

RESEARCH

Open Access



Substantial structural variation and repetitive DNA content contribute to intraspecific plastid genome evolution

Alfredo López-Caamal^{1*}, Tyler Gandee², Laura F. Galloway¹ and Karen B. Barnard-Kubow²

Abstract

Background Plastids have highly conserved genomes in most land plants. However, in several families, plastid genomes exhibit high rates of nucleotide substitution and structural rearrangements among species. This elevated rate of evolution has been posited to lead to increased rates of plastid-nuclear incompatibilities (PNI), potentially acting as a driver of speciation. However, the extent to which plastid structural variation exists within a species is unknown. This study investigates whether plastid structural variation, observed at the interspecific level in Campanulaceae, also occurs within *Campanula americana*, a species with strong intraspecific PNI. We assembled multiple plastid genomes from three lineages of *C. americana* that exhibit varying levels of PNI when crossed. We then investigated the structural variation and repetitive DNA content among these lineages and compared the repetitive DNA content with that of other species within the family.

Results We found significant variation in plastid genome size among the lineages of *C. americana* (188,309–201,788 bp). This variation was due in part to multiple gene duplications in the inverted repeat region. Lineages also varied in their repetitive DNA content, with the Appalachian lineage displaying the highest proportion of tandem repeats (~ 10%) compared to the Eastern and Western lineages (~ 6%). In addition, genes involved in transcription and protein transport showed elevated sequence divergence between lineages, and a strong correlation was observed between genome size and repetitive DNA content. *Campanula americana* was found to have one of the most repetitive plastid genomes within Campanulaceae.

Conclusions These findings challenge the conventional view of plastid genome conservation within a species and suggest that structural variation, differences in repetitive DNA content, and divergence of key genes involved in transcription and protein transport may play a role in PNI. This study highlights the need for further research into the genetic mechanisms underlying PNI, a key process in the early stages of speciation.

Keywords Campanulaceae, Chloroplast genome evolution, Comparative genomics, Cyto-nuclear incompatibility, Plastid-nuclear incompatibility, Repetitive DNA

Background

Plastids are crucial for a number of metabolic processes in plants, including photosynthesis, starch biosynthesis, and the modulation of stress responses [1, 2]. These organelles originated from a cyanobacterial endosymbiont and thus retain a bacterial-like circular genome. Over evolutionary history, a significant portion of the genes from the ancestral cyanobacterium have been transferred

*Correspondence:

Alfredo López-Caamal
hqc5dj@virginia.edu

¹ Department of Biology, University of Virginia, Charlottesville, VA, USA

² Department of Biology, James Madison University, Harrisonburg, VA, USA



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

to the nuclear genome [3]. This transfer has led to the reduced plastid genomes observed in land plants today. Consequently, the coordinated expression of genes in both plastid and nuclear genomes is essential for sustaining metabolic functions [4, 5]. This intergenomic coordination results in strong selective pressure between nuclear and organellar genomes to maintain these functions. As a result, plastid genomes are highly conserved among most land plants [6], with ferns, gymnosperms, and most angiosperms exhibiting largely collinear structure [7]. Unsurprisingly, intraspecific variation in the plastid genomes is usually low, often limited to a small number of single nucleotide polymorphisms (SNPs) or short insertions and deletions (indels) [8, 9].

Changes in structure or sequence in the plastid genome are expected to lead to the fixation of compensatory mutations in the nuclear genome [5]. The absence of compensatory mutations can lead to a loss of organelle function due to incompatibilities between genomes, a process known as cytonuclear incompatibility, or more specifically, plastid-nuclear incompatibility (PNI). PNIs are thought to play a key role in speciation [5, 10, 11], with incompatible hybrids exhibiting low-fitness chlorotic or albino phenotypes. While PNIs are considered to be one of the earliest barriers to arise during speciation, most documented cases involve well-differentiated species, such as those in the genera *Oenothera* [12, 13] and *Pelargonium* [14]. However, several plant families, e.g., Geraniaceae [15], Fabaceae [16], Onagraceae [12], and Campanulaceae [17, 18], are known for high plastid substitution rates and structural rearrangements, and therefore may demonstrate the plastid genome evolution expected to underlie PNI.

Intraspecific PNI is required for this mechanism of speciation. One of two known cases of intraspecific PNI is *Campanula americana* (Campanulaceae) [19, 20]. The family Campanulaceae sensu lato (*s.l.*) is well known for its dynamic plastid evolution, including rearrangements, duplications, inversions, and gene losses among taxa [18]. In fact, these rearrangements are so ubiquitous that they have proven useful as an additional tool for phylogenetic inference in the family [17, 21]. For example, *Trachelium caeruleum* exhibits 18 rearranged, inverted, or relocated regions relative to the ancestral gene order of plastid genomes [22, 23]. The presence of dispersed repeats and tRNA gene duplications across the genome is thought to facilitate homologous recombination, which facilitates further plastid evolution. This idea is supported by the concentration of these elements near rearrangement endpoints. Additionally, some plastid genes in the Campanulaceae (*clpP*, *ycf1*, *ycf2*, and *ndhK*) show high levels of divergence compared to those in other plant species, suggesting high nucleotide substitution rates [23, 24].

