. DOI: 10.1093/nar/27.2.573.

551 33. Kurtz S, Choudhuri JV, Ohlebusch E, Schleiermacher C, Stoye J, Giegerich R. REPuter: the  
552 manifold applications of repeat analysis on a genomic scale. *Nucleic Acids Res*.  
553 2001;29(22):4633-42. Epub 2001/11/20. DOI: 10.1093/nar/29.22.4633.

554 34. Frazer KA, Pachter L, Poliakov A, Rubin EM, Dubchak I. VISTA: computational tools for  
555 comparative genomics. *Nucleic Acids Res*. 2004;32(Web Server issue):W273-9. Epub 2004/06/25.  
556 DOI: 10.1093/nar/gkh458.

557 35. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, et al. Clustal W  
558 and Clustal X version 2.0. *Bioinformatics*. 2007;23(21):2947-8. Epub 2007/09/12. DOI:  
559 10.1093/bioinformatics/btm404.

560 36. Kimura M. A simple method for estimating evolutionary rates of base substitutions through  
561 comparative studies of nucleotide sequences. *J Mol Evol.* 1980;16(2):111-20. Epub 1980/12/01. DOI:  
562 10.1007/bf01731581.

563 37. Darling AE, Mau B, Perna NT. progressiveMauve: multiple genome alignment with gene gain,  
564 loss and rearrangement. *PLoS One.* 2010;5(6):e11147. Epub 2010/07/02. DOI:  
565 10.1371/journal.pone.0011147.

566 38. Wicke S, Schneeweiss GM, dePamphilis CW, Müller KF, Quandt D. The evolution of the plastid  
567 chromosome in land plants: gene content, gene order, gene function. *Plant Mol Biol.*  
568 2011;76(3-5):273-97. Epub 2011/03/23. doi: 10.1007/s11103-011-9762-4.

569 39. Chen H, Shao J, Zhang H, Jiang M, Huang L, Zhang Z, et al. Sequencing and Analysis of  
570 *Strobilanthes cusia* (Nees) Kuntze Chloroplast Genome Revealed the Rare Simultaneous  
571 Contraction and Expansion of the Inverted Repeat Region in Angiosperm. *Front Plant Sci.*  
572 2018;9:324. Epub 2018/03/30. DOI: 10.3389/fpls.2018.00324.

573 40. Xue J, Wang S, Zhou SL. Polymorphic chloroplast microsatellite loci in *Nelumbo*  
574 (Nelumbonaceae). *Am J Bot.* 2012;99(6):e240-4. Epub 2012/05/23. DOI: 10.3732/ajb.1100547.

575 41. Powell W, Morgante M, McDevitt R, Vendramin GG, Rafalski JA. Polymorphic simple sequence  
576 repeat regions in chloroplast genomes: applications to the population genetics of pines. *Proc Natl  
577 Acad Sci U S A.* 1995;92(17):7759-63. Epub 1995/08/15. DOI: 10.1073/pnas.92.17.7759.

578 42. Wicke S, Müller KF, de Pamphilis CW, Quandt D, Wickett NJ, Zhang Y, et al. Mechanisms of  
579 functional and physical genome reduction in photosynthetic and nonphotosynthetic parasitic plants  
580 of the broomrape family. *Plant Cell.* 2013;25(10):3711-25. Epub 2013/10/22. DOI:  
581 10.1105/tpc.113.113373.

582 43. Wang RJ, Cheng CL, Chang CC, Wu CL, Su TM, Chaw SM. Dynamics and evolution of the  
583 inverted repeat-large single copy junctions in the chloroplast genomes of monocots. *BMC Evol Biol.*  
584 2008;8:36. Epub 2008/02/02. DOI: 10.1186/1471-2148-8-36.

585 44. Jansen RK, Cai Z, Raubeson LA, Daniell H, Depamphilis CW, Leebens-Mack J, et al. Analysis  
586 of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies  
587 genome-scale evolutionary patterns. *Proc Natl Acad Sci U S A.* 2007;104(49):19369-74. Epub  
588 2007/12/01. DOI: 10.1073/pnas.0709121104.

589 45. Moore MJ, Bell CD, Soltis PS, Soltis DE. Using plastid genome-scale data to resolve enigmatic  
590 relationships among basal angiosperms. *Proc Natl Acad Sci U S A.* 2007;104(49):19363-8. Epub  
591 2007/12/01. DOI: 10.1073/pnas.0708072104.

