

RESEARCH ARTICLE

The complete chloroplast genome sequences of *Lychnis wilfordii* and *Silene capitata* and comparative analyses with other Caryophyllaceae genomes

Jong-Soo Kang^{1a,b}, Byoung Yoon Lee, Myounghai Kwak*

Plant Resources Division, National Institute of Biological Resources, Incheon, Republic of Korea

^{1a} Current address: State Key Laboratory of Systematic and Evolutionary Botany, Institute of Botany, Chinese Academy of Sciences, Beijing, China

^{1b} Current address: University of the Chinese Academy of Sciences, Beijing, China

* mhkwak1@korea.kr



OPEN ACCESS

Citation: Kang J-S, Lee BY, Kwak M (2017) The complete chloroplast genome sequences of *Lychnis wilfordii* and *Silene capitata* and comparative analyses with other Caryophyllaceae genomes. PLoS ONE 12(2): e0172924. doi:10.1371/journal.pone.0172924

Editor: Shihui Yang, National Renewable Energy Laboratory, UNITED STATES

Received: December 15, 2016

Accepted: February 10, 2017

Published: February 27, 2017

Copyright: © 2017 Kang et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: This work was supported by a grant from the National Institute of Biological Resources (NIBR) funded by the Ministry of Environment of the Republic of Korea (NIBR201503102).

Competing interests: The authors have declared that no competing interests exist.

Abstract

The complete chloroplast genomes of *Lychnis wilfordii* and *Silene capitata* were determined and compared with ten previously reported Caryophyllaceae chloroplast genomes. The chloroplast genome sequences of *L. wilfordii* and *S. capitata* contain 152,320 bp and 150,224 bp, respectively. The gene contents and orders among 12 Caryophyllaceae species are consistent, but several microstructural changes have occurred. Expansion of the inverted repeat (IR) regions at the large single copy (LSC)/IRb and small single copy (SSC)/IR boundaries led to partial or entire gene duplications. Additionally, rearrangements of the LSC region were caused by gene inversions and/or transpositions. The 18 kb inversions, which occurred three times in different lineages of tribe *Sileneae*, were thought to be facilitated by the intermolecular duplicated sequences. Sequence analyses of the *L. wilfordii* and *S. capitata* genomes revealed 39 and 43 repeats, respectively, including forward, palindromic, and reverse repeats. In addition, a total of 67 and 56 simple sequence repeats were discovered in the *L. wilfordii* and *S. capitata* chloroplast genomes, respectively. Finally, we constructed phylogenetic trees of the 12 Caryophyllaceae species and two Amaranthaceae species based on 73 protein-coding genes using both maximum parsimony and likelihood methods.

Introduction

Chloroplasts are important photosynthetic organelles that provide energy for the synthesis of glucose, fatty acids, and amino acids [1,2]. The chloroplast genome is the smallest of the plant genomes, ranging from 135 to 160 kb in most plants [3–5]. Most angiosperm chloroplast genomes have a quadripartite circular structure and contain two copies of inverted repeat (IR) regions, separating a large single copy (LSC) region and small single copy (SSC) region [5]. Recently, with the rapid development of next-generation sequencing platforms, many chloroplast genome sequences have been reported and used to help resolve plant phylogenies [6,7]. Chloroplast genomic data are widely used in various studies, such as those on molecular

phylogeny, molecular identification (DNA barcoding), and genetic diversity [8–10]. The structure and gene order of the chloroplast genome are stable, and the rates of nucleotide substitution are generally slow in angiosperms [11–14].

Rearrangements in the chloroplast genome were considered to have occurred rarely enough in evolution that they can be used to demarcate major groups [4]; however, recently some lineages have revealed various patterns of changes in chloroplast genomes, for example large-scale rearrangements, gene duplications, and even loss of IR regions [15–20]. Scattered angiosperm lineages show extensive rearrangement of plastid genomes, and these gene order changes are correlated with increased rates of nucleotide substitutions and gene and intron losses [6]. Rearrangements of the chloroplast genome are often associated with repeated sequences [5].

The family Caryophyllaceae consists of 75–80 genera and approximately 2,000 species, which are widely distributed, mainly in the temperate or warm-temperate regions of the northern hemisphere [21]. The genera *Lychnis* and *Silene* are sister genera belonging to tribe *Sileneae*, but the taxonomic identities and limitations between these two genera remain unclear [21–23], which is why the genus *Lychnis* was nested within *Silene* using nuclear ribosomal internal transcribed spacer (nrITS), five chloroplast genes, and intergenic spacers (IGS) [24]. Previous studies have shown that the *Sileneae* underwent accelerated plastid genome evolution, including inversions, shifts in IR boundaries, large indels, intron losses, and rapid rates of amino acid sequence substitution [25,26]. Interestingly, the *psaA-ycf3:psaI-ycf4* inversion and intron losses in *clpP-1* and *clpP-2* were suggested to be independent events that occurred three times [26].

A total of ten Caryophyllaceae chloroplast genomes have been reported [25–28]. In the genus *Lychnis*, only the chloroplast genome of *L. chalcedonica* has been reported [26], whereas in *Silene*, chloroplast genomes from a total of six species have been reported [25,26]. Therefore, in this study, we sequenced the complete chloroplast genomes of *L. wilfordii* and *S. capitata* and then analyzed them to identify their genetic characteristics and differences compared with other Caryophyllaceae species. The specific goals of the present study were to (1) present the complete chloroplast genome sequences of two *Sileneae* species, (2) investigate any significant characteristics suggesting extensive genome rearrangement in this tribe, and (3) explore significant changes in gene content and intron losses in the tribe *Sileneae*.

Materials and methods

Plant materials, DNA extraction, sequencing and genome assembly

Leaf materials from *Lychnis wilfordii* and *Silene capitata* were obtained from living plants by seed germination in a greenhouse at the Korean Botanical Garden. The voucher specimens of *L. wilfordii* (NIBRVP0000542331) and *S. capitata* (NIBRVP0000542433) were deposited in the National Institute of Biological Resources Herbarium (KB). Total genomic DNA was extracted using the Genome Wizard kit (Promega, Madison, WI, USA). Sequencing libraries were prepared using the NEXTflex Rapid DNA-seq kit (Bioo Scientific, Austin, TX, USA). Paired-end sequencing libraries containing insert sizes of approximately 350–450 bp were sequenced on the Illumina HiSeq 2500 platform (Illumina Inc., San Diego, CA, USA) at the National Instrumentation Center for Environmental Management (Seoul, South Korea), yielding 27,739,600 reads from *L. wilfordii* and 22,127,152 reads from *S. capitata*, each with a read length of 250 bp. These paired-end reads were aligned with sequences from *Silene vulgaris* (JF715057). After screening these paired-end reads through alignment with *S. vulgaris* plastid genome, 585,206 (2.1%) reads of *L. wilfordii* and 661,807 (2.9%) reads of *S. capitata* were extracted with mean of coverage 980× and 1082×, respectively. *De novo* assembly was performed using Geneious v. 7.1.3 (Biomatters, Auckland, New Zealand). The consensus sequences were extracted and gap-

filled by PCR amplification using specific primers based on the gaps between sequences. The PCR products were purified and sequenced by Sanger sequencing.

Genome annotation and comparative analyses

The initial annotation of the two Caryophyllaceae chloroplast genomes was performed using Dual Organellar GenoMe Annotator (DOGMA) [29]. From this initial annotation, putative starts, stops, and intron positions were determined by comparison with homologous genes in other Caryophyllaceae chloroplast genomes. The tRNA genes were annotated using DOGMA and tRNAscan-SE [30]. The circular chloroplast genome map was drawn using the OGDRAW program [31]. The complete chloroplast genomes of *L. wilfordii* and *S. capitata* were compared with those of ten other Caryophyllaceae species using the mVISTA program in Shuffle-LAGAN mode (Table 1) [32]. *Agrostemma githago* (KF527884) was used as a reference.

Repeat sequence analysis

Simple sequence repeats (SSRs or microsatellites; mono-, di-, tri-, tetra-, penta-, and hexanucleotide repeats) were detected using Phobos v. 3.3.12 [33] with thresholds of ten repeat units for mononucleotide SSRs, five repeat units for di- and trinucleotide SSRs, and three repeat units for tetra-, penta-, and hexanucleotide SSRs. REPuter [34] was also used to analyze the repeat sequences, which included forward, reverse, palindromic, and complementary sequences with a minimal length of 30 bp and 90% sequence identities (Hamming distance of three). Moreover, we constructed a phylogenetic trees based on the sequences of the pairs of repeat regions to investigate the relationship between the distributions of repeat sequences and structural inversions. Maximum parsimony (MP) analysis was conducted using PAUP v. 4.0a150 [35], and branch support was assessed using 1000 bootstrap replicates.

Phylogenetic analysis

Phylogenetic analyses based on 73 protein-coding genes were also performed for 12 Caryophyllaceae species, using two Amaranthaceae species (*Beta vulgaris* and *Salicornia europaea*) as the outgroup (Table 1). Among 77 whole protein-coding genes, *ycf1*, *ycf2*, *accD*, *clpP* genes were excluded from data matrix, since those genes were reported fast evolving genes with high

Table 1. GenBank accession numbers and references used in this study.