However, the presence of variation in plastid genomes within a species in this family is unknown. By assembling and analyzing their genomes, we can shed light on intraspecific plastid genome evolution and the potential role of structural rearrangements and sequence divergence in plant speciation through PNI.

In this study, we assembled the plastid genomes of 18 individuals of *Campanula americana*, representing the three distinct lineages within the species. PNI occurs in crosses between these three lineages, evidenced by chlorotic offspring and hybrid breakdown [11]. Plastid sequence divergence and the strength of the incompatibility are correlated in *C. americana* [11], suggesting that the plastid drives the incompatibility. Combined with the dynamic plastid evolution that typifies this family, this system therefore provides an exceptional opportunity to explore intraspecific plastid genome variation and its potential role in speciation. We hypothesized that plastid structural variation, in addition to sequence divergence, exists among the *C. americana* lineages given the strong PNI among lineages. Additionally, we explored whether differences in repetitive DNA content are found between lineages, which may also contribute to PNI. The findings from this research provide insight into the evolutionary dynamics of plastid genomes and the potential role of structural and sequence divergence in plant speciation.

Materials and Methods

Study species

Campanula americana L. (= *Campanulastrum americanum* Small) is an insect-pollinated monocarpic herb native to the eastern United States, typically found along forest edges and in shaded disturbed sites. This species has three distinct plastid lineages across its range: an Appalachian lineage (A) restricted to the Appalachian Mountains, an Eastern lineage (E) located east of the Appalachian Mountains, and a widespread Western lineage (W) occurring throughout most of the range (Fig. 1A). The split between A and the clade comprising the closely related W and E lineages is estimated to have occurred approximately 2 million years ago (mya) [25]. Previous research identified strong PNI within the species during inter-lineage crosses. Crosses where W maternal plants are pollinated by A plants result in F1 progeny with albino phenotypes, showing a fitness reduction of up to 94% compared to the parental populations. Similarly, offspring from E maternal plants crossed with A paternal individuals also exhibit a fitness reduction in the F1 generation, though to a lesser extent than the W × A crosses. However, the reciprocal crosses (i.e., A as the maternal and either W or E as paternal) show no chlorosis or reduction in survival [26], indicating that the A plastid genome is compatible with both W and E nuclear

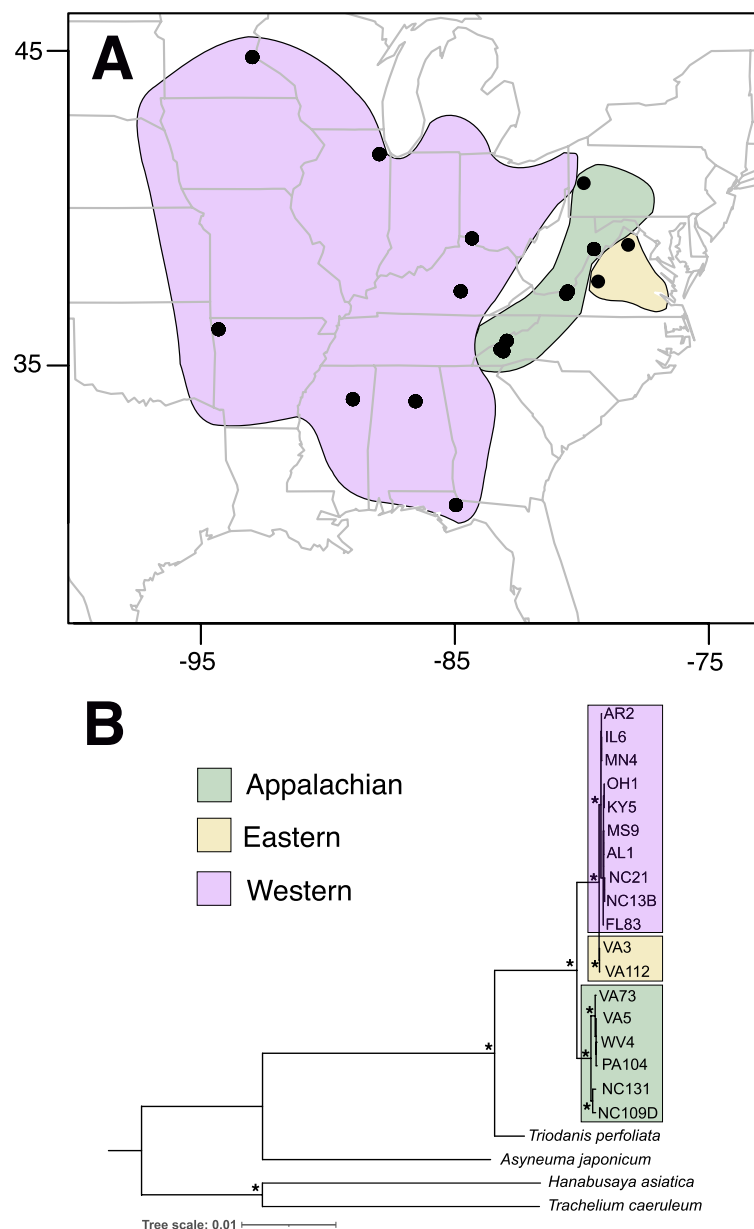


Fig. 1 **A**, Distribution of the *Campanula americana* lineages and populations sampled for sequencing. **B**, Maximum likelihood phylogenetic tree of 38 concatenated plastid coding sequences of *C. americana*. Western, Eastern and Appalachian lineages are recovered as monophyletic. Asterisks denote bootstrap values > 95

backgrounds. Previous research also found an elevated rate of nucleotide substitution in the *ycf1*, *ycf2*, *clpP*, and *rps* genes [27], which may be involved in *C. americana*'s observed PNI.