592 46. Jiang M, Chen H, He S, Wang L, Chen AJ, Liu C. Sequencing, Characterization, and  
593 Comparative Analyses of the Plastome of *Caragana rosea* var. *rosea*. *Int J Mol Sci.* 2018;19(5).  
594 Epub 2018/05/12. DOI: 10.3390/ijms19051419.

595 47. Nguyen PA, Kim JS, Kim JH. The complete chloroplast genome of colchicine plants (*Colchicum*  
596 *autumnale* L. and *Gloriosa superba* L.) and its application for identifying the genus. *Planta.*  
597 2015;242(1):223-37. Epub 2015/04/24. DOI: 10.1007/s00425-015-2303-7.

598 48. Millen RS, Olmstead RG, Adams KL, Palmer JD, Lao NT, Heggie L, et al. Many parallel losses  
599 of *infA* from chloroplast DNA during angiosperm evolution with multiple independent transfers to the  
600 nucleus. *Plant Cell.* 2001;13(3):645-58. Epub 2001/03/17. DOI: 10.1105/tpc.13.3.645.

601 49. Lu Q, Ye W, Lu R, Xu W, Qiu Y. Phylogenomic and Comparative Analyses of Complete  
602 Plastomes of *Croomia* and *Stemonaceae* (*Stemonaceae*). *Int J Mol Sci.* 2018;19(8). Epub 2018/08/15.  
603 DOI: 10.3390/ijms19082383.

604 50. Hansen DR, Dastidar SG, Cai Z, Penaflor C, Kuehl JV, Boore JL, et al. Phylogenetic and  
605 evolutionary implications of complete chloroplast genome sequences of four early-diverging  
606 angiosperms: *Buxus* (Buxaceae), *Chloranthus* (Chloranthaceae), *Dioscorea* (Dioscoreaceae), and  
607 *Illicium* (Schisandraceae). *Mol Phylogenet Evol.* 2007;45(2):547-63. Epub 2007/07/24. DOI:  
608 10.1016/j.ympev.2007.06.004.

609 51. Liu Y, Huo N, Dong L, Wang Y, Zhang S, Young HA, et al. Complete chloroplast genome  
610 sequences of Mongolia medicine *Artemisia frigida* and phylogenetic relationships with other plants.  
611 *PLoS One.* 2013;8(2):e57533. Epub 2013/03/06. DOI: 10.1371/journal.pone.0057533.

612 52. Walker JF, Zanis MJ, Emery NC. Comparative Analysis of complete chloroplast genome  
613 sequence and inversion variation in *Lasthenia burkei* (Madieae, Asteraceae). *Am J Bot.*  
614 2014;101(4):722-9. Epub 2014/04/05. DOI: 10.3732/ajb.1400049.

615 53. Provan J, Corbett G, McNicol JW, Powell W. Chloroplast DNA variability in wild and cultivated  
616 rice (*Oryza* spp.) revealed by polymorphic chloroplast simple sequence repeats. *Genome.*  
617 1997;40(1):104-10. Epub 1997/02/01. DOI: 10.1139/g97-014.

618 54. Wang L, Zhang H, Jiang M, Chen H, Huang L, Liu C. Complete plastome sequence of *Iodes*  
619 *cirrhosa* Turcz., the first in the Icacinaceae, comparative genomic analyses and possible split of  
620 *Idoes* species in response to climate changes. *PeerJ.* 2019;7:e6663. Epub 2019/04/12. DOI:  
621 10.7717/peerj.6663.

## 622 Supporting information

623 **S1 File. The barcode sequences of *Stemona sessilifolia* available in GeneBank.**

624 **S2 File. The sequence of *Stemona sessilifolia*, *Carludovica palmata*, and *Asparagus*  
625 *officinalis*.**

626 **S1 Table. List of chloroplast genomes used in this study.**

627 **S2 Table. The length of introns and exons for intron-containing genes.**

628 **S3 Table. K2p distances for intron regions among *Stemona sessilifolia*, *Carludovica*  
629 *palmata*, and *Asparagus officinalis*.**

630 **S4 Table. K2p distances for intergenic regions among *Stemona sessilifolia*, *Carludovica*  
631 *palmata*, and *Asparagus officinalis*.**

