

**Article title: Structural variation and evolutionary dynamics of plastid genomes in the subgenus *Lychnis* (*Silene*, Caryophyllaceae): insights from genome rearrangements and substitution rates**

**Authors:** Seongjun Park, Ogyeong Son

## **Supplementary Material**

**Figure S1.** Plastome maps for six *Silene* species within the subgenus *Lychnis*.

**Figure S2.** Characteristics of complete *Silene* plastomes.

**Figure S3.** Nucleotide and amino acid sequence alignments of the plastid-encoded *accD* gene among *Silene* species within the subgenus *Lychnis*, including outgroup *Agrostemma*.

**Figure S4.** Summary of plastid gene and intron content across sampled Caryophyllaceae species.

**Figure S5.** Sequence alignment and functional domain analysis of nuclear-encoded *INFA* and *RPL23* ORFs.

**Figure S6.** Nuclear gene structures and coding-sequence alignments of *INFA* and *RPL23* in *Silene conica* and *S. latifolia*.

**Figure S7.** Alignments of IR<sub>B</sub>/SSC boundary regions in *Silene* subgenus *Lychnis* and the outgroup *Agrostemma*.

**Figure S8.** Sequence divergence in plastid-encoded gene groups among selected Caryophyllaceae, showing the correlation between  $d_N$  and  $d_S$ .

**Figure S9.** Amino acid sequence alignment of plastid-encoded *accD* across Caryophyllaceae, highlighting the AAR-containing region.

**Table S1.** Repeat characteristics of *Silene* species within the subgenus *Lychnis*, including an outgroup.

**Table S2.** Predicted transit peptide scores for putative nuclear-encoded plastid-targeted genes (*INFA* and *RPL23*) using TargetP 2.0.

**Table S3.** Genomic locations and structures of nuclear *INFA* and *RPL23* homologs in *Silene conica* and *S. latifolia*.