Plant material and DNA extraction

We collected and grew seeds from 18 populations across *C. americana*'s range (Fig. 1A, Table S1), including all lineages identified in a previous plastid phylogeographic

study [25]. Additionally, we grew *Triodanis perfoliata*, the sister species to *C. americana* [28]. We germinated seeds in a 3:1 mixture of peat moss and turf, placing the trays in growth chambers under controlled conditions with a 12-h light–dark cycle (21 °C during the day and 14 °C at night). After approximately four weeks, ~ 300 mg of fresh leaf tissue was collected from one individual per population and stored at – 80 °C until DNA extraction.

We extracted high-molecular-weight genomic DNA using a modified CTAB protocol (Doyle & Doyle 1987) (Additional File 1). Briefly, we ground the leaf tissue with a pestle in liquid nitrogen and added a lysis buffer containing CTAB, sorbitol, and sarkosyl to the homogenate, followed by two chloroform extractions and isopropanol precipitation. We assessed the purity of the DNA using spectrophotometry with a NanoDrop and Qubit fluorometry. For samples with low-quality DNA ($A_{260}/A_{280} > 1.9$ and $A_{260}/A_{230} < 1.7$), we performed additional cleaning using an in-house magnetic bead-based protocol.

Long-read sequencing using Oxford Nanopore Technologies platform

We prepared sequencing libraries for all *C. americana* and *T. perfoliata* samples using the Rapid Barcoding Kit (SQK-RBK-004) according to the manufacturer's protocol with an input of ~100 ng of high-molecular weight DNA. We multiplexed up to six barcoded libraries at a time. Sequencing was performed on a MinION Mk1B sequencer (72 h sequencing run) using MinION flow cells (R10.4). Some of the sequencing was done as part of the genetics laboratory course at James Madison University with the support of Oxford Nanopore's Education Beta program.

Chloroplast-specific basecalling model training and de novo plastid genome assembly

To develop a basecaller model, we first needed a high-quality reference dataset for model training and accuracy assessment. We constructed a reference chloroplast genome assembly using PacBio HiFi reads from an individual belonging to the Appalachian lineage (population VA73, Table S1). Library preparation and sequencing were performed through Phase Genomics, Inc (Seattle, WA). Raw HiFi reads were mapped to a subset of plastid genes (*matK*, *rbcL*, *ndhF*) available for several Campanulaceae species using minimap2 v2.27 [29] (Additional File 2). We then assembled the mapped reads with Flye v2.9.3 [30]. Next, we mapped the raw reads to this initial assembly and used the mapping reads for a second round of Flye assembly. The contig corresponding to the chloroplast genome was extracted from the assembly graph using `get_organelle_from_assembly.py`, a script included in the GetOrganelle suite [31]. We considered this assembly the "true" plastid genome of *C. americana*.

We then used Bonito v0.7.3, a basecaller developed by Oxford Nanopore Technologies (ONT), to train a custom basecalling model for *C. americana* (hereafter referred to as the custom model). For this purpose, we used the raw current signal FAST5 file from a 72-h MinION sequencing run of an individual from the same population

(VA73). We performed the initial basecalling using a pretrained model (`dna_r9.4.1_e8_hac@v3.3`) with the parameters `-reference` and `-save-ctc`, which identified the true sequence of each read by aligning it to our PacBio HiFi reference genome. We then used Bonito to train a custom basecalling model with the parameters `-batch 10,000`, `-chunks 1,000,000`, `-epochs 12`, and `-lr 1e-4`. We conducted the training on the University of Virginia's High-Performance Computing system (Rivanna) using a single node with NVIDIA Volta V100 GPUs. We subsequently used this custom model to basecall all samples, followed by de novo chloroplast genome assembly using the previously described pipeline. We obtained final consensus sequences with Medaka v1.5.0 (<https://github.com/nanoporetech/medaka>). Additionally, we basecalled the raw Nanopore signal using ONT's built-in basecalling model, Guppy (hereafter referred to as the Guppy model).

We performed gene annotation of all assemblies by mapping available exon sequences previously published for *C. americana* [27]. We used GeSeq (Tillich et al. 2017) to annotate tRNA, rRNA, and gene fragments, and manually checked and edited intron–exon boundaries in Geneious Prime 2024.0.7 (<https://www.geneious.com>).

Phylogenetic analyses

To investigate the phylogenetic relationships among the *C. americana* lineages based on chloroplast data, we constructed a maximum likelihood (ML) tree. We included the sister species *T. perfoliata*, as well as the plastid genomes of the closely related species *Asyneuma japonicum* (OR805474), *Hanabusaya asiatica* (NC_024732), and *Trachelium caeruleum* (NC_010442), which are available in GenBank. We extracted coding sequences (CDS) from the GenBank files and aligned them using MAFFT v7.490 [32] with the `-auto` option enabled. We manually corrected the alignments in Geneious Prime and trimmed them to remove gappy regions, allowing up to 10% gaps within a column. We excluded gene alignments with more than 95% identical sites from the analysis. We concatenated the resulting 36 CDS alignments and constructed a ML phylogenetic tree with IQ-TREE [33], using the GTR + F + G4 model identified by ModelFinder [34] and 1000 ultrafast bootstrap replicates.