632

633

634

635

636

637

638

639

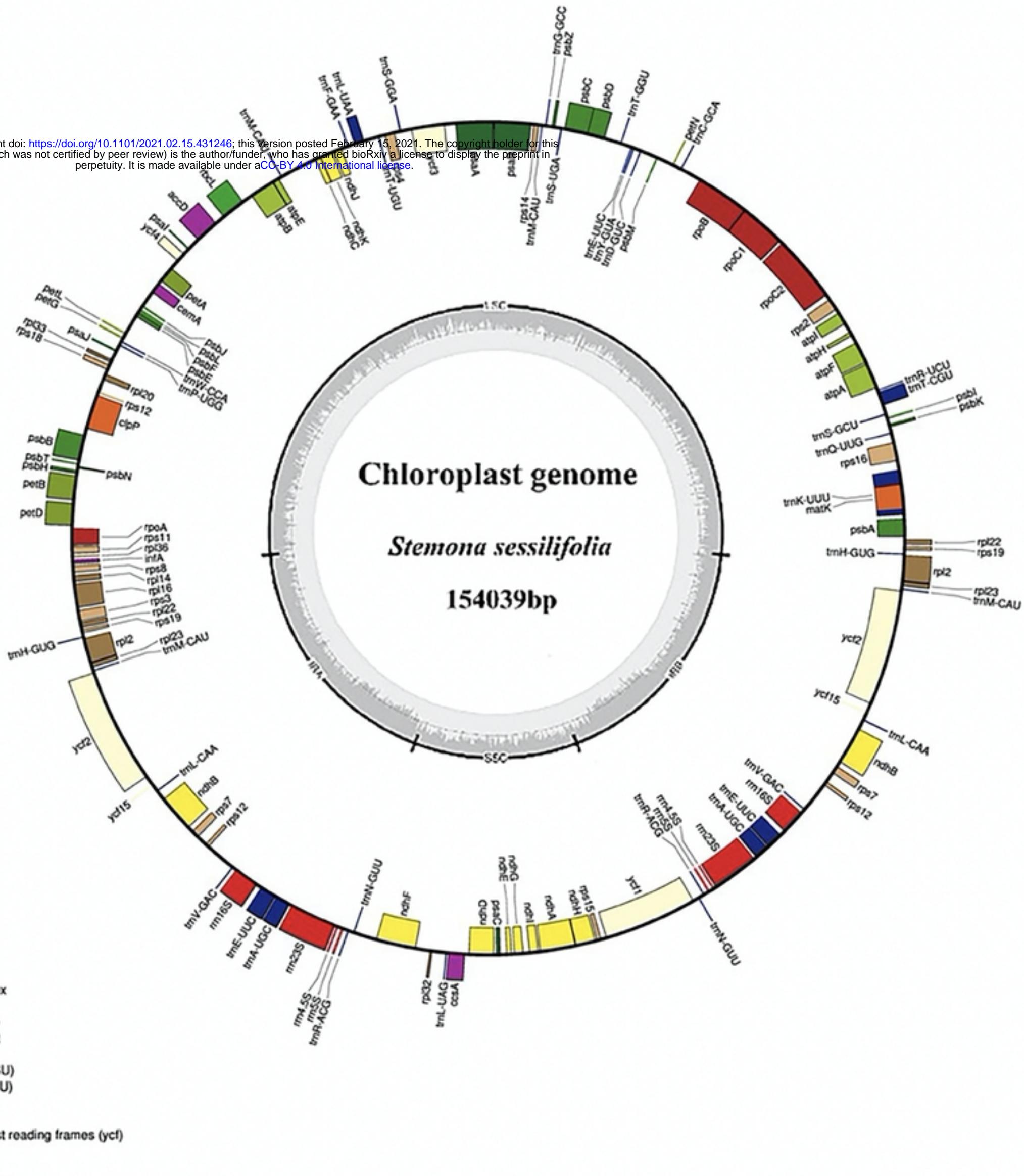