Family name	Scientific name	Accession numbers	Reference
Caryophyllaceae	<i>Agrostemma githago</i>	KF527884	Sloan et al., 2014
	<i>Colobanthus quitensis</i>	KT737383	Kang et al., 2015
	<i>Dianthus longicalyx</i>	KM668208	Gurusamy et al., 2016
	<i>Lychnis chalconica</i>	KF527886	Sloan et al., 2014
	<i>Lychnis wilfordii</i>	KT727929	In this study
	<i>Silene capitata</i>	KT727930	In this study
	<i>Silene conica</i>	JF715054	Sloan et al., 2012
	<i>Silene conoidea</i>	KF527885	Sloan et al., 2014
	<i>Silene latifolia</i>	JF715055	Sloan et al., 2012
	<i>Silene noctiflora</i>	JF715056	Sloan et al., 2012
	<i>Silene paradoxa</i>	KF527887	Sloan et al., 2014
	<i>Silene vulgaris</i>	JF715057	Sloan et al., 2012
Amaranthaceae	<i>Beta vulgaris</i>	KR230391	Stadermann et al., 2015
	<i>Salicornia europaea</i>	KJ629116	Unpublished

doi:10.1371/journal.pone.0172924.t001

substitution rate within tribe *Sileneae* [26]. Consequently, a total of 54,271 bp were aligned using MAFFT [36]. MP analysis was conducted using PAUP v. 4.0a150 [35], and branch support was assessed using 1000 bootstrap replicates. Before Maximum likelihood (ML) analysis, a search for the best fitting substitution model was performed using jModeltest v. 2.1.5 [37]. Based on the Akaike Information Criterion (AIC) and Akaike Information Criterion with Correction (AICc), GTR+I+G was the best model. ML analysis was performed using RAxML v. 7.4.2 with 1000 bootstrap replicates and the GTR+I+G model [38]. Bayesian inference was performed using MrBayes 3.2 [39].

Results and discussion

Genome organization and features

The complete sizes of the *L. wilfordii* and *S. capitata* chloroplast genomes are 152,320 and 150,224 bp, respectively (Fig 1, Table 2). The size of the *L. wilfordii* chloroplast genome is the longest among the reported Caryophyllaceae species. The *L. wilfordii* and *S. capitata* genomes include a pair of IRs of 27,709 bp and 25,371 bp separated by a SSC region of 12,914 bp and 17,313 bp and a LSC region of 83,988 bp and 82,169 bp, respectively (Fig 1, Table 2), similar to the published Caryophyllaceae chloroplast genomes [25–28]. The *L. wilfordii* chloroplast genome contains 110 unique genes, 17 of which are duplicated in the IR region, giving a total of 127 genes (Fig 1, Table 2, S1 Table). The *S. capitata* chloroplast genome contains 111 unique genes, 19 of which are duplicated in the IR region, giving a total of 130 genes (Fig 1, Table 2, S2 Table). The chloroplast genomes of these two species contain 30 distinct tRNAs, seven of which are duplicated in the IR region. Seventeen genes contain one or two introns: 14 contain

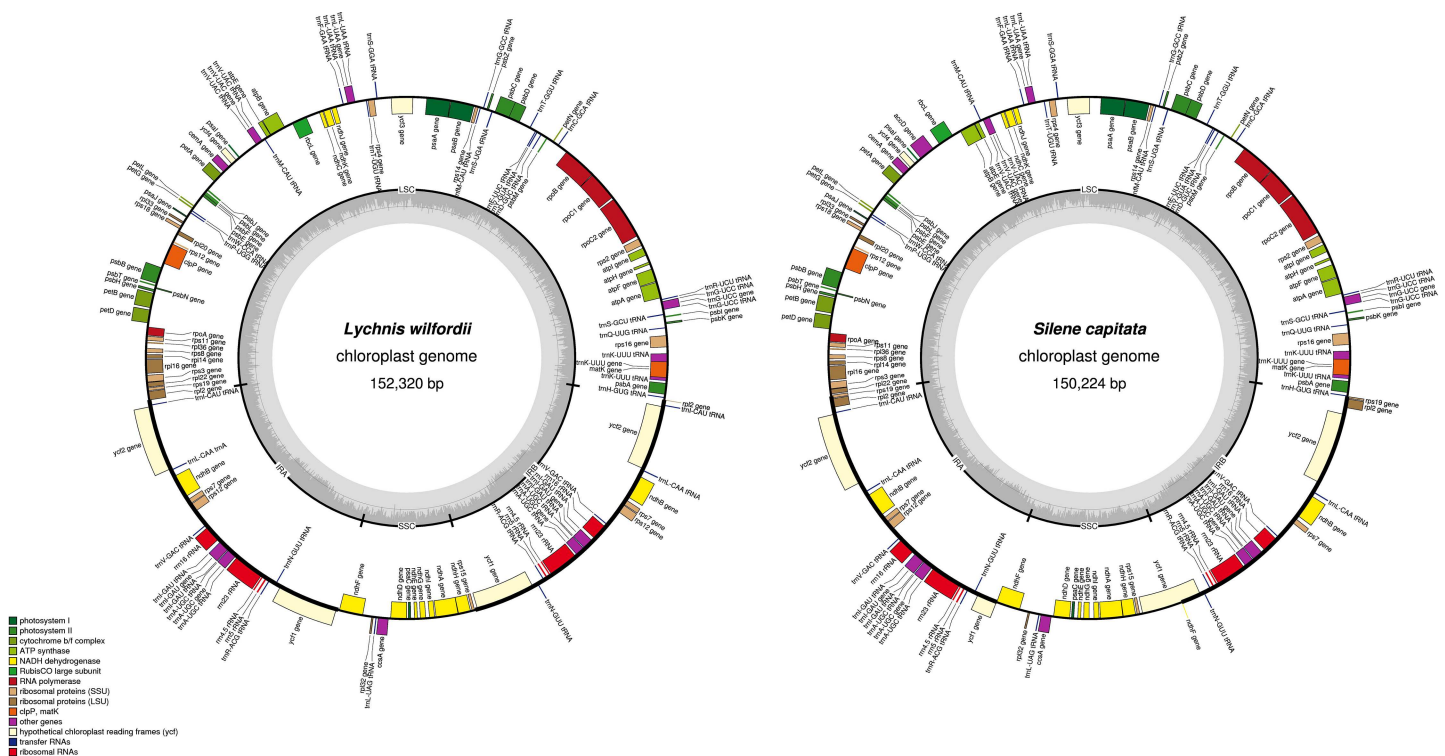


Fig 1. The chloroplast genomes of *Lychnis wilfordii* and *Silene capitata*. Genes inside the circle are transcribed clockwise, while genes outside are transcribed counter-clockwise. The dark gray inner circle corresponds to the GC content and the light-gray circle to the AT content.

doi:10.1371/journal.pone.0172924.g001

Table 2. Summary of chloroplast genome characteristics of two caryophyllaceae genomes.

Genome features	<i>Lychnis wilfordii</i>	<i>Silene capitata</i>
Size (bp)	152,320	150,224
LSC length (bp)	83,988	82,169
SSC length (bp)	12,914	17,313
IR length (bp)	27,709	25,371
Number of genes	110	111
Protein-coding genes	76+6	77+8
tRNA genes	30+7	30+7
rRNA genes	4+4	4+4
Number of genes duplicated in IR	17	19
Overall GC content (%)	36.6	36.3
GC content in LSC (%)	34.9	34.2
GC content in SSC (%)	31.2	29.8
GC content in IR (%)	40.5	42.2

doi:10.1371/journal.pone.0172924.t002

one intron and three (*rps12*, *clpP*, and *ycf3*) two introns. Six of the genes containing one intron are tRNAs (S1 and S2 Tables).

In addition, while the *L. wilfordii* and *S. capitata* chloroplast genomes both have lost the *infA* gene, the *accD* gene was pseudogenized only in *L. wilfordii*. The lack or pseudogenization of the *infA* gene has been discovered in many taxa outside of Caryophyllaceae, such as the Brassicaceae, Fabaceae, Liliaceae, Malvaceae, and Onagraceae [25,26,40–44]. Loss or pseudogenization of the *accD* gene in the plastid genome or *accD* gene transfer to the nucleus has also been reported in various angiosperm lineages, including Poaceae, Orobanchaceae, Ericaceae, and Primulaceae [45–48].

Comparative chloroplast genomic analysis

We compared gene arrangements in the chloroplast genomes of *L. wilfordii* and *S. capitata* with those of the ten previously reported Caryophyllaceae species (Fig 2). The chloroplast genome of *S. capitata* has an identical gene order with those of the genera *Agrostemma*, *Colobanthus*, and *Dianthus*, but the chloroplast genome of *L. wilfordii* has unique structural changes compared with previously reported Caryophyllaceae chloroplast genomes (Fig 2). The gene rearrangements present in the LSC regions were a result of inversions and/or transpositions (Fig 2). The chloroplast genome of *L. wilfordii* revealed an inversion of the *trnV-rbcL* region compared with the genomes of other genera (Fig 2), whereas the *L. chalconica* genome had twice the number of inversions and transpositions in the *accD-psaI* and *ycf3* regions compared with the genomes of other genera (Fig 2). Interestingly, truncated partial sequences of *clpP-2* and *accD* were found in the IGS region between *trnV* and *psaI*. The 5' upstream non-genic region and a partial 348 bp sequence of the *accD* gene, as well as the exon 1 and partial intron 1 sequences of *clpP-2*, have remained, but the downstream regions of both genes were truncated in the *L. wilfordii* chloroplast genome. Compared with the gene orders in other chloroplast genomes, these disruptions in the *accD* and *clpP-2* genes may have occurred by inversion of the *trnV-rbcL* fragment. Thus, we deduced that duplication of *clpP* occurred before diversification of *L. chalconica* from *L. wilfordii*, and that transposition of *psaI-accD* and the loss of introns in *clpP-1* and *clpP-2* in *L. chalconica* may have occurred after species diversification.