**Table S4.** Information on collection sources and voucher specimens

**Table S5.** NCBI Sequence Read Archive (SRA) accession numbers for raw sequencing data used in this study.

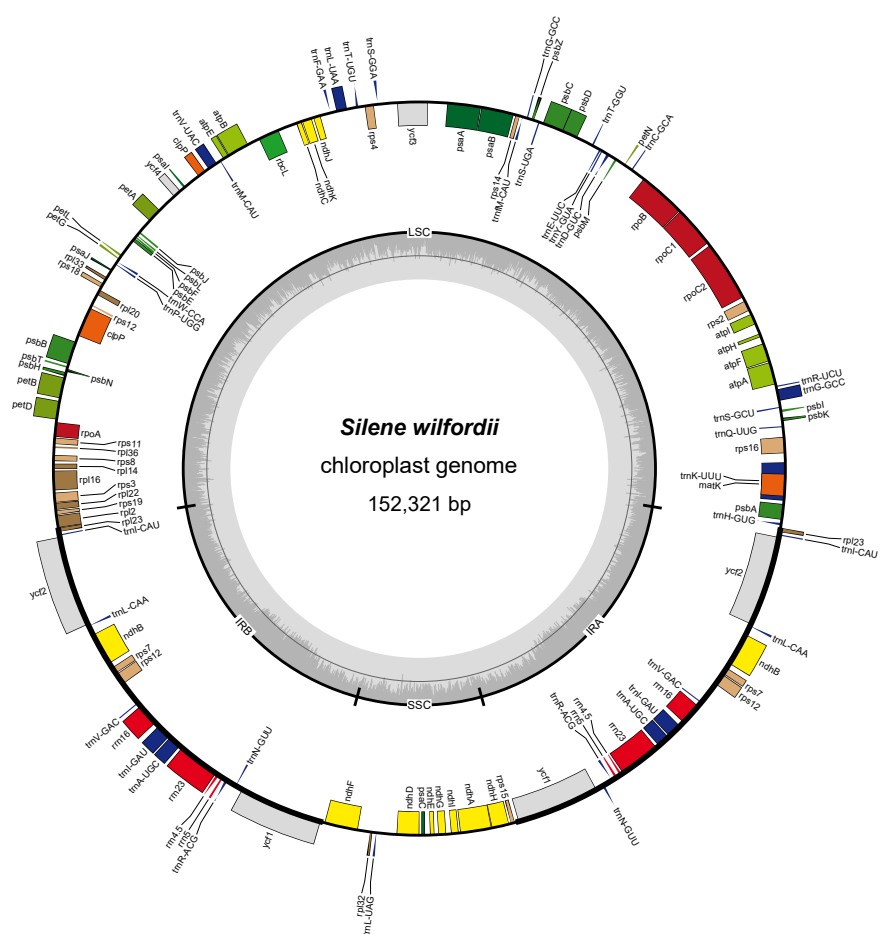


Figure S1. (continued)