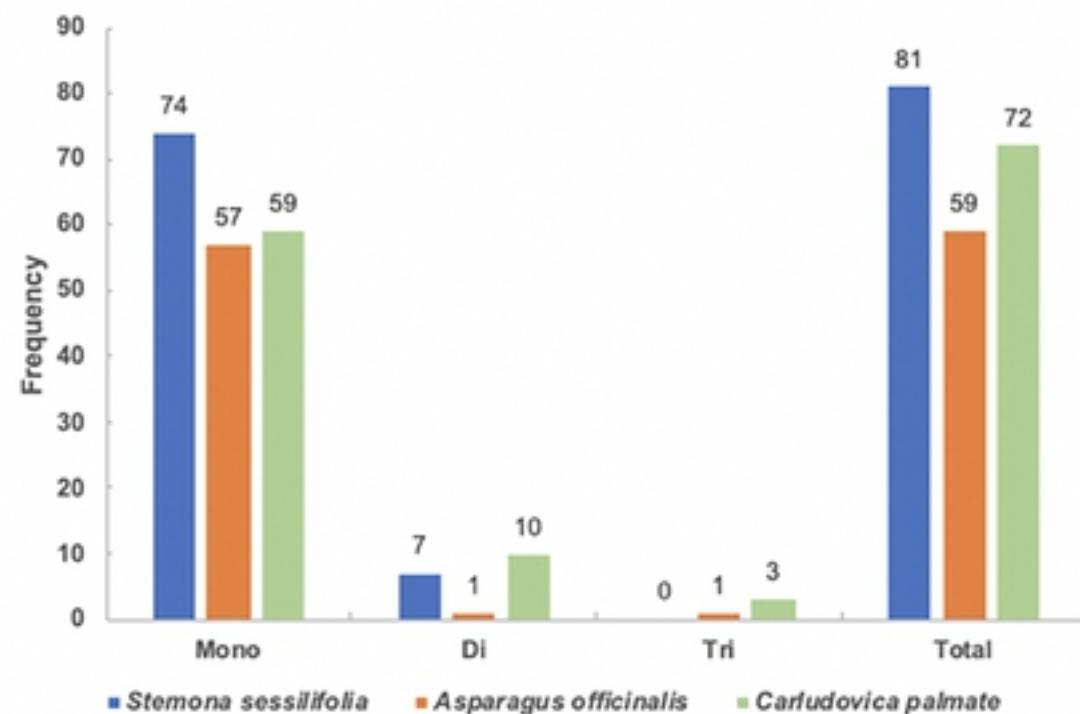