In the genus *Silene*, we identified three types of chloroplast genomes (Fig 2). These are a) the common type of chloroplast genome observed in most Caryophyllaceae (*Agrostemma*, *Colobanthus*, and *Dianthus*) (seen in *S. capitata*, *S. latifolia*, and *S. vulgaris*); b) chloroplast

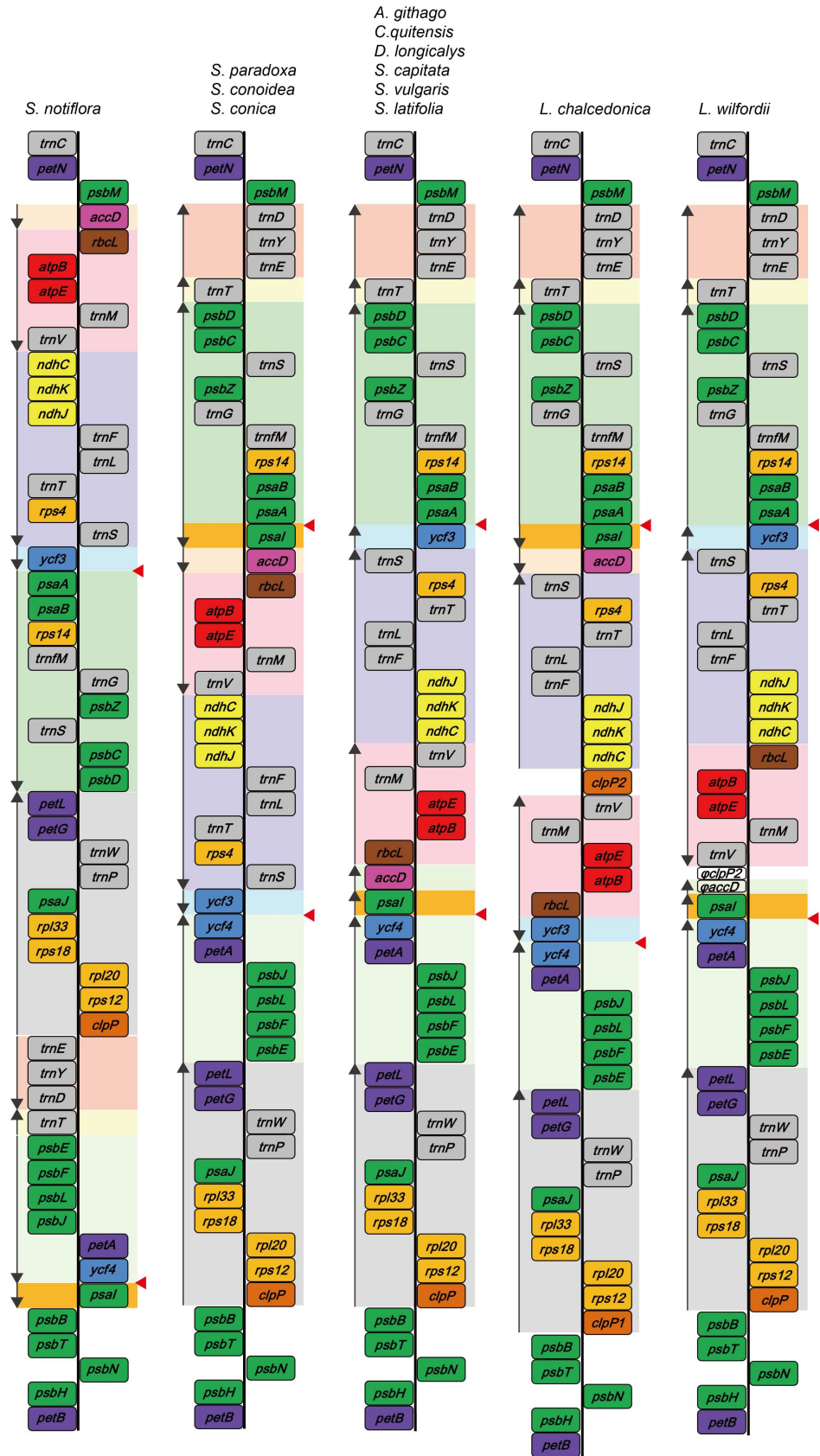


Fig 2. Comparison of gene rearrangements in the large single copy region among 12 Caryophyllaceae. Genes are indicated in the colored boxes. Green: photosystem; blue: hypothetical chloroplast reading frame (*ycf* series); yellow: NADH-dehydrogenase; light orange: ribosomal subunit; dark orange: protease; brown: rubisco subunit; red: ATP synthase; purple: cytochrome *b/f* complex; pink: acetyl-CoA carboxylase; gray: tRNA; white: pseudogene. The larger boxes indicate that the inversion or transposition fragments have been identified. The arrows to the left of the large boxes indicate the direction of inversion compared with the ancestral large single copy gene order of this region. The red triangle to the right of the large squares indicates the breaking point of an 18 kb inversion with intermolecular duplicated sequences.

doi:10.1371/journal.pone.0172924.g002

genomes exhibiting an inversion of the *ycf3-psaI* regions (seen in *S. paradoxa*, *S. conoidea*, and *S. conica*); c) chloroplast genomes exhibiting transpositions and/or inversions of the *psbD-accD*, *petL-clpP*, *trnD-T*, and *psaI-psbE* regions (seen in *S. noctiflora*). *Silene noctiflora* currently has the most complicated chloroplast genome among the Caryophyllaceae.

Overall sequence identity was analyzed with mVISTA program among the 12 chloroplast genomes of Caryophyllaceae, using the *Agrostemma githago* genome as a reference (Fig 3). The results revealed higher divergence in the LSC regions than in the IRs and SSCs, as a result of gene rearrangements (Figs 2 and 3), and greater conservation in the coding regions than in the non-coding regions (Fig 3). The most divergent coding regions were the *ycf1*, *ycf2*, *accD*, and *clpP* genes, which are similar to results from previous studies [25,26,49], showing lower (under 50%) similarity compared with other protein-coding regions (Fig 3). Consequently, we suggest that these genes evolve rapidly in Caryophyllaceae (including the tribe *Sileneae*). These genes are either absent or highly variable in the genomes of Campanulaceae, Geraniaceae, and Poaceae [6].

Boundaries between single copy and inverted repeat regions

The size variations among angiosperm chloroplast genomes are mostly the result of expansion or contraction of the IR region [50]. Additionally, the expansion or contraction of the IR region differs among various plant species [51]. In this study, the LSC-IR and IR-SSC boundaries of the 12 sequenced Caryophyllaceae genomes were compared (Fig 4). IR locations have changed substantially in *Lychnis* and *Silene* as a result of movement of the boundaries between the IR and SC regions (Fig 4). The IR and SC boundaries of *S. capitata* are consistent with those of the *S. vulgaris* and *S. latifolia* genomes, as well as the Caryophyllaceae genera *Agrostemma*, *Colobanthus*, and *Dianthus* (Fig 4).

The expansion of the IR at the SSC/IR boundary that duplicates the entire *ycf1* gene was found only in the genome of *L. wilfordii* and three *Silene* species (*S. conica*, *S. conoidea*, and *S. noctiflora*). This event was observed in non-core Caryophyllales [52]. In the *S. noctiflora* chloroplast genome, the *ycf1* and *rps15* genes are duplicated within the IR region (Fig 4), and this species contains the longest IR region (29,891 bp) among the 12 Caryophyllaceae species. However, the contraction of the IR at the LSC/IR boundary that duplicates a part of the *rpl2* gene was found only in *Silene* (*S. conica*, *S. conoidea*, and *S. noctiflora*) and *Lychnis* (*L. wilfordii* and *L. chalcedonica*). *Lychnis chalcedonica* has the shortest IR region (23,540 bp) among 12 Caryophyllaceae species due to contraction of the IR region at the LSC/IR boundary and lack of expansion of the IR region at the IR/SSC boundary.