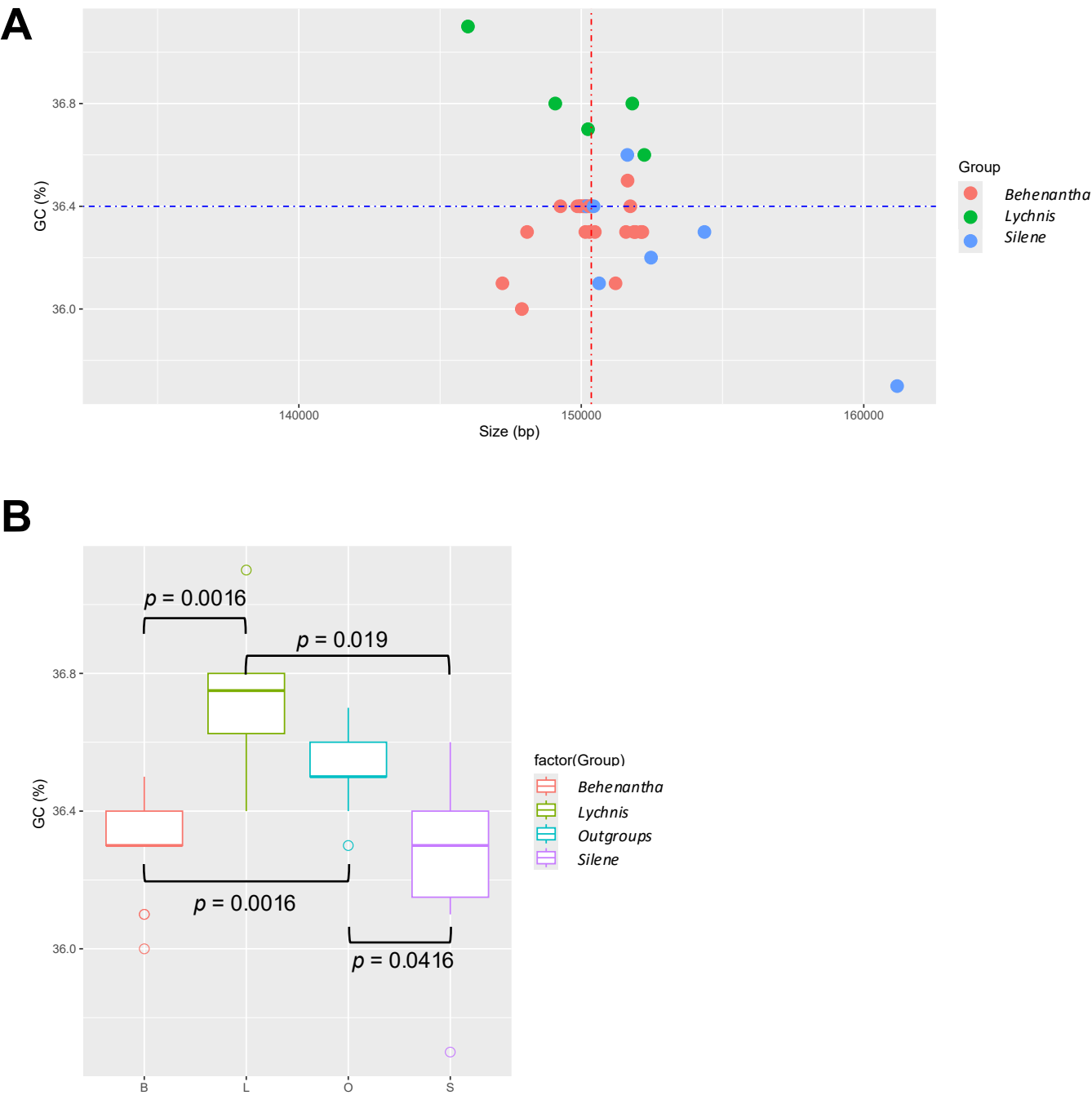
C



# E



**Figure S2. Characteristics of complete *Silene* plastomes.** (A) Relationship between plastome size and GC content (%) across *Silene* species and the outgroup. Points are colour-coded by *Silene* subgenus; outgroup taxa are shown in grey. (B) Boxplot summarizing the distribution of total plastome sizes. Boxes indicate the interquartile range, horizontal lines within boxes denote the median, whiskers extend to the minimum and maximum non-outlier values, and circles represent outliers. Differences among groups were assessed using pairwise Wilcoxon rank-sum tests implemented in R.



**Figure S3. Nucleotide and amino acid sequence alignments of the plastid-encoded *accD* gene among *Silene* species within the subgenus *Lychnis*, including outgroup *Agrostemma*.** The alignments highlight regions of pronounced structural variation (Indels) which contribute to *accD* instability in *Lychnis*. Colored rectangles indicate specific insertion (blue and green) and deletion (red) events. A large insertion (bracketed) is shared by *S. cognata*, *S. fulgens*, *S. kiusiana* and *S. chalcidonica*, and is hypothesized to represent an amino acid repeat (AAR; green) expansion that may precede gene loss in this lineage. The 109-bp N-terminal expansion specific to *S. cognata* and *S. fulgens* is also indicated by an arrow.



**Figure S4. Summary of plastid gene and intron content across sampled Caryophyllaceae species.** Presence, absence, and pseudogenization status of plastid (A) protein-coding genes, (B) tRNA genes, and (C) introns across the 39 sampled Caryophyllaceae taxa. Green boxes indicate the presence of an intact gene or intron, white boxes indicate complete loss (absence), and the symbol  $\psi$  denotes pseudogenized (non-functional) copies.

**A** Protein-coding genes

[illegible]

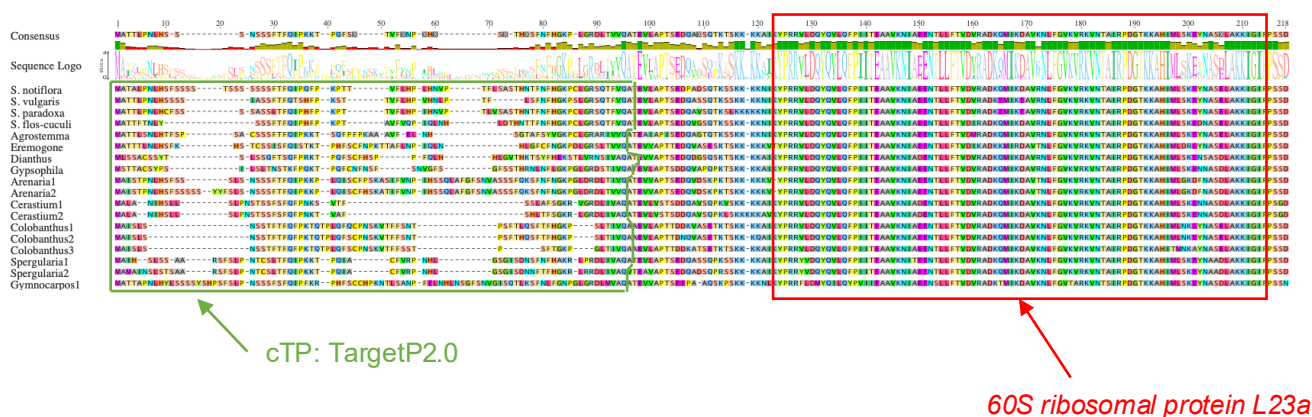
**B** tRNA genes

[illegible]