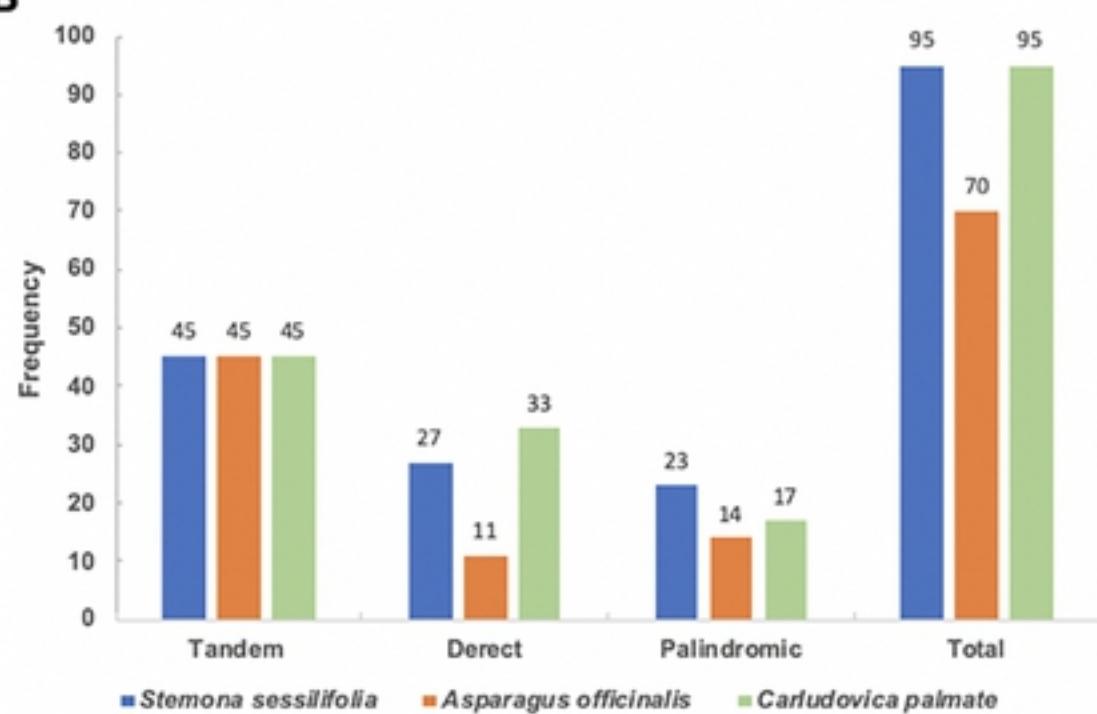
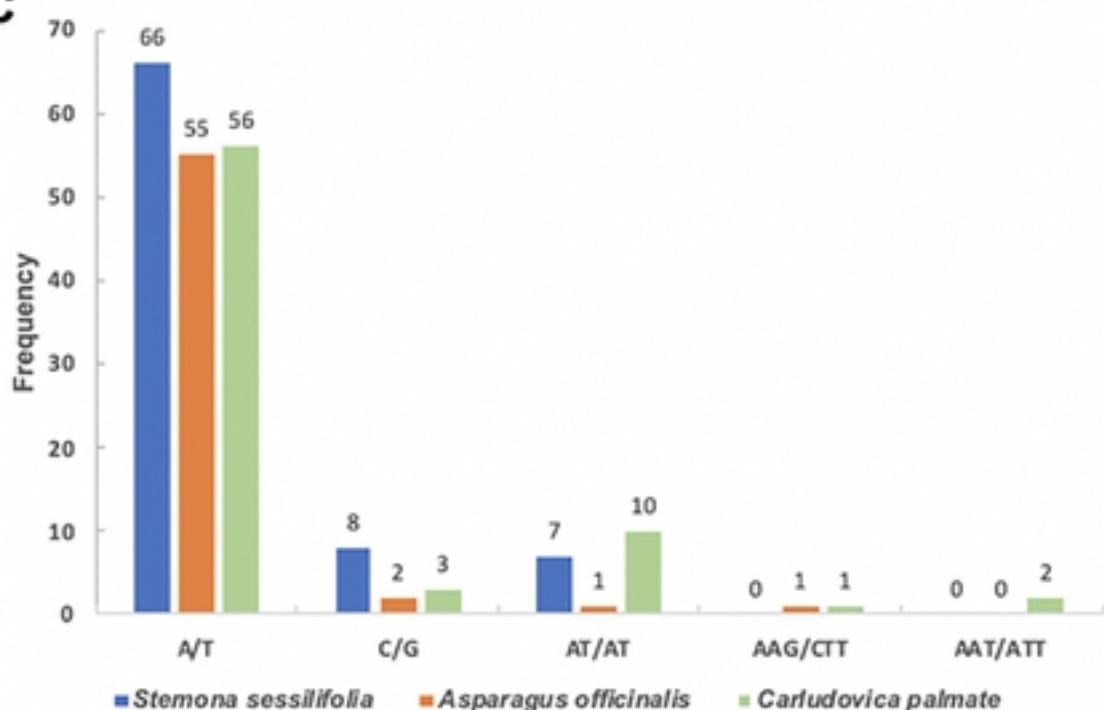