Repeat sequence analysis and short inverted repeats as inversion hotspots

We analyzed repeat sequences from the chloroplast genomes of *L. wilfordii* and *S. capitata* and observed forward, reverse, palindromic, and complementary repeats using REPuter. *Lychnis*

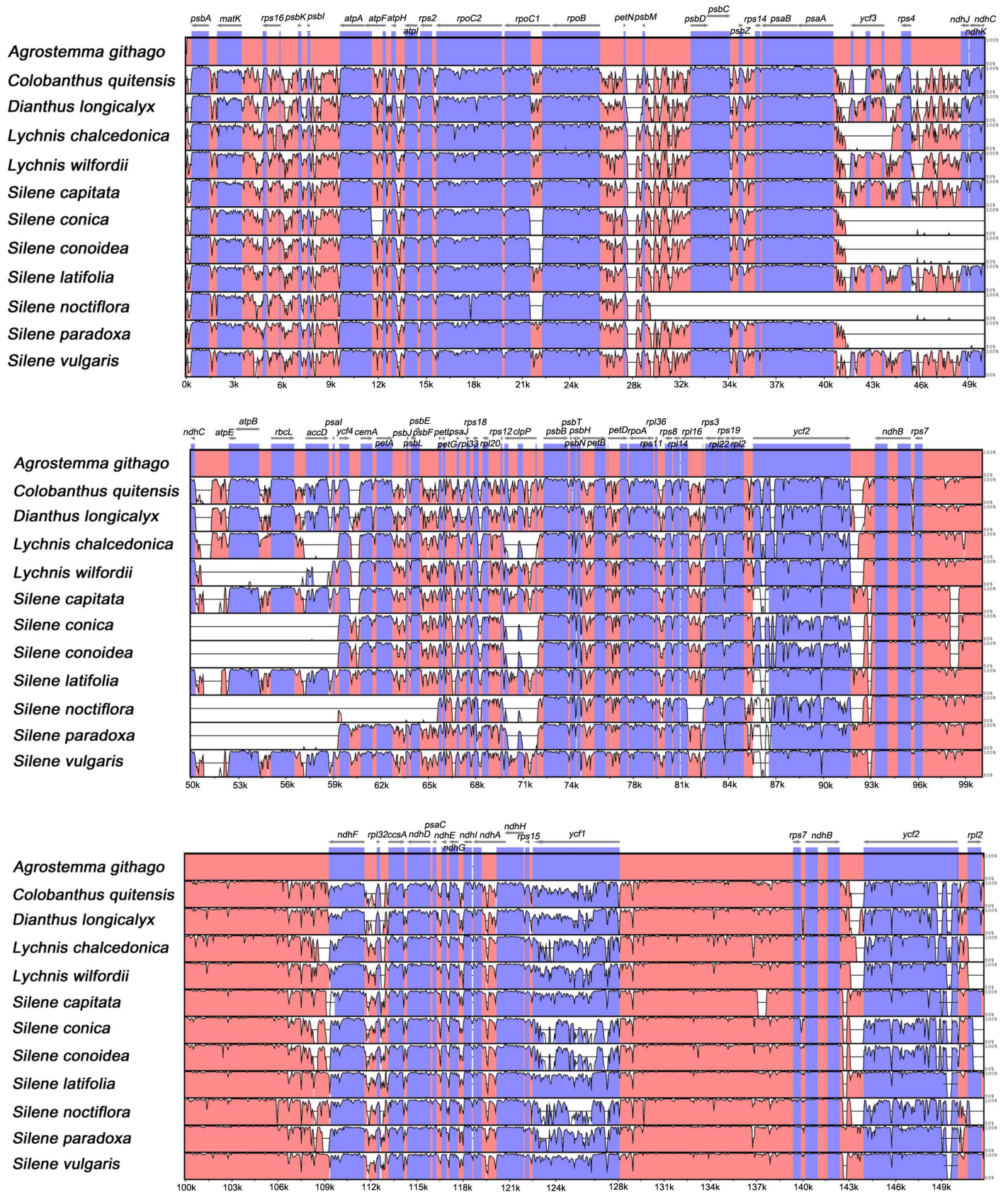


Fig 3. Sequence alignment of 12 Caryophyllaceae genomes in mVISTA, using the *Agrostemma githago* genome as a reference. The vertical scale indicates the identity percentage, ranging from 50 to 100%.

doi:10.1371/journal.pone.0172924.g003

wilfordii contains 17 forward repeats and 22 palindromic repeats, whose lengths range from 40 to 462 bp (Fig 5, S3 Table). *Silene capitata* contains 15 forward repeats, 27 palindromic repeats, and only one reverse repeat, whose lengths range from 30 to 64 bp (Fig 5, S4 Table). Most of the *L. wilfordii* repeats are located in IGS regions (56.4%), and less than half were located in genes (30.8%; *ycf1* and *ycf2*) and introns (12.8%; *ndhA* and *clpP* intron). In contrast, the majority of *S. capitata* repeats are located in genes (44.0%; *ycf2*, *ycf4*, *psaA*, *psaB*, *ccsA*, *trnS-GGA*, *trnS-GCU*, *trnS-UGA*, *trnG-UCC*, and *trnG-GCC*), with fewer located in IGSs (40.0%) and introns (16.0%; *ycf3*, *rpl16*, *rpoC1*, and *ndhA* introns).

We then analyzed the SSRs (or microsatellites), which are increasingly evaluated in molecular genetic studies because of their high reproducibility, ease of scoring, and fast throughput compared with other marker techniques [53]. In the *L. wilfordii* and *S. capitata* chloroplast genomes, the most abundant SSRs were A or T mononucleotide repeats, which accounted for approximately 77.6% and 76.8% of the total SSRs, followed by tetranucleotides (10.4% and 16.1%) and dinucleotides (10.4% and 7.1%), respectively (Table 3, S5 and S6 Tables). SSRs in the chloroplast genome are commonly composed of A or T repeats and rarely G or C repeats [54,55]. Furthermore, the majority of *L. wilfordii* and *S. capitata* SSRs are located in IGS regions (49.3% and 55.4%), followed by genes (37.3% and 26.8%) and introns (13.4% and 17.9%), respectively (S5 and S6 Tables). SSRs located in coding regions were found mainly in *ycf1* and *rpoC2*, with the remaining SSRs found in *matK*, *rpoA*, *psbF*, *atpB*, and *atpF*. Among the SSRs in genes, part or all of those in *matK*, *rpoC2*, *rpoA*, *psbF*, *ycf1*, and *rrn23* were shared by the two Caryophyllaceae species.

Under the assumption that the common chloroplast types observed in most Caryophyllaceae (*Agrostemma*, *Colobanthus*, and *Dianthus*, *S. capitata*, *S. latifolia*, and *S. vulgaris*) are ancestral, the inversion of the *ycf3-psaI* fragment might have occurred independently at least three times: in *L. chalconica*, *S. notiflora*, and the lineage containing *S. conoidea* and *S. conica*, consistent with previous results [26]. Interestingly, loss of introns in the *clpP* gene is always coupled with these inversions. In Caryophyllaceae, all 12 species possess imperfect palindromic repeats on both sides of the *ycf3-psaI* fragment (Fig 2), whereas only one homologous sequence corresponding to these repeats was found in the intergenic region between *psaI* and *ycf4* in two Amaranthaceae species. In all cases, the repeat sequences were overlapped by partial *ycf4* coding region sequences (63 bp). Thus, the partial *ycf4* and upstream sequences might have been duplicated in the IGS between *psaA* and *ycf3* before diversification of Caryophyllaceae. Even in the repeats between *psaA* and *ycf3*, expected to be non-genic sequences, intermolecular duplicated sequences were grouped together in *A. githago*, *C. quitensis*, *D. longicalix*, and *S. paradoxa* based on the maximum parsimony tree (Fig 6). The intermolecular duplicated sequences were not grouped together in the other *Silene* species. A large fragment inversion mediated by short IRs was reported in several plant species. The 22 kb inversion in Asteraceae [56], the 42 kb inversion in *Abies* of the Pinaceae [57], the 21 kb inversion in *Jasminae* of the Oleaceae [16], and the 36 kb inversion in the core Genistoids are thought to be induced by IRs in tRNAs or repeat elements several base pairs long. These dispersed repeats were shown to promote inversions via intermolecular recombination [5,58,59]. Thus, we suggest that this short IR in Caryophyllaceae might mediate intramolecular flip-flop recombination events, and thus, independent identical inversion events of the *ycf3-psaI* 18 kb fragment might be facilitated independently in different lineages.