**C** introns

[illegible]

# A



**Figure S6. Nuclear gene structures and coding-sequence alignments of *INFA* and *RPL23* in *Silene conica* and *S. latifolia*.** Schematic diagrams depict exon–intron organization and coding sequence (CDS) boundaries for each nuclear locus. Pairwise nucleotide alignments are shown below the gene models to visualize sequence similarity across species. For *INFA*, alignments were generated using the full nucleotide sequence of the gene model. For *RPL23*, alignments were performed using CDS-only sequences (concatenated exons) to enable direct comparison of coding regions independent of intron length and structure.

*INFA*



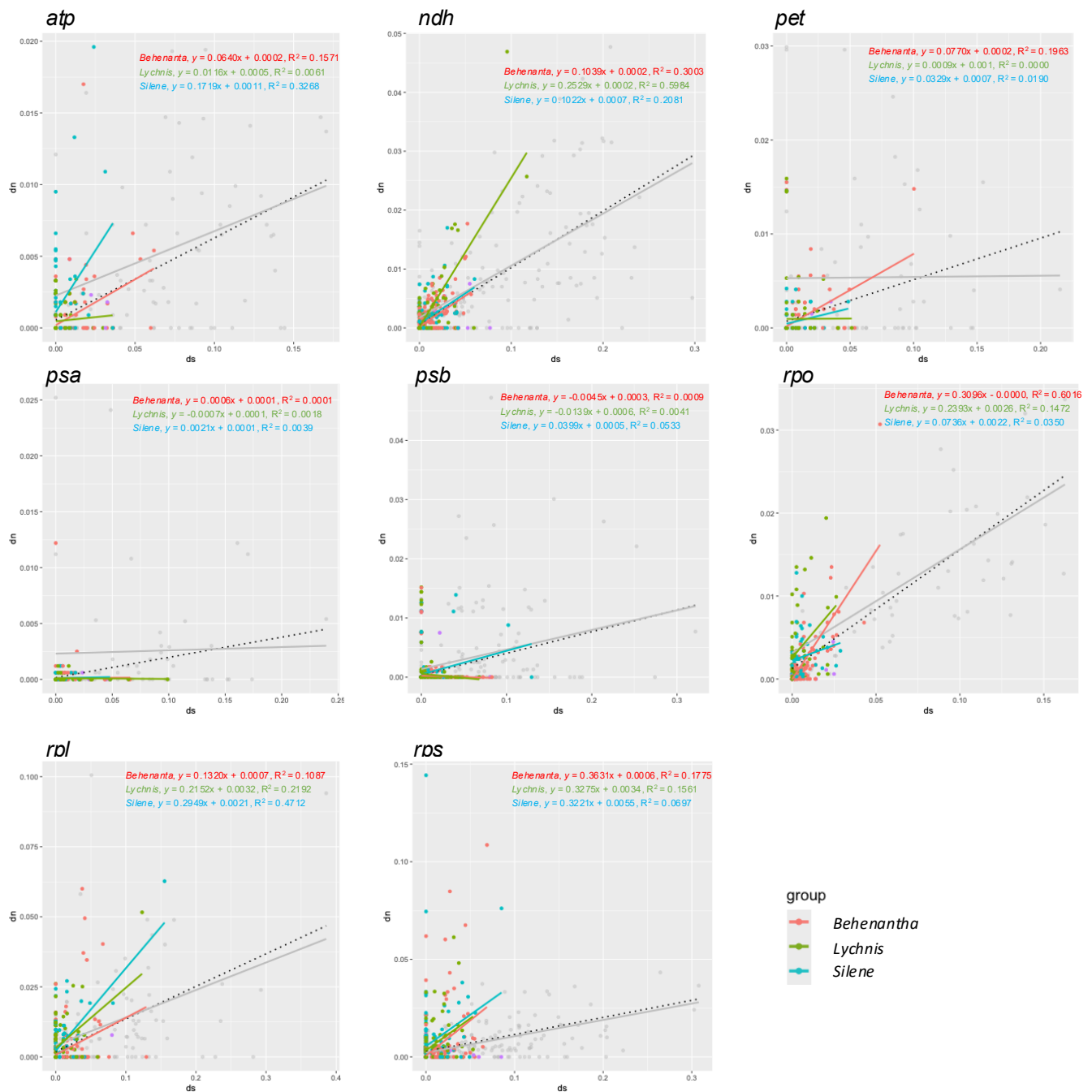
*RPL23*



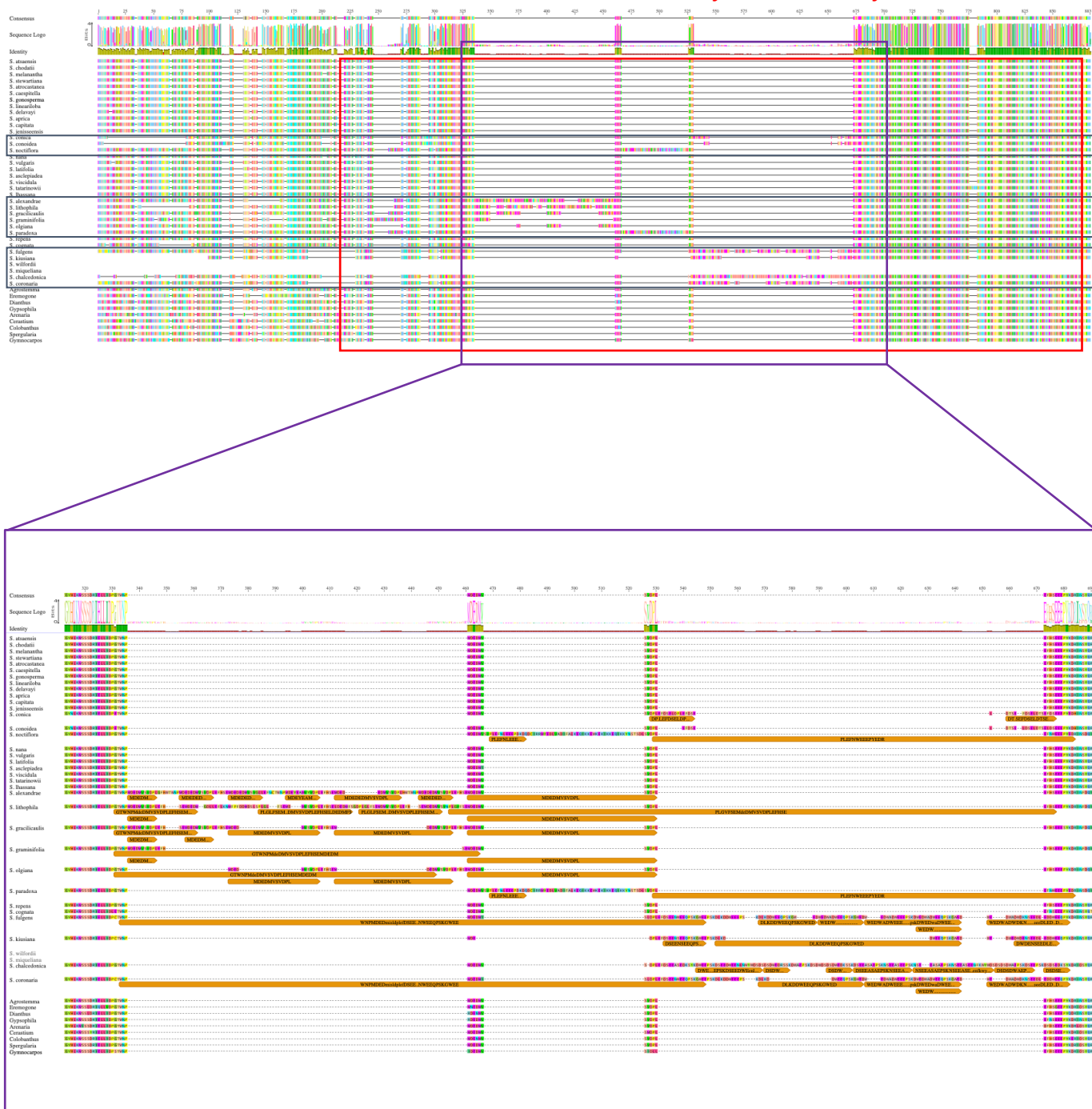
**Figure S7. Alignments of IR<sub>B</sub>/SSC boundary regions in *Silene* subgenus *Lychnis* and the outgroup *Agrostemma*.** Comparative alignment of genomic regions spanning the IRB/SSC junctions for seven *Silene* species in subgenus *Lychnis* and the outgroup *Agrostemma*. The full alignment is shown above, with five representative IRB/SSC boundary segments enlarged below to highlight shifts in the junction position and local sequence context among species.