Figure 1

**A****B****C****Figure2**

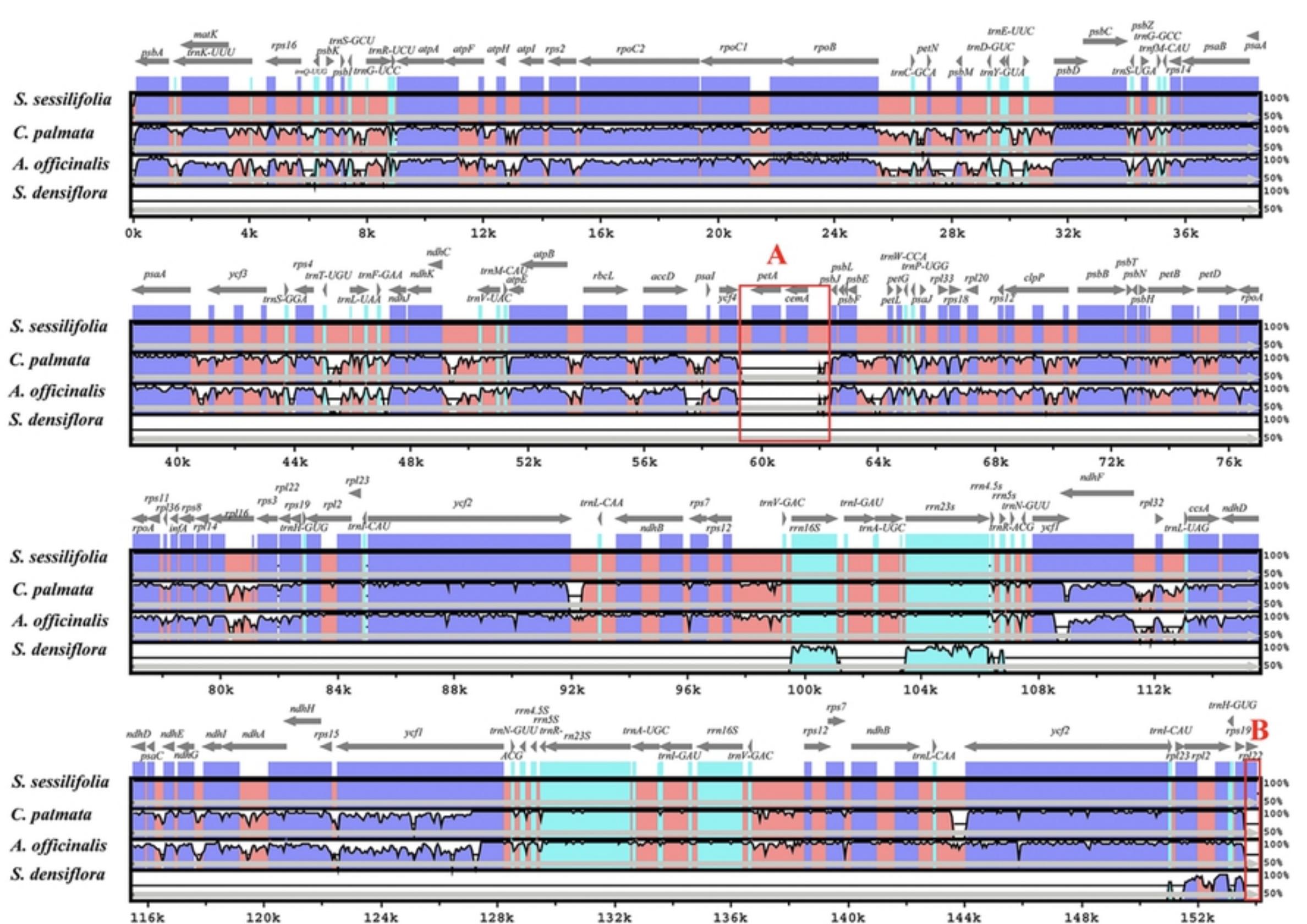