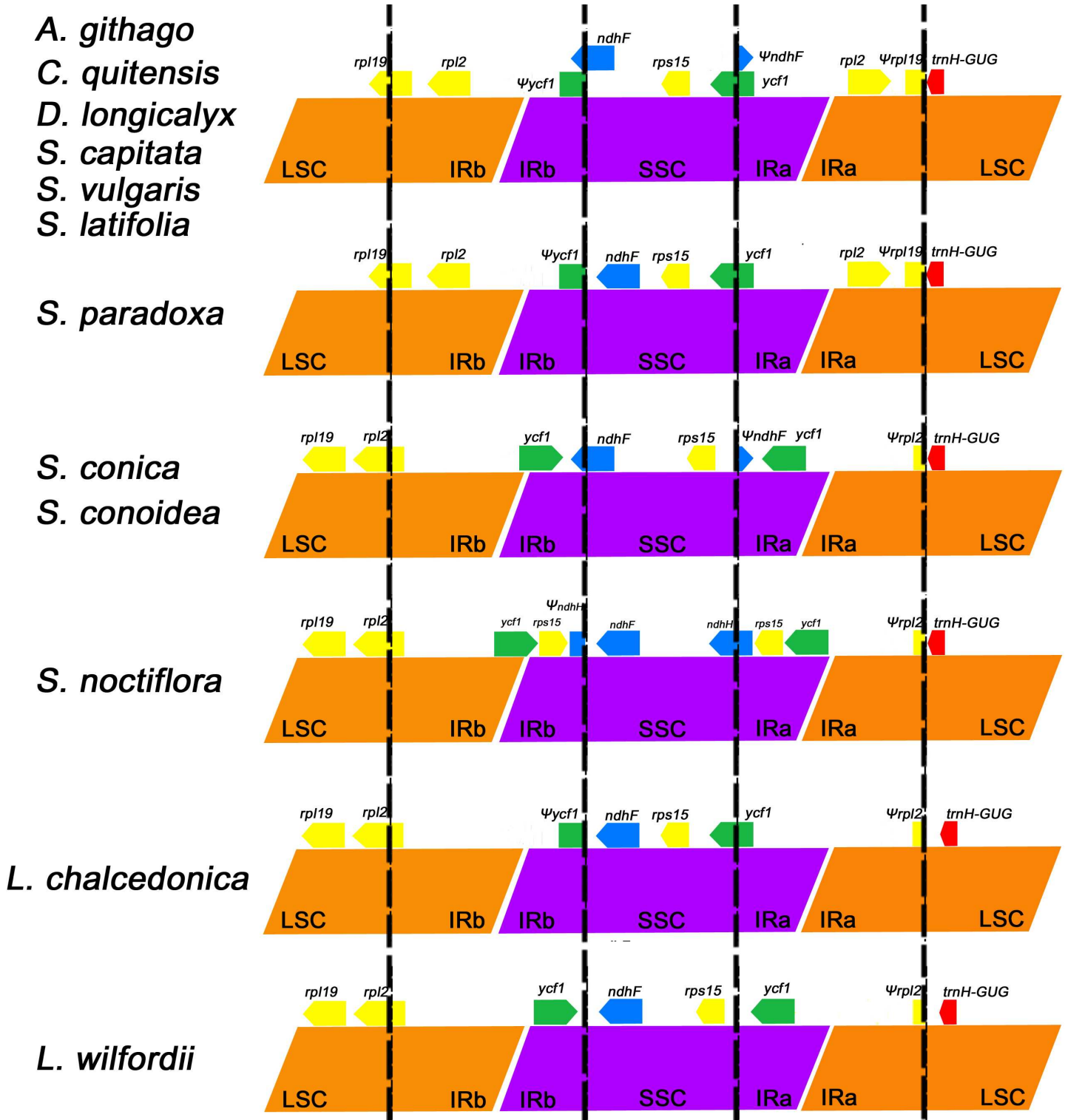


Fig 4. Comparison of the large single copy, inverted repeat, and small single copy border regions among 12 Caryophyllaceae chloroplast genomes.

doi:10.1371/journal.pone.0172924.g004

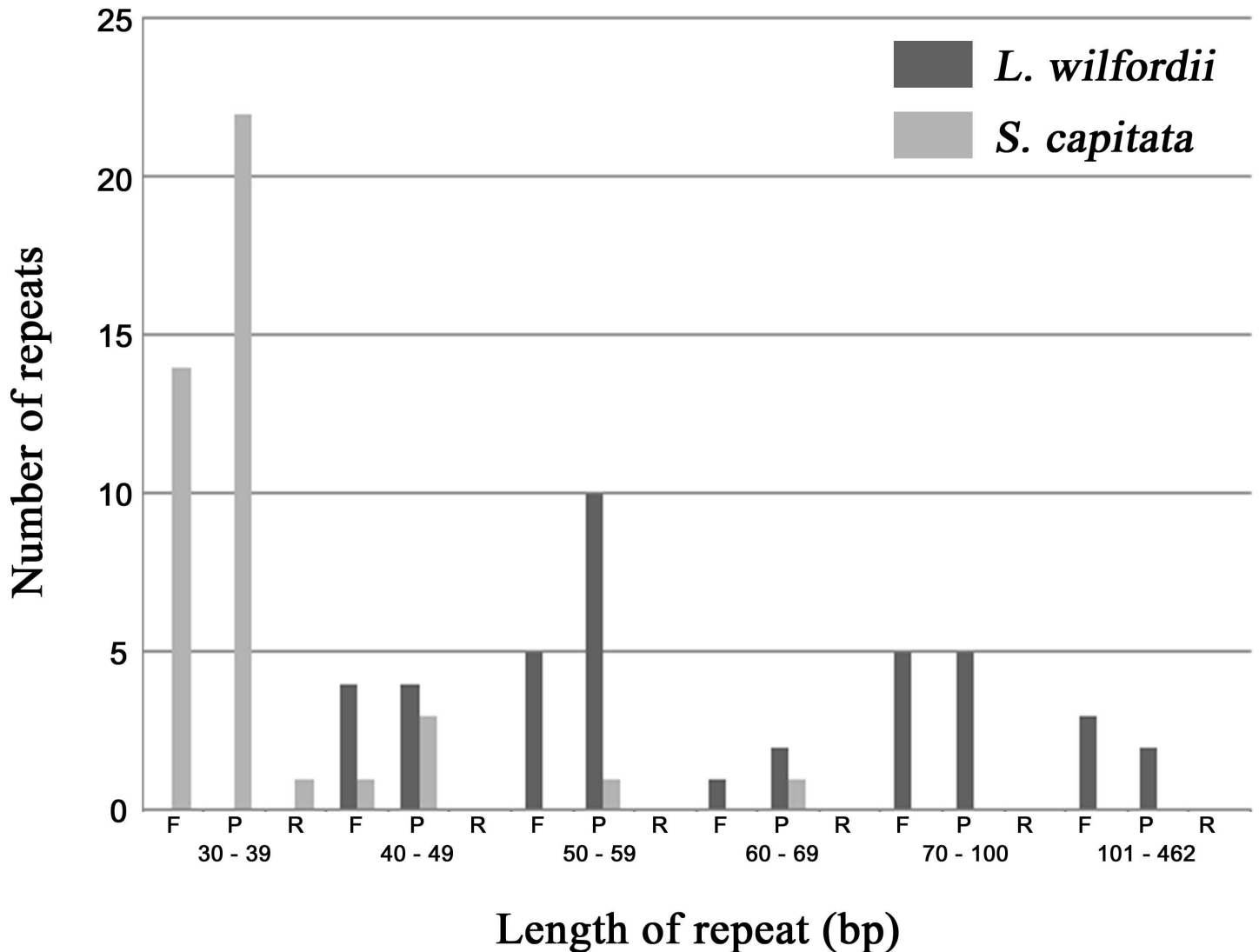


Fig 5. Frequency of repeat sequences in the chloroplast genomes of *Lychnis wilfordii* and *Silene capitata* using REPuter.

doi:10.1371/journal.pone.0172924.g005

Phylogenetic analysis

Both the MP and ML trees of the 12 Caryophyllaceae species and two Amaranthaceae species based on 73 protein-coding genes showed consistent phylogenetic patterns (Fig 7, S1 Fig). In

Table 3. The types and number of SSRs in chloroplast genomes of *L. wilfordii* and *S. capitata*.

	<i>L. wilfordii</i>	<i>S. capitata</i>
Mononucleotide repeats	52	43
Dinucleotide repeats	7	4
Trinucleotide repeats	1	0
Tetranucleotide repeats	7	9
Total SSRs in cp genome	67	56
Total SSRs in protein-coding genes	25	14

Penta- and hexanucleotides were not discovered over than three repeats.