**Figure S8. Sequence divergence in plastid-encoded gene groups among selected Caryophyllaceae, showing the correlation between  $d_N$  and  $d_S$ .** Scatterplots showing the relationship between nonsynonymous ( $d_N$ ) and synonymous ( $d_S$ ) substitution rates for plastid-encoded gene groups across selected Caryophyllaceae taxa. Points are colour-coded by *Silene* subgenus, with outgroup taxa shown in grey. For each subgenus, a least-squares regression line is fitted, and the corresponding slope and coefficient of determination ( $R^2$ ) are reported in the panels to summarize differences in the  $d_N$ – $d_S$  relationship among groups.



*acetyl-CoA carboxylase beta subunit*



**Table S1. Repeat characteristics of *Silene* species within the subgenus *Lychnis*, including an outgroup.**

| Species                | Repeat pairs | Repeat length | Repeat portion |
|------------------------|--------------|---------------|----------------|
| <i>S. cognata</i>      | 33           | 8,278         | 5.55           |
| <i>S. fulgens</i>      | 32           | 11,194        | 7.45           |
| <i>S. kiusiana</i>     | 47           | 11,364        | 7.49           |
| <i>S. wilfordii</i>    | 26           | 4,788         | 3.14           |
| <i>S. miqueliana</i>   | 23           | 4,789         | 3.28           |
| <i>S. coronaria</i>    | 6            | 622           | 0.41           |
| <i>S. chalcedonica</i> | 28           | 5,089         | 3.44           |
| <i>Agrostemma</i>      | 6            | 554           | 0.37           |

**Table S2. Predicted transit peptide scores for putative nuclear-encoded plastid-targeted genes (*INFA* and *RPL23*) using TargetP 2.0.**