Structural variation analysis

Structural variation and size differences in plastid genomes are typically closely associated with expansions and contractions of the inverted repeat regions. Therefore, we first compared the boundaries between the single-copy elements and inverted repeat regions of *C. americana* and *T. perfoliata* using IRplus [35]. Next, we investigated the presence of structural variation across

the genomes of all three lineages of *C. americana* and *T. perfoliata*. To accomplish this, we performed pairwise whole genome alignments using the nucmer application in MUMer4 [36]. We filtered alignments with an identity of less than 90% and shorter than 100 bp using the delta-filter application. We identified syntenic regions and genomic rearrangements (duplications, translocations, and inversions) using SyRI v1.7 [37]. Finally, we visualized genomic rearrangements using plotsr v1.1 [38] and the R package gggenes v0.4.1 (Wilkins 2020).

Genetic diversity and repetitive DNA content analysis

We investigated nucleotide diversity (π) both between and within the *C. americana* lineages. To do this, we performed a whole genome alignment, removing all columns with more than 10% gaps. We then conducted a sliding window analysis of π across the genome using windows of 800 bp and a step size of 600 bp in DNAsp v6 [39]. In addition, to explore gene sequence divergence between lineages, we aligned each of the 72 protein-coding genes in *C. americana* using MAFFT v7.490 with the `-auto` option enabled. We manually corrected the alignments in Geneious Prime and calculated π for each gene alignment using the R-package `pegas` v1.3 [40].

Next, we examined differences in repetitive DNA content among the *C. americana* lineages. To compare the repetitive DNA of *C. americana* with that of other species, we downloaded 70 plastid genomes from various Campanulaceae species, the reference genome of *Helianthus annuus*, and several species from Rouseaceae, the sister family to Campanulaceae (Table S2). We analyzed tandem repeats in all genomes using Tandem Repeats Finder v4.09 [41] with the following parameters: match = 2, mismatch = 7, delta = 7, match probability = 80, and indel probability = 10. We set the minimum alignment score to 50 and the maximum period size to 500. We analyzed both direct and palindromic dispersed repeats with Vmatch (Kurtz 2017), using a minimum repeat length of 30 bp, a Hamming distance of 3, and a minimum identity of 98, while recording the best 500 matches. For all genomes, we masked the repetitive DNA and calculated the percentage of the genome (including only one inverted repeat) containing either tandem or dispersed repeats. Finally, we explored the relationship between genome size and the percentage of repetitive DNA content.

Results

Campanula americana reference chloroplast genome and performance of custom basecalling model.

The complete chloroplast genome of *C. americana* assembled from PacBio HiFi reads has a length of 193,622 bp and a GC content of 37.1%. This assembly,

considered the “true” chloroplast genome for *C. americana*, was used to train a custom basecalling model. The length of the assembly obtained with the custom model (193,597 bp) was similar to that of the reference, while the Guppy model produced a slightly shorter assembly (193,341 bp). Interestingly, both models showed no mismatches relative to the reference genome. However, the Guppy model had a higher number of single-nucleotide indels (255) compared to the custom model [42] (Table S3). These indels occurred in low-complexity regions throughout the genome, resulting in frameshift mutations and early stop codons in 35 genes for the Guppy model versus 4 genes for the custom model (Table S3). Therefore, while the custom model greatly enhanced assembly quality compared to the Guppy model, a small number of sequencing and assembly errors remained. As these errors can lead to erroneous interpretations of the reading frame, we limited our analysis to structural variation and nucleic acid sequence and excluded analysis of the amino acid sequence, which will be addressed in future research.

Plastid genome structure and phylogenetic relationships of Triodanis perfoliata and Campanula americana.

The chloroplast genomes of the W (188,309–201,788 bp), E (190,713–197,234 bp), and A (190,375–197,108 bp) lineages of *Campanula americana*, as well as the sister taxon *Triodanis perfoliata* (180,196 bp), were fully assembled into circular contigs (Fig. 2, Table S4, Figure S1). All assemblies exhibited a GC content of approximately 37% and displayed the characteristic quadripartite structure typical of most plant chloroplast genomes, comprising a large single-copy (LSC) region, a small single-copy (SSC) region, and two identical inverted repeat (IR) regions. *Campanula americana* showed striking variation in genome size, ranging from 188,309 to 201,788 bp among lineages.