doi:10.1371/journal.pone.0172924.t003

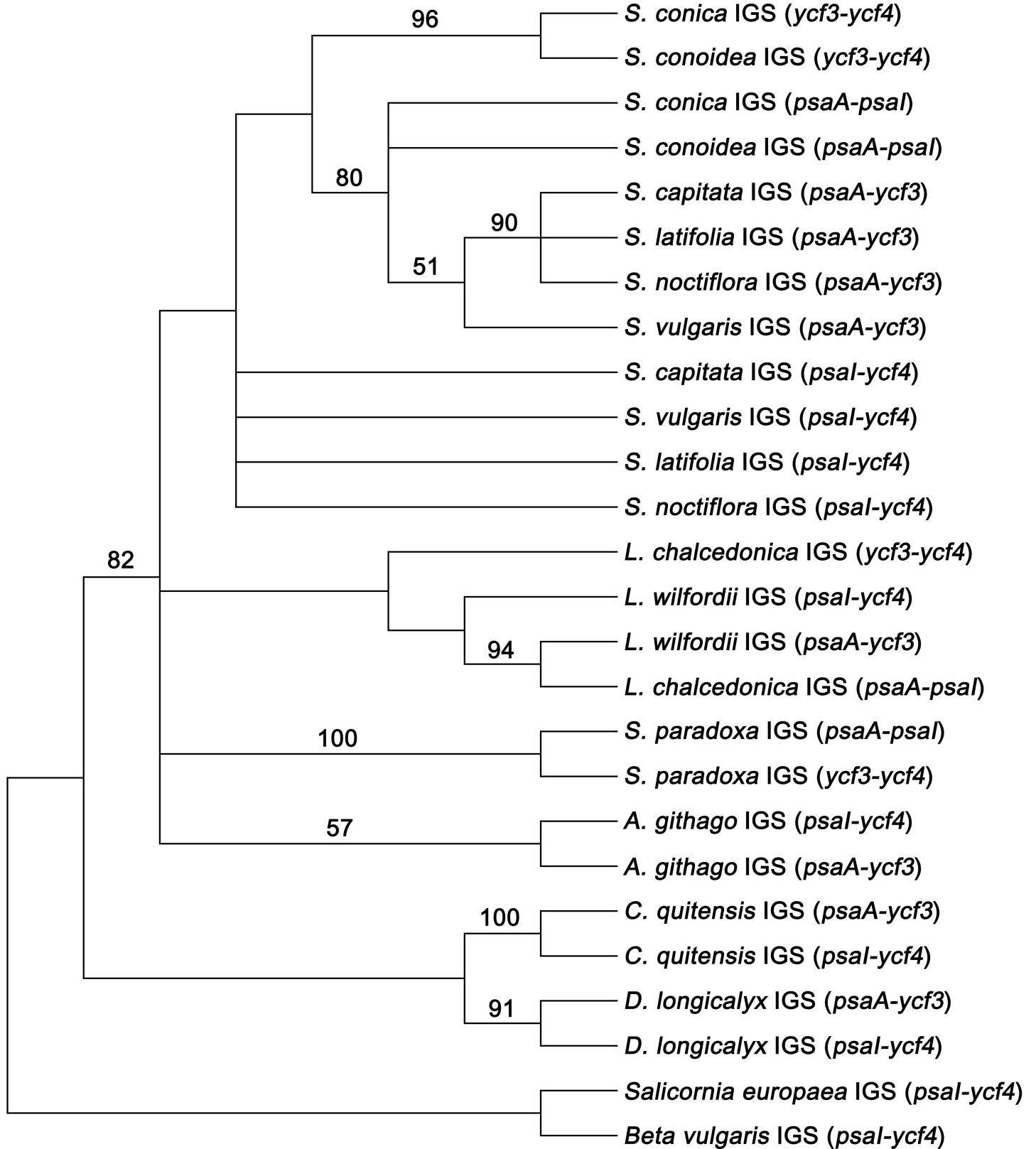


Fig 6. A phylogenetic tree of the inverted repeats at the end of the 18 kb inversion of the *psal-ycf3* fragment. Two intermolecular repeat sequences on both sides of the inversion were extracted. For two Amaranthaceae species (*Salicornia europaea* and *Beta vulgaris*), only one homologous sequence was present. Gene names in parentheses indicate the intergenic location between the two genes where both repeat sequences are present.

doi:10.1371/journal.pone.0172924.g006

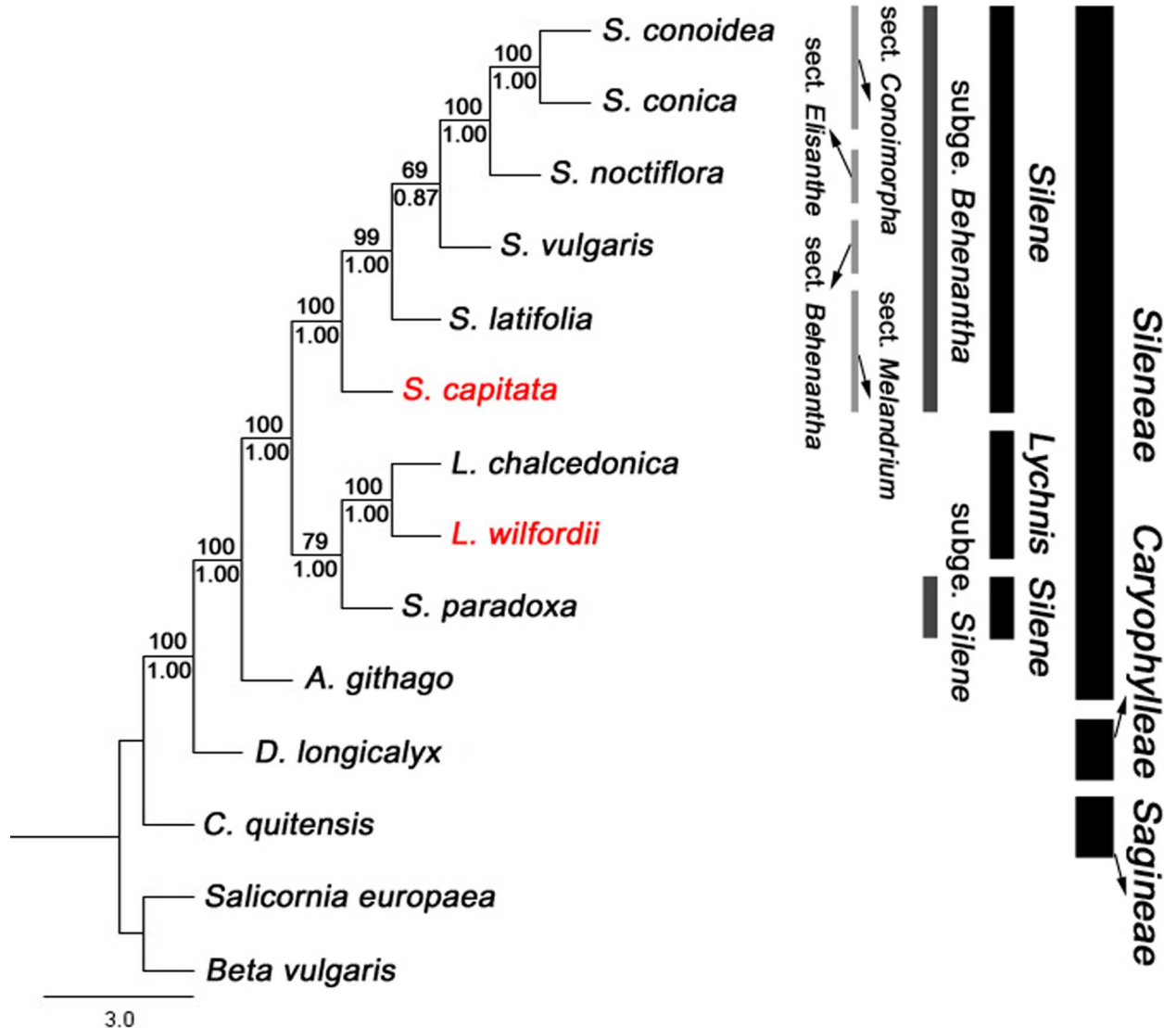


Fig 7. Phylogenetic tree of 14 taxa based on 73 protein-coding genes using the maximum likelihood method. Taxa in red are the new genomes reported in this study. Bootstrap values greater than 50% are shown above the nodes, and the Bayesian posterior probabilities are shown below the nodes.

doi:10.1371/journal.pone.0172924.g007

the ML tree, bootstrap analysis indicated that eight of ten nodes were supported by bootstrap values $\geq 99\%$ and the other two nodes by values $> 65\%$. In previous studies, the genus *Lychnis* was shown to be nested within the genus *Silene* based on internal transcribed spacer (ITS) sequences of nuclear genome and chloroplast DNA data [24,60] and based on chloroplast genome data [26]. *Lychnis* species were nested within *Silene*, close to *S. paradoxa* in the subgenus *Silene*, which is consistent with previous studies [26] (Fig 7). The subgenus *Behenantha* and monophyly of sect. *Melandrium* (*S. capitata* and *S. latifolia*) were not supported, whereas *S. conoidea* and *S. conica* of sect. *Conoimorpha* form a monophyletic group were found to be closely related to *S. noctiflora* of sect. *Elisanthe* (Fig 7). However, we need additional chloroplast genome data from more *Sileneae* species to resolve the relationship between *Lychnis* and *Silene*, as well as the infrageneric relationships of *Silene*.

Supporting information

S1 Fig. Phylogenetic tree of 14 taxa based on 73 protein-coding genes obtained using the maximum parsimony (MP) method in PAUP. Bootstrap values greater than 50% are shown above the nodes.

(TIF)

S1 Table. List of genes present in the chloroplast genome of *Lychnis wilfordii*.

(DOCX)

S2 Table. List of genes present in the chloroplast genome of *Silene capitata*.

(DOCX)

S3 Table. List of repeat sequences in the chloroplast genome of *Lychnis wilfordii*.

(DOCX)

S4 Table. List of repeat sequences in the chloroplast genome of *Silene capitata*.

(DOCX)

S5 Table. List of simple sequence repeats in the chloroplast genome of *Lychnis wilfordii*.

(DOCX)

S6 Table. List of simple sequence repeats in the chloroplast genome of *Silene capitata*.

(DOCX)

Acknowledgments

This work was supported by a grant from the National Institute of Biological Resources (NIBR) funded by the Ministry of Environment of the Republic of Korea (NIBR201503102). The authors thank Young Chul Kim for help with collecting the samples.

Author Contributions

Conceptualization: MK.

Data curation: JK.

Formal analysis: JK.

Funding acquisition: MK BL.

Investigation: JK.

Methodology: MK JK.

Project administration: MK BL.

Resources: JK.

Software: JK.

Supervision: MK.

Validation: MK.

Visualization: JK.

Writing – original draft: MK JK.

Writing – review & editing: BL.