| Species                        | <i>INFA</i> |                                         |             |                                | <i>RPL23</i> |                                         |             |                                |
|--------------------------------|-------------|-----------------------------------------|-------------|--------------------------------|--------------|-----------------------------------------|-------------|--------------------------------|
|                                | len (aa)    | Chloroplast transfer peptide likelihood | Probability | Potential cleavage length (aa) | len (aa)     | Chloroplast transfer peptide likelihood | Probability | Potential cleavage length (aa) |
| <i>Silene noctiflora</i>       | 141         | 0.3761                                  | -           | -                              | 195          | 0.9869                                  | 0.5173      | 74                             |
| <i>Silene vulgaris</i>         | 131         | 0.5919                                  | 0.2323      | 47                             | 183          | 0.9987                                  | 0.6115      | 62                             |
| <i>Silene paradoxa</i>         | 135         | 0.4292                                  | -           | -                              | 191          | 0.9982                                  | 0.5067      | 69                             |
| <i>Silene flos-cuculi L.</i>   | 137         | 0.5270                                  | 0.2434      | 53                             | 181          | 0.9978                                  | 0.5192      | 60                             |
| <i>Agrostemma githago</i>      | 134         | 0.4522                                  | -           | -                              | 187          | 0.9997                                  | 0.6749      | 65                             |
| <i>Eremogone hookeri</i>       | 140         | 0.4999                                  | 0.1255      | 93                             | 187          | 0.9993                                  | 0.6177      | 67                             |
| <i>Dianthus superbus</i>       | 136         | 0.4921                                  | -           | -                              | 185          | 0.9985                                  | 0.5878      | 65                             |
| <i>Gypsophila repens</i>       | 137         | 0.615                                   | 0.1409      | 51                             | 186          | 0.9984                                  | 0.4634      | 65                             |
| <i>Arenaria serpyllifolia</i>  | 139         | 0.6105                                  | 0.1003      | 57                             | 204          | 0.9977                                  | 0.5626      | 83                             |
|                                | 134         | 0.6866                                  | 0.1887      | 52                             | 210          | 0.9952                                  | 0.5606      | 89                             |
| <i>Cerastium alpinum</i>       | 132         | 0.5551                                  | 0.1739      | 49                             | 170          | 0.9988                                  | 0.6233      | 48                             |
|                                | 132         | 0.4639                                  | -           | -                              | 171          | 0.9988                                  | 0.6063      | 48                             |
| <i>Colobanthus quitensis</i>   | 138         | 0.6992                                  | 0.1194      | 52                             | 179          | 0.9994                                  | 0.6048      | 58                             |
|                                | 147         | 0.3601                                  | -           | -                              | 179          | 0.9994                                  | 0.6064      | 58                             |
|                                |             |                                         |             |                                | 172          | 0.9994                                  | 0.6485      | 51                             |
| <i>Spergularia marina</i>      | 137         | 0.8307                                  | 0.1643      | 55                             | 187          | 0.996                                   | 0.7152      | 65                             |
|                                | 138         | 0.8129                                  | 0.1884      | 56                             | 190          | 0.9987                                  | 0.6834      | 68                             |
| <i>Gymnocarpus przewalskii</i> | 144         | 0.8226                                  | 0.5788      | 62                             | 212          | 0.9884                                  | 0.6183      | 91                             |
|                                | 138         | 0.544                                   | 0.4323      | 56                             |              |                                         |             |                                |

**len (aa):** The length of the predicted ORF in amino acids. **Chloroplast transfer peptide likelihood:** The likelihood score for the presence of a cTP, as determined by TargetP 2.0. **Probability:** The probability score associated with the cTP prediction, indicating the confidence level. **Potential cleavage length (aa):** The predicted length (number of amino acids) of the cTP before the predicted cleavage site.

**Table S3. Genomic locations and structures of nuclear *INFA* and *RPL23* homologs in *Silene conica* and *S. latifolia*.** For each species, the best-supported nuclear loci were localized to the indicated chromosome/scaffold IDs, and their genomic coordinates and strand orientation are reported. Gene length denotes the span from the annotated start to end coordinates (including introns), whereas CDS length represents the summed coding sequence length. Exons/introns indicates the inferred gene structure based on the genomic locus (e.g., 3/2 = three exons and two introns).