Both *C. americana* and *T. perfoliata* plastid genomes encode the same set of 110 unique genes, 18 of which are duplicated in the IR region. They share 72 intact protein-coding genes of known function, including four *ycf* genes (*ycf1*, *ycf2*, *ycf3*, and *ycf4*), 30 tRNAs, and four rRNAs (23S, 16S, 4.5S, and 5.5S). Eight genes contain a single intron, while only two (*ycf3* and *clpP*) contain two. As in most angiosperms, the gene *rps12* undergoes trans-splicing during transcription. Notably, the gene *rps16*, which typically contains an intron, is absent in both *C. americana* and *T. perfoliata*. Additionally, as observed in other Campanulaceae species, the genes *ycf15*, *accD*, *rpl23*, and *infA* are truncated and likely non-functional [23] (Table S5). Most protein-coding genes utilize the standard start codon for methionine (ATG), with only three exceptions: *ndhD* and *psbL* use ACG, which is known to be RNA-edited to AUG [43],

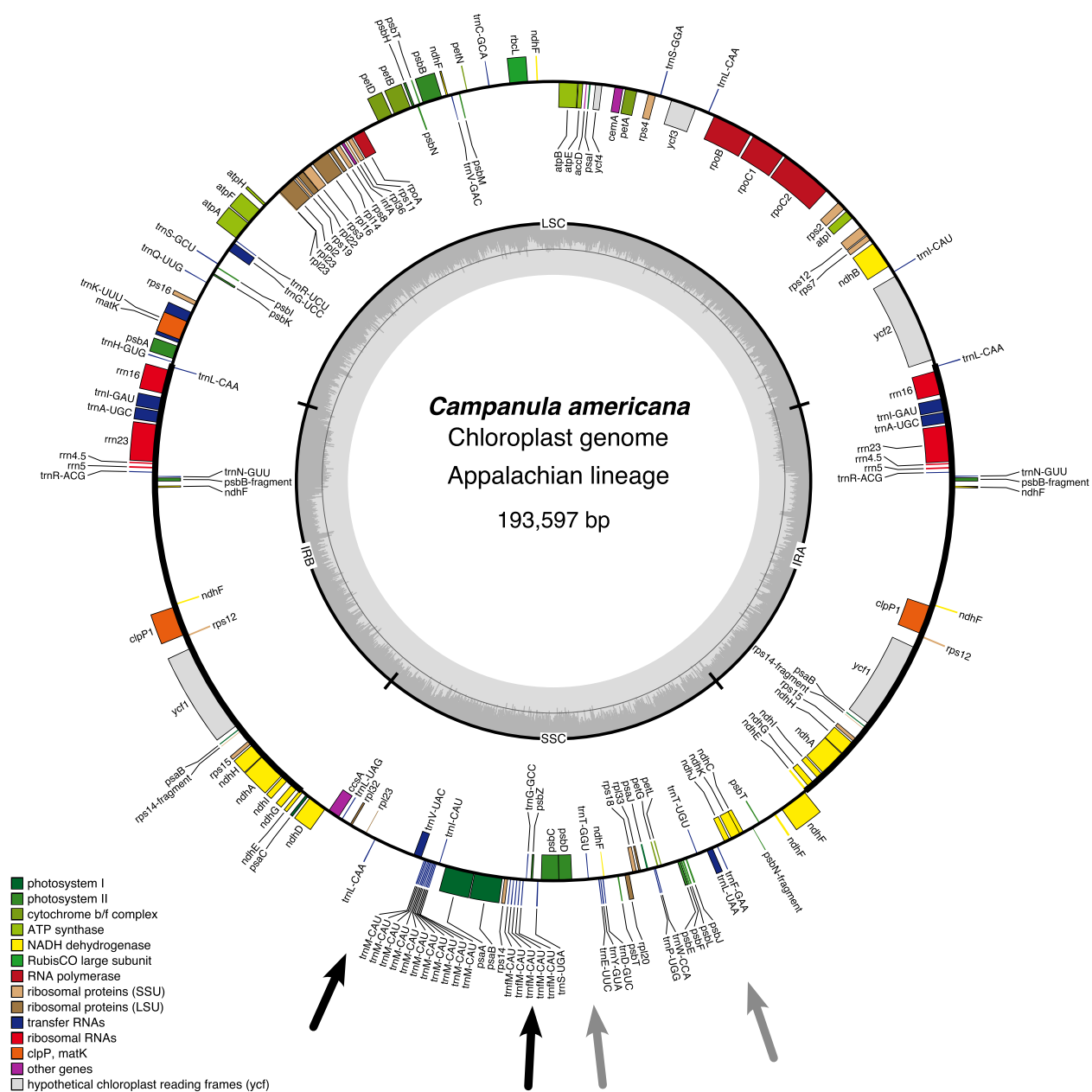


Fig. 2 Gene map of the *C. americana* chloroplast genome (Appalachian lineage). Genes shown on the inside are transcribed clockwise, while those on the outside counterclockwise. Black (inversion 1) and gray arrows (inversion 2) in the SSC region mark inversion endpoints relative to *T. perfoliata*

while *rps19* uses the bacterial alternative start codon GTG. While all genomes share the same set of protein-coding, tRNA, and rRNA genes, we found differences in the number of partial or complete gene duplications. For example, *T. perfoliata* contains three copies of the entire *psbT* gene, eight copies of a fragment of *ndhF*, three copies of *trnL-CAA*, and two copies of *trnM-CAU*. Within *C. americana*, we found variation in gene copy numbers

among lineages. The W and E lineages had the highest number of protein-coding gene copies, with 10–14 copies of a *ndhF* fragment, one copy each of a *psaB* fragment and *psbB*, three of *psbN*, 4–6 copies of the complete *psbT* gene, and one partial duplication of *rps14*. Meanwhile, the A lineage showed only 5–6 copies of a *ndhF* fragment and 2–3 copies of the entire *psbT* gene but shared the *psbN* and *rps14* gene copies with the W and E lineages. However, the Appalachian lineage displayed a higher

number of tRNA gene duplications, with 4–7 copies of *trnM-CAU* and 4–9 copies of *trnM-CAU* in the SSC region, compared to 1–4 copies of these genes in the W and E lineages. As shown below, these duplications are associated with repetitive DNA content and inversion endpoints relative to *T. perfoliata*.