References

1. Neuhaus HE, Emes MJ. Nonphotosynthetic metabolism in plastids. *Annu Rev Plant Biol.* 2000; 51: 111–140.
2. Rodríguez-Ezpeleta N, Brickmann H, Burey SC, Roure B, Burger G, Löffelhardt W, et al. Monophyly of primary photosynthetic eukaryotes: green plants, red algae, and glaucophytes. *Current Biology.* 2005; 15(14): 1325–1330. doi: [10.1016/j.cub.2005.06.040](https://doi.org/10.1016/j.cub.2005.06.040) PMID: [16051178](https://pubmed.ncbi.nlm.nih.gov/16051178/)
3. Downie SR, Palmer JD. Use of chloroplast DNA rearrangements in reconstructing plant phylogeny. In: Soltis PS, Soltis DE, Doyle JJ. *Molecular Systematics of Plants*, Springer US; 1992. pp. 14–35.
4. Judd WS, Campbell CS, Kellogg EA, Stevens PF, Donoghue MJ. *Plant systematics: a phylogenetic approach*. 2nd ed. Sinauer Associates, Inc., Sunderland, Massachusetts. USA; 2002.
5. Palmer JD. Plastid chromosomes: structure and evolution. In: Vasil LK, Bogorad L. *Cell Culture and Somatic Cell Genetics in Plants, the Molecular Biology of Plastid 7A*. Academic Press, San Diego; 1991. pp. 5–53.
6. Jansen RK, Cai Z, Raubeson LA, Daniell H, dePamphilis CW, Leebens-Mack J, et al. Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns. *Proc Natl Acad Sci. U.S.A.* 2007; 104, 19369–19374. doi: [10.1073/pnas.0709121104](https://doi.org/10.1073/pnas.0709121104) PMID: [18048330](https://pubmed.ncbi.nlm.nih.gov/18048330/)
7. Shaw J, Shafer HL, Leonard OR, Kovach MJ, Schorr M, Morris AB. Chloroplast DNA sequence utility for the lowest phylogenetic and phylogeographic inferences in angiosperms: the tortoise and the hare IV. *American Journal of Botany.* 2014; 101: 1987–2004. doi: [10.3732/ajb.1400398](https://doi.org/10.3732/ajb.1400398) PMID: [25366863](https://pubmed.ncbi.nlm.nih.gov/25366863/)
8. Burke SV, Grennan CP, Duvall MR. Plastome sequences of two New World bamboos—*Arundinaria gigantea* and *Cryptochloa strictiflora* (Poaceae)—extend phylogenomic understanding of Bambusoideae. *American Journal of Botany.* 2012; 99(12): 1951–1961. doi: [10.3732/ajb.1200365](https://doi.org/10.3732/ajb.1200365) PMID: [23221496](https://pubmed.ncbi.nlm.nih.gov/23221496/)
9. Huang H, Shi C, Liu Y, Mao SY, Gao LZ. Thirteen *Camellia* chloroplast genome sequences determined by high-throughput sequencing: genome structure and phylogenetic relationships. *BMC Evol Biol.* 2014; 14: 151. doi: [10.1186/1471-2148-14-151](https://doi.org/10.1186/1471-2148-14-151) PMID: [25001059](https://pubmed.ncbi.nlm.nih.gov/25001059/)
10. Walker JF, Zanis MJ, Emery NC. Comparative analysis of complete chloroplast genome sequence and inversion variation in *Lasthenia burkei* (Madieae, Asteraceae). *American Journal of Botany.* 2014; 101(4), 722–729. doi: [10.3732/ajb.1400049](https://doi.org/10.3732/ajb.1400049) PMID: [24699541](https://pubmed.ncbi.nlm.nih.gov/24699541/)
11. Wolfe KH, Li WH, Sharp PM. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *Proceedings of the National Academy of Sciences.* 1987; 84(24): 9054–9058.
12. Raubeson LA, Jansen RK. Chloroplast genomes of plants. In: Henry RJ. *Plant Diversity and Evolution: Genotypic and Phenotypic Variation in Higher Plants*. CABI, Wallingford, UK; 2005. pp. 45–68.
13. Drouin G, Daoud H, Xia J. Relative rates of synonymous substitutions in the mitochondrial, chloroplast and nuclear genomes of seed plants. *Molecular Phylogenetics and Evolution.* 2008; 49(3): 827–831. doi: [10.1016/j.ympev.2008.09.009](https://doi.org/10.1016/j.ympev.2008.09.009) PMID: [18838124](https://pubmed.ncbi.nlm.nih.gov/18838124/)
14. Bell CD, Soltis DE, Soltis PS. The age and diversification of the angiosperms re-revisited. *American Journal of Botany* 2010; 97(8): 1296–1303. doi: [10.3732/ajb.0900346](https://doi.org/10.3732/ajb.0900346) PMID: [21616882](https://pubmed.ncbi.nlm.nih.gov/21616882/)
15. Cosner ME, Raubeson LA, Jansen RK. Chloroplast DNA rearrangements in Campanulaceae: phylogenetic utility of highly rearranged genomes. *BMC Evol Biol.* 2004; 4(1): 1.
16. Lee HL, Jansen RK, Chumley TW, Kim KJ. Gene relocations within chloroplast genomes of *Jasminum* and *Menodora* (Oleaceae) are due to multiple, overlapping inversions. *Mol Biol Evol.* 2007; 24: 1161–1180. doi: [10.1093/molbev/msm036](https://doi.org/10.1093/molbev/msm036) PMID: [17329229](https://pubmed.ncbi.nlm.nih.gov/17329229/)
17. Cai Z, Guisinger M, Kim HG, Ruck E, Blazier JC, McMurtry V, et al. Extensive reorganization of the plastid genome of *Trifolium subterraneum* (Fabaceae) is associated with numerous repeated sequences and novel DNA insertions. *Journal of Molecular Evolution.* 2008; 67(6): 696–704. doi: [10.1007/s00239-008-9180-7](https://doi.org/10.1007/s00239-008-9180-7) PMID: [19018585](https://pubmed.ncbi.nlm.nih.gov/19018585/)
18. Guisinger MM, Chumley TW, Kuehl JV, Boore JL, Jansen RK. Implications of the plastid genome sequence of *Typha* (Typhaceae, Poales) for understanding genome evolution in Poaceae. *Journal of Molecular Evolution.* 2010; 70(2): 149–166. doi: [10.1007/s00239-009-9317-3](https://doi.org/10.1007/s00239-009-9317-3) PMID: [20091301](https://pubmed.ncbi.nlm.nih.gov/20091301/)
19. Guisinger MM, Kuehl JV, Boore JL, Jansen RK. Extreme reconfiguration of plastid genomes in the angiosperm family Geraniaceae: rearrangements, repeats, and codon usage. *Molecular Biology and Evolution.* 2011; 28(1): 583–600. doi: [10.1093/molbev/msq229](https://doi.org/10.1093/molbev/msq229) PMID: [20805190](https://pubmed.ncbi.nlm.nih.gov/20805190/)
20. Martin GE, Rousseau-Gueutin M, Cordonnier S, Lima O, Michon-Coudouel S, Naquin D, et al. The first complete chloroplast genome of the Genistoid legume *Lupinus luteus*: evidence for a novel major lineage-specific rearrangement and new insights regarding plastome evolution in the legume family. *Annals of Botany.* 2014; 113(7): 1197–1210. doi: [10.1093/aob/mcu050](https://doi.org/10.1093/aob/mcu050) PMID: [24769537](https://pubmed.ncbi.nlm.nih.gov/24769537/)

21. Lu D, Wu Z, Zhou L, Chen S, Gilbert MG. Caryophyllaceae. In: Wu ZY, Raven RH, Hong DY (eds). www.eFloras.org, Flora of China, vol. 6. Accessed 2016 August 20th.
22. Greuter W. *Silene* (Caryophyllaceae) in Greece: a subgeneric and sectional classification. *Taxon*. 1995; 44(4): 543–581.
23. Lidén M, Popp M, Oxelman B. A revised generic classification of the tribe *Sileneae* (Caryophyllaceae). *Nordic Journal of Botany*. 2000; 20(5): 513–518.
24. Greenberg AK, Donoghue MJ. Molecular systematics and character evolution in Caryophyllaceae. *Taxon*. 2011; 60(6): 1637–1652.
25. Sloan DB, Alverson AJ, Wu M, Palmer JD, Taylor DR. Recent acceleration of plastid sequence and structural evolution coincides with extreme mitochondrial divergence in the angiosperm genus *Silene*. *Genome Biol Evol*. 2012; 4(3): 294–306. doi: [10.1093/gbe/evs006](https://doi.org/10.1093/gbe/evs006) PMID: [22247429](https://pubmed.ncbi.nlm.nih.gov/22247429/)
26. Sloan DB, Triant DA, Forrester NJ, Bergner LM, Wu M, Taylor DR. A recurring syndrome of accelerated plastid genome evolution in the angiosperm tribe *Sileneae* (Caryophyllaceae). *Molecular Phylogenetics and Evolution*. 2014; 72: 82–89. doi: [10.1016/j.ympev.2013.12.004](https://doi.org/10.1016/j.ympev.2013.12.004) PMID: [24373909](https://pubmed.ncbi.nlm.nih.gov/24373909/)
27. Kang Y, Lee H, Kim MK, Shin SC, Park H, Lee J. The complete chloroplast genome of Antarctic pearlwort, *Colobanthus quitensis* (Kunth) Bartl. (Caryophyllaceae). *Mitochondrial DNA*. 2015.
28. Gurusamy R, Lee DH, Park SJ. The complete chloroplast genome sequence of *Dianthus superbus* var. *longicalycinus*. *Mitochondrial DNA Part A*. 2016; 27(3): 2015–2017.
29. Wyman SK, Jansen RK, Boore JL. Automatic annotation of organellar genomes with DOGMA. *Bioinformatics*. 2004; 20(17): 3252–3255. doi: [10.1093/bioinformatics/bth352](https://doi.org/10.1093/bioinformatics/bth352) PMID: [15180927](https://pubmed.ncbi.nlm.nih.gov/15180927/)
30. Schattner P, Brooks AN, Lowe TM. The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. *Nucleic Acids Research*. 2005; 33, W686–W689. doi: [10.1093/nar/gki366](https://doi.org/10.1093/nar/gki366) PMID: [15980563](https://pubmed.ncbi.nlm.nih.gov/15980563/)
31. Lohse M, Drechsel O, Kahlau S, Bock R. OrganellarGenomeDRAW—a suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets. *Nucleic Acids Res*. 2013.
32. Frazer KA, Pachter L, Poliakov A, Rubin EM, Dubchak I. VISTA: computational tools for comparative genomics. *Nucleic Acids Res*. 2004; 32, W273–W279. doi: [10.1093/nar/gkh458](https://doi.org/10.1093/nar/gkh458) PMID: [15215394](https://pubmed.ncbi.nlm.nih.gov/15215394/)
33. Mayer C. Phobos Version 3.3.12. A tandem repeat search program. 20 p. 2010; Available:http://www.rub.de/spezzoo/cm/cm_phobos.htm. Accessed 2016 September 30th.
34. Kurtz S, Choudhuri JV, Ohlebusch E, Schleiermacher C, Stoye J, Giegerich R. REPuter: the manifold applications of repeat analysis on a genome scale. *Nucleic Acids Res*. 2001; 29(22): 4633–4642. PMID: [11713313](https://pubmed.ncbi.nlm.nih.gov/11713313/)
35. Swofford DL. PAUP*. Phylogenetic analysis using parsimony (*and other methods). Version 4. Sunderland, MA: Sinauer Associates. 2003.
36. Katoh K, Misawa K, Kuma KI, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res*. 2002; 30: 3059–3066. PMID: [12136088](https://pubmed.ncbi.nlm.nih.gov/12136088/)
37. Darriba D, Taboada GL, Doallo R, Posada D. jModelTest 2: more models, new heuristics and parallel computing. *Nat Meth*. 2012; 9: 772.
38. Stamatakis A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics*. 2006; 22(21): 2688–2690. doi: [10.1093/bioinformatics/btl446](https://doi.org/10.1093/bioinformatics/btl446) PMID: [16928733](https://pubmed.ncbi.nlm.nih.gov/16928733/)
39. Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, et al. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology*. 2012; 61(3): 539–542. doi: [10.1093/sysbio/sys029](https://doi.org/10.1093/sysbio/sys029) PMID: [22357727](https://pubmed.ncbi.nlm.nih.gov/22357727/)
40. Sato S, Nakamura Y, Kaneko T, Asamizu E, Tabata S. Complete structure of the chloroplast genome of *Arabidopsis thaliana*. *DNA Research*. 1999; 6(5): 283–290. PMID: [10574454](https://pubmed.ncbi.nlm.nih.gov/10574454/)
41. Hupfer H, Swiatek M, Hornung S, Herrmann RG, Maier RM, Chiu WL, et al. Complete nucleotide sequence of the *Oenothera elata* plastid chromosome, representing plastome I of the five distinguishable *Oenothera* plastomes. *Molecular and General Genetics MGG*. 2000; 263(4): 581–585. PMID: [10852478](https://pubmed.ncbi.nlm.nih.gov/10852478/)
42. Kato T, Kaneko T, Sato S, Nakamura Y, Tabata S. Complete structure of the chloroplast genome of a legume, *Lotus japonicas*. *DNA Res*. 2000; 7(6): 323–330. PMID: [11214967](https://pubmed.ncbi.nlm.nih.gov/11214967/)
43. Ibrahim RIH, Azuma JI, Sakamoto M. Complete nucleotide sequence of the cotton (*Gossypium barbadense* L.) chloroplast genome with a comparative analysis of sequences among 9 dicot plants. *Genes & Genetic Systems*. 2006; 81(5): 311–321.