Figure 3

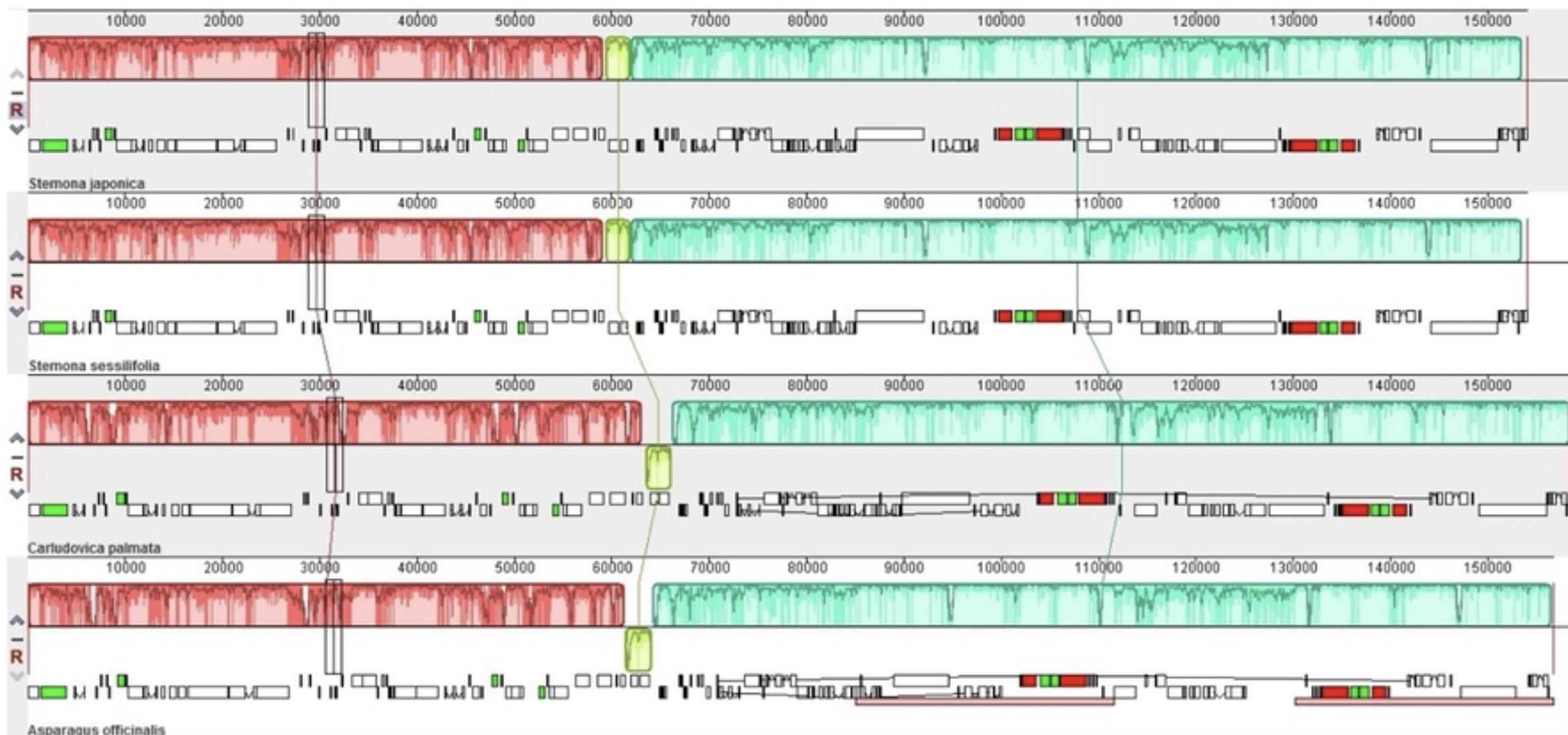


Figure4

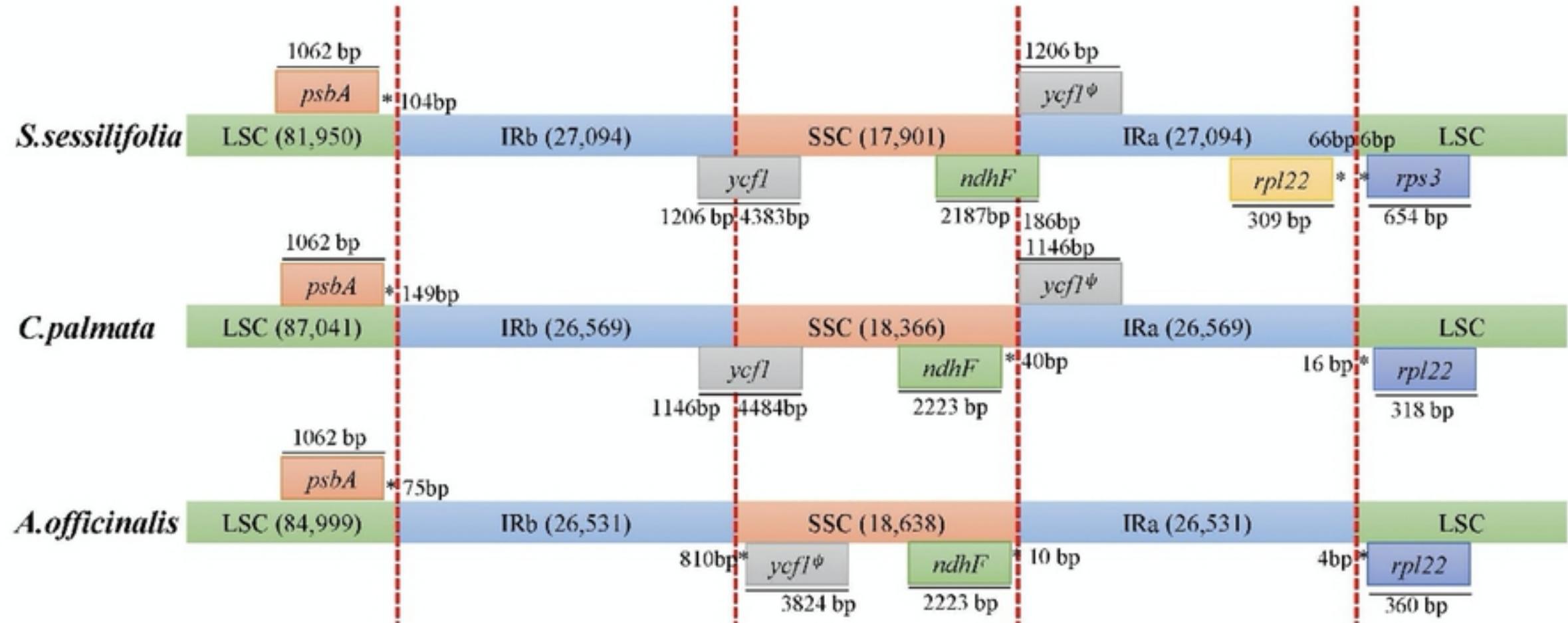