Consistent with previous studies [25], our phylogenetic analysis indicates that *T. perfoliata* is sister to *C. americana*. Additionally, each lineage of *C. americana* was recovered as a monophyletic clade with highly supported nodes. The A lineage is basal to the clade formed by the closely related W and E lineages (Fig. 1B).

Absence of structural variation at the boundaries of single-copy and IR regions

The boundaries between the single-copy regions and the IR regions showed no substantial differences within *C. americana* or between *C. americana* and *T. perfoliata* (Fig. 3, Figure S2). The junction between the LSC region and the IRB occurs between the genes *trnH-GUG* and *trnL-CAA* in all genomes. The junction of the small single-copy (SSC) region and the IRB is located within the *ndhF* gene, while the junction between the SSC and IRA is situated next to the truncated copy of *ndhE* in the IRA and the *ndhF* gene of the SSC (Fig. 3).

Structural variation between the T. perfoliata and C. americana plastid genomes.

The genome of *C. americana* exhibits two large inversions in the SSC of 15,328 bp and 10,928 bp in length, relative to *T. perfoliata* (Fig. 4A). In addition, two translocations between *T. perfoliata* and *C. americana* were found: a 378 bp sequence including *trnL-CAA* translocated from the IR of *T. perfoliata* to the LSC of the *C. americana*, and a 676 bp region, including a copy of *psbT*, was translocated from the LSC of *T. perfoliata* to the SSC of *C. americana*. The synteny analysis showed that a portion of the IR and the endpoints of the two large

inversions are highly divergent regions between *T. perfoliata* and the A lineage of *C. americana* (Fig. 4A).

Structural variation within and between the C. americana lineages.

Syntenic analysis showed that the *C. americana* plastid genomes are syntenic within the E lineage. The W and A lineages, although largely syntenic, exhibited additional duplications in the IR region in some individuals (Figure S3). However, when comparing synteny between lineages, we observed a number of rearrangements, particularly in the IR region and at the inversion endpoints relative to *T. perfoliata*. For instance, the comparison between the Appalachian and Eastern lineages revealed that the inversion endpoints relative to *T. perfoliata* and a portion of the IR are highly diverged sequences, producing a lack of synteny between them (Fig. 4A). Interestingly, in this region of the IR, the Eastern lineage has several copies of two Appalachian sequences that flank this highly diverged region: a 514 bp sequence, which includes a fragment of *ndhF* and a fragment of *clpP* exon 1, is found six times in the IR, where three copies of a 396 bp region containing a fragment of *ndhF* are also found (Fig. 4B). Meanwhile, the Western and Eastern lineages were largely syntenic, with only one major duplication of a 2,386 bp region in the IR, originally occurring in the SSC of the Eastern lineage, including a fragment of *ndhF* and a complete copy of *psbT* (Fig. 4A, B).

Most of the duplicated sequences in *C. americana* are present in *T. perfoliata* as single-copy elements. For example, the 514 bp sequence, including the *ndhF* fragment and *clpP* exon 1, is located within the original *clpP* gene sequence in *T. perfoliata*, suggesting that recombination or errors during genome replication are responsible for the duplication of this sequence in *C. americana*. Notably, the observed structural variation and repetitive elements occur in non-coding sequences, with no repetitive elements found within operons. Most of the

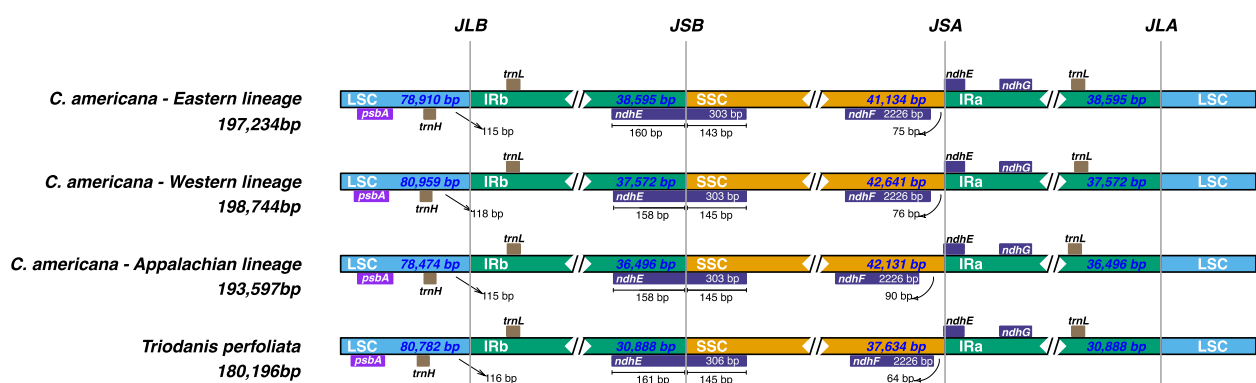


Fig. 3 Comparison of the junction between the large single copy (LSC), small single copy (SSC) and inverted repeats (IR) of the *C. americana* lineages and *T. perfoliata* plastid genomes