| Species              | <i>Silene conica</i>  |                       | <i>Silene latifolia</i> |                       |
|----------------------|-----------------------|-----------------------|-------------------------|-----------------------|
| NCBI RefSeq assembly | GCA_029255685.2       |                       | GCF_048544455.1         |                       |
| Gene                 | <i>INFA</i>           | <i>RPL23</i>          | <i>INFA</i>             | <i>RPL23</i>          |
| Chromosome ID        | Chr7                  | Chr8                  | Chr4                    | Chr2                  |
|                      | CM055317.2            | CM055318.2            | NC_133529.1             | NC_133527.1           |
| Genomic coordinates  | 64152139-<br>64152546 | 28633499-<br>28635816 | 15918312-<br>15918719   | 49644045-<br>49648029 |
| Strand               | +                     | +                     | +                       | +                     |
| Gene length (bp)     | 408                   | 2,318                 | 408                     | 3,985                 |
| CDS length (bp)      | 408                   | 609                   | 408                     | 573                   |
| Exons/introns        | 1/0                   | 3/2                   | 1/0                     | 3/2                   |

**Table S4. Information on collection sources and voucher specimens.**

| Species              | collection sources                                                                   | Voucher specimen<br>[Herbarium code] | Identified by     | NCBI<br>Accession number |
|----------------------|--------------------------------------------------------------------------------------|--------------------------------------|-------------------|--------------------------|
| <i>S. cognata</i>    | Mt. Bohyeon, Hwabuk-myeon, Yeongcheon-si, Gyeongsangbuk-do, Republic of Korea [Wild] | <i>SPark 2023</i><br>[YNUH]          | Dr. Seongjun Park | PQ533854                 |
| <i>S. fulgens</i>    | Korea National Arboretum, Pocheon-si, Kyonggi-do, Republic of Korea [Cultivated]     | -                                    | Dr. Sungwon Son   | PQ533856                 |
| <i>S. kiusiana</i>   | Korea National Arboretum, Pocheon-si, Kyonggi-do, Republic of Korea [Cultivated]     | -                                    | Dr. Sungwon Son   | PQ533857                 |
| <i>S. wilfordii</i>  | Korea National Arboretum, Pocheon-si, Kyonggi-do, Republic of Korea [Cultivated]     | -                                    | Dr. Sungwon Son   | PQ533859                 |
| <i>S. miqueliana</i> | Shirokanedai, Minato City, Tokyo, Japan [Wild]                                       | <i>SPark 2018</i><br>[YNUH]          | Dr. Seongjun Park | PQ533858                 |
| <i>S. coronaria</i>  | Sowon-myeon, Taeon-gun, Chungcheongnam-do, Republic of Korea [Cultivated]            | <i>SPark s.n.</i><br>[YNUH]          | Dr. Seongjun Park | PQ533855                 |

*Silene fulgens*, *S. kiusiana*, and *S. wilfordii* leaf materials were obtained from living collections maintained at Korea National Arboretum (Pocheon-si, Gyeonggi-do, Republic of Korea). Voucher specimens were not prepared since these plants are actively cultivated and maintained as part of the Arboretum’s botanical garden collection.

**Table S5. NCBI Sequence Read Archive (SRA) accession numbers for raw sequencing data used in this study.**

| Species                        | NCBI Accession number |
|--------------------------------|-----------------------|
| <i>Agrostemma githago</i>      | SRR999316             |
| <i>Arenaria serpyllifolia</i>  | SRR6787511            |
| <i>Cerastium alpinum</i>       | SRR6435353            |
| <i>Colobanthus quitensis</i>   | SRR5678326            |
| <i>Dianthus superbus</i>       | SRR342032             |
| <i>Eremogone hookeri</i>       | SRR6787478            |
| <i>Gymnocarpos przewalskii</i> | SRR5340021            |
| <i>Gypsophila repens</i>       | SRR6787498            |
| <i>Silene flos-cuculi</i> L.   | SRR29469937           |
| <i>Silene noctiflora</i>       | SRR29469929           |
| <i>Silene paradoxa</i>         | SRR999299             |
| <i>Silene vulgaris</i>         | SRR2106278            |
| <i>Spergularia marina</i>      | SRR6435300            |