Figure 5

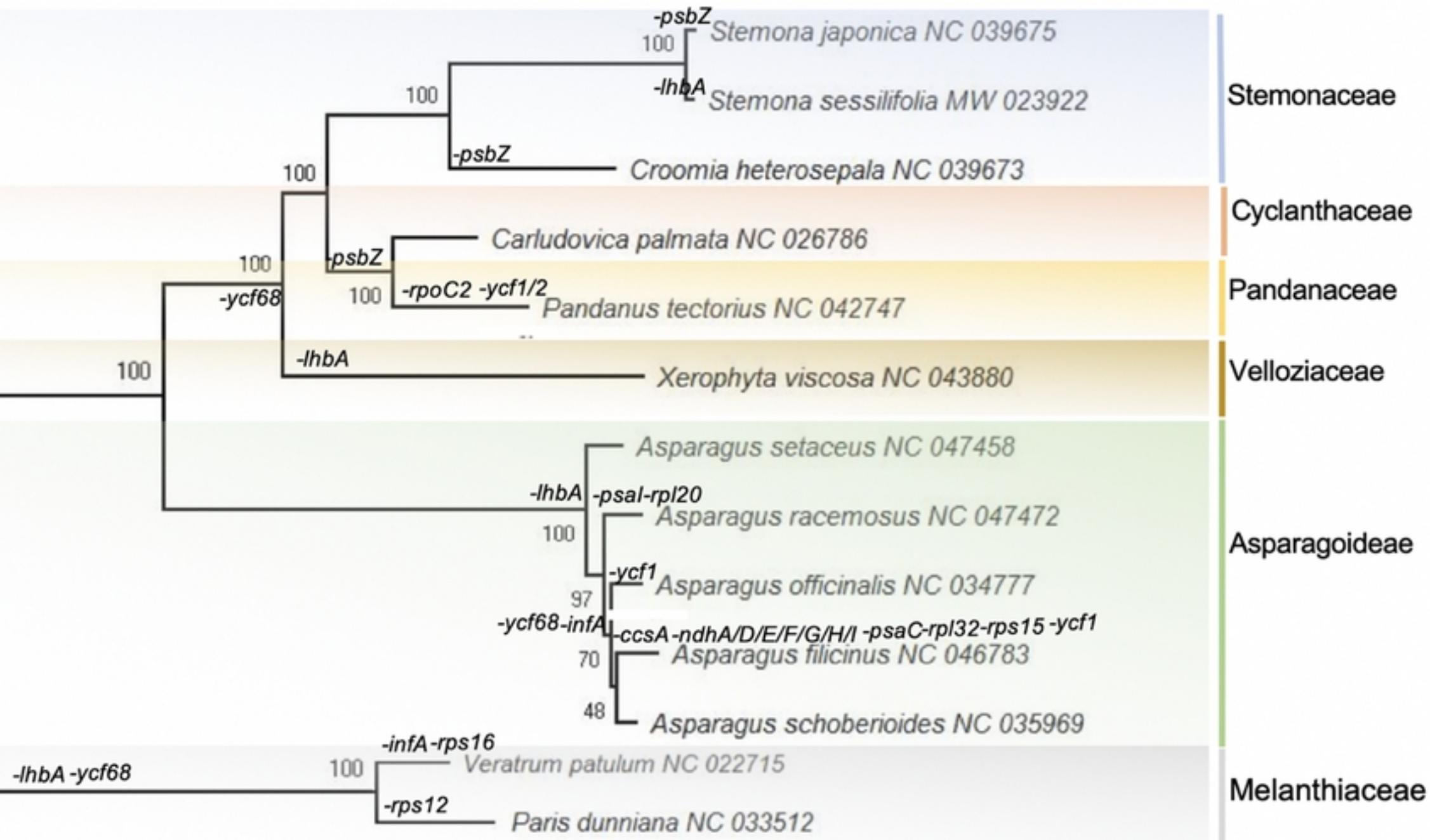


Figure 6