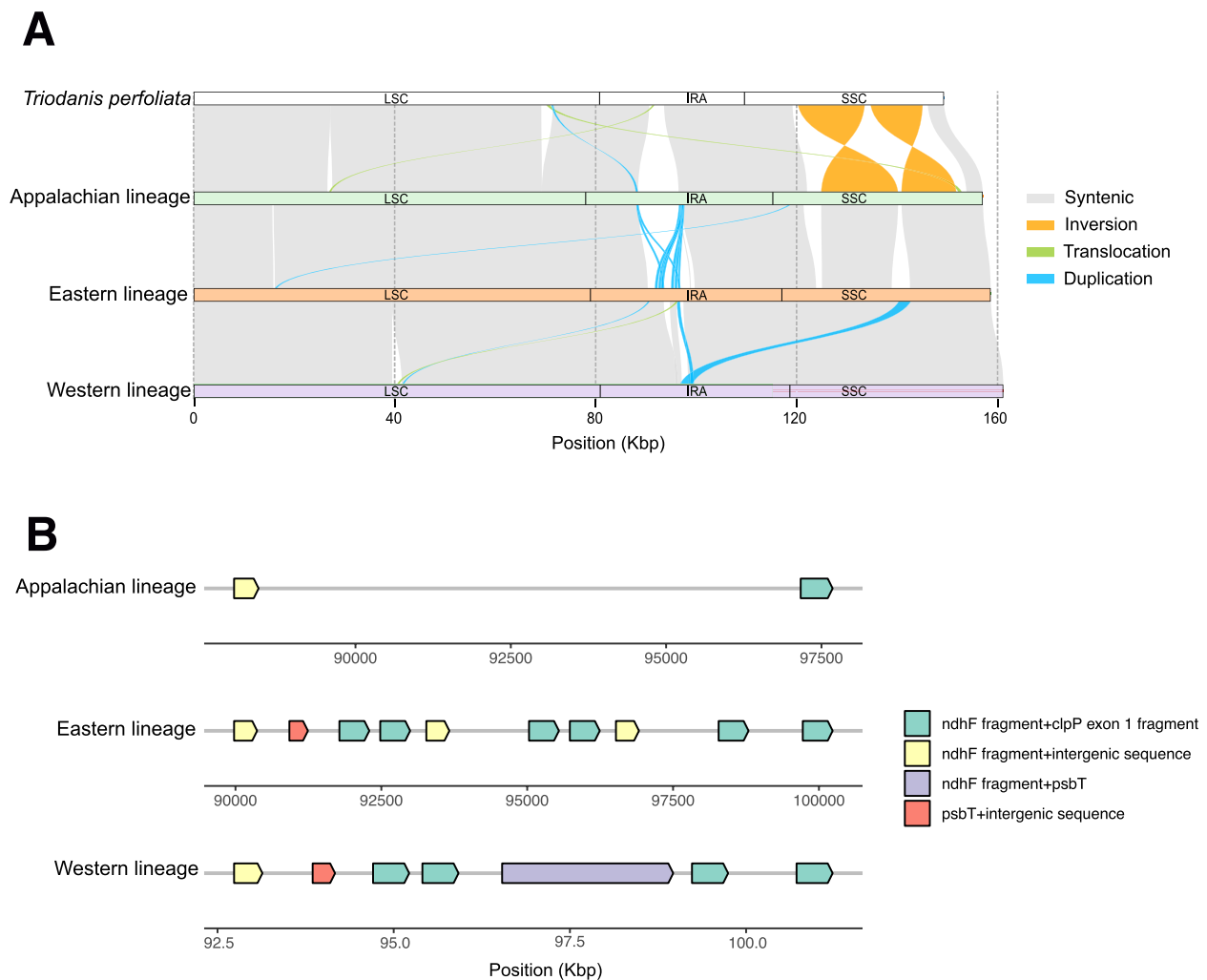


Fig. 4 Structural variation between *T. perfoliata* and *C. americana*. **A**, Synteny analysis reveals two large inversions between *T. perfoliata* and *C. americana*. The *C. americana* lineages also exhibit several duplications and rearrangements in the Inverted Repeat (IR) regions. **B**, Detailed view of a portion of the IR regions in the Appalachian, Western, and Eastern lineages of *C. americana*. Each duplicated region is represented by a different color, with corresponding annotations indicated

structural variation is confined to the IR region mentioned above and the inversion endpoints relative to *T. perfoliata*. Therefore, we have no evidence suggesting that gene transcription could be affected due to structural variation.

Sequence divergence and repetitive DNA content in Campanula americana.