44. Do HDK, Kim JS, Kim JH. Comparative genomics of four Liliales families inferred from the complete chloroplast genome sequence of *Veratrum patulum* O. Loes. (Melanthiaceae). *Gene*. 2013; 530(2): 229–235. doi: [10.1016/j.gene.2013.07.100](https://doi.org/10.1016/j.gene.2013.07.100) PMID: [23973725](https://pubmed.ncbi.nlm.nih.gov/23973725/)
45. Harris ME, Meyer G, Vandergon T, Vandergon VO. Loss of the acetyl-CoA carboxylase (*accD*) gene in Poales. *Plant Molecular Biology Reporter*. 2013; 31(1): 21–31.
46. Li X, Zhang TC, Qiao Q, Ren Z, Zhao J, Yonezawa T, et al. Complete chloroplast genome sequence of holoparasite *Cistanche deserticola* (Orobanchaceae) reveals gene loss and horizontal gene transfer from its host *Haloxylon ammodendron* (Chenopodiaceae). *PLoS One* (2013; 8(3): e58747. doi: [10.1371/journal.pone.0058747](https://doi.org/10.1371/journal.pone.0058747) PMID: [23554920](https://pubmed.ncbi.nlm.nih.gov/23554920/)
47. Martínez-Alberola F, del Campo EM, Lázaro-Gimeno D, Mezquita-Claramonte S, Molins A, Mateu-Andrés I, et al. Balanced gene losses, duplications and intensive rearrangements led to an unusual regularly sized genome in *Arbutus unedo* chloroplasts. *PLoS One*. 2013; 8(11): e79685. doi: [10.1371/journal.pone.0079685](https://doi.org/10.1371/journal.pone.0079685) PMID: [24260278](https://pubmed.ncbi.nlm.nih.gov/24260278/)
48. Liu TJ, Zhang CY, Yan HF, Zhang L, Ge XJ, Hao G. Complete plastid genome sequence of *Primula sinensis* (Primulaceae): structure comparison, sequence variation and evidence for *accD* transfer to nucleus. *PeerJ*. 2016; 4: e2101. doi: [10.7717/peerj.2101](https://doi.org/10.7717/peerj.2101) PMID: [27375965](https://pubmed.ncbi.nlm.nih.gov/27375965/)
49. Erixon P, Oxelman B. Reticulate or tree-like chloroplast DNA evolution in *Sileneae* (Caryophyllaceae)? *Molecular Phylogenetics and Evolution*. 2008; 48: 313–325. doi: [10.1016/j.ympev.2008.04.015](https://doi.org/10.1016/j.ympev.2008.04.015) PMID: [18490181](https://pubmed.ncbi.nlm.nih.gov/18490181/)
50. Ravi V, Khurana JP, Tyagi AK, Khurana P. An update on chloroplast genomes. *Plant Systematics and Evolution*. 2008; 271: 101–122.
51. Ni L, Zhao Z, Xu H, Chen S, Dorje G. The complete chloroplast genome of *Gentiana straminea* (Gentianeaceae), an endemic species to the Sino-Himalayan subregion. *Gene*. 2016; 577(2): 281–288. doi: [10.1016/j.gene.2015.12.005](https://doi.org/10.1016/j.gene.2015.12.005) PMID: [26680100](https://pubmed.ncbi.nlm.nih.gov/26680100/)
52. Logacheva MD, Penin AA, Valiejo-Roman CM, Antonov AS. Structure and evolution of junctions between inverted repeat and small single copy regions of chloroplast genome in non-core Caryophyllales. *Molecular Biology*. 2009; 43: 757–765.
53. Sun QB, Li LF, Li Y, Wu GJ, Ge XJ. SSR and AFLP markers reveal low genetic diversity in the biofuel plant in China. *Crop Science*. 2008; 48(5): 1865–1871.
54. Kuang DY, Wu H, Wang YL, Gao LM, Zhang SZ, Lu L. Complete chloroplast genome sequence of *Magnolia kwangsiensis* (Magnoliaceae): implications for DNA barcoding and population genetics. *Genome*. (2011; 54(8): 663–673. doi: [10.1139/G11-026](https://doi.org/10.1139/G11-026) PMID: [21793699](https://pubmed.ncbi.nlm.nih.gov/21793699/)
55. Qian J, Song J, Gao H, Zhu Y, Xu J, Pang X, et al. The complete chloroplast genome sequence of the medicinal plant *Salvia miltiorrhiza*. *PLoS One*. 2013; 8(2): e57607. doi: [10.1371/journal.pone.0057607](https://doi.org/10.1371/journal.pone.0057607) PMID: [23460883](https://pubmed.ncbi.nlm.nih.gov/23460883/)
56. Kim KJ, Choi KS, Jansen RK. Two chloroplast DNA inversions originated simultaneously during the early evolution of the sunflower family (Asteraceae). *Molecular Biology and Evolution*. 2005; 22(9): 1783–1792. doi: [10.1093/molbev/msi174](https://doi.org/10.1093/molbev/msi174) PMID: [15917497](https://pubmed.ncbi.nlm.nih.gov/15917497/)
57. Tsumura Y, Suyama Y, Yoshimura K. Chloroplast DNA inversion polymorphism in populations of *Abies* and *Tsuga*. *Molecular Biology and Evolution*. 2000; 17(9): 1302–1312. PMID: [10958847](https://pubmed.ncbi.nlm.nih.gov/10958847/)
58. Ogihara Y, Terachi T, Sasakuma T. Intramolecular recombination of chloroplast genome mediated by short direct-repeat sequences in wheat species. *Proceedings of the National Academy of Sciences*. 1988; 85(22): 8573–8577.
59. Knox EB, Downie SR, Palmer JD. Chloroplast genome rearrangements and the evolution of giant lobelias from herbaceous ancestors. *Molecular Biology and Evolution*. 1993; 10(2): 414–430.
60. Fior S, Karis PO, Casazza G, Minuto L, Sala F. Molecular phylogeny of the Caryophyllaceae (Caryophyllales) inferred from chloroplast *matK* and nuclear rDNA ITS sequences. *American Journal of Botany*. 2006; 93(3): 399–411. doi: [10.3732/ajb.93.3.399](https://doi.org/10.3732/ajb.93.3.399) PMID: [21646200](https://pubmed.ncbi.nlm.nih.gov/21646200/)