Our results indicate a strong correlation between genome size and total repetitive DNA content, including dispersed and tandem repeats, in Campanulaceae. *Campanula americana* has one of the most repetitive plastid genomes within the family, with ~20% of its genome consisting of repetitive DNA (Fig. 5A). We observed no dramatic differences between the *C. americana* lineages, except that four W lineage populations (AL1, KY5, OH1,

FL83) showed a higher proportion of repetitive DNA compared to the other populations. These populations also exhibited additional duplications in the IR, as shown above in the synteny analysis. However, when analyzing only the tandem repeat content, the A lineage had the highest proportion of repeats (~10%), while the W and E lineages had ~6% tandem repeat content. The proportion of tandem repeats in the A lineage is one of the highest in Campanulaceae, where most species have less than 5% of their genome as tandem repeats. *Adenophora racemosa* is the species with the highest tandem repeat content in the family (Fig. 5B).

The nucleotide sequence of the repeats differs between the *C. americana* lineages, leading to high nucleotide diversity hotspots across the plastid genome (Fig. 6A).

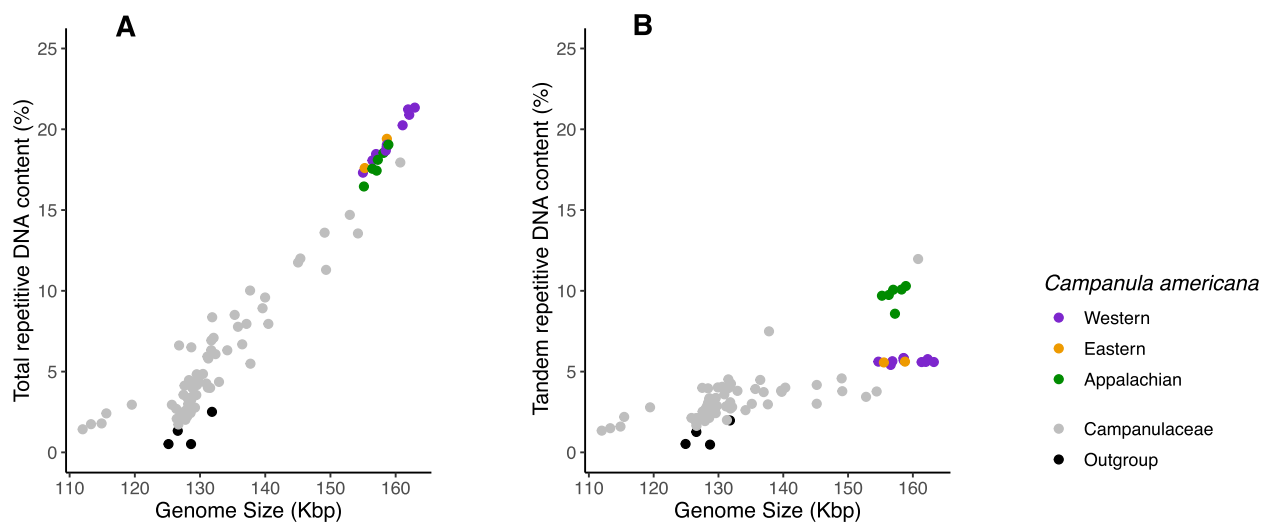


Fig. 5 Relationship between plastid genome size and (A) total (dispersed + tandem repeats) and (B) tandem repetitive DNA content. Species within Campanulaceae are shown in gray, while outgroups (*Carpodetus serratus*, *Helianthus annuus*, *Pentaphragma spicatum*, and *Stylidium debile*) are shown in black. The lineages of *C. americana*, shown in purple (Western), orange (Eastern) and green (Appalachian) show high amount of repetitive DNA

We identified six diversity hotspots in intergenic regions in the all-lineage comparison ($\pi = 0.08\text{--}0.27$). Two of these were found in the LSC, one in the IR, and three in the SSC region (Fig. 6A). These hotspots correspond to regions where duplications and rearrangements have occurred. Interestingly, these diversity hotspots are enriched in repetitive DNA motifs unique to each lineage (Fig. 6B). In addition, the diversity hotspots in the SSC region align with the endpoints of the large inversions relative to *T. perfoliata*, while the one in the IR matches the region where sequence duplications have occurred. The hotspots in the LSC were frequently found in regions where duplications and translocations have occurred among *C. americana* lineages, as revealed by the synteny analysis (Fig. 4A). The nucleotide diversity pattern observed is driven by the differences between the A lineage and both the W and E lineages, while the W and E lineages showed low sequence divergence between them (Figure S4). When examining nucleotide diversity within individual lineages, we found that the diversity hotspots in the A and W lineages generally occur in the same regions as those observed when comparing all lineages, though with lower π values (0.02–0.12). In contrast, the Eastern lineage showed very low levels of nucleotide diversity across the entire genome (Fig. 6A), though sampling was also smaller in this lineage.

Finally, the analysis of nucleotide diversity in individual gene alignments showed that the genes encoding the large and small subunits of the ribosome, *rpl* and *rps*, have the highest diversity values (mean $\pi = 3.79 \times 10^{-3}$ and 2.93×10^{-3} , respectively), followed by the *ycf* genes

(mean $\pi = 2.38 \times 10^{-3}$). In contrast, the genes encoding subunits of photosystems I (mean $\pi = 0.31 \times 10^{-3}$) and photosystem II (mean $\pi = 0.29 \times 10^{-3}$), as well as those encoding subunits of the NADH dehydrogenase complex (mean $\pi = 0.98 \times 10^{-3}$), showed lower mean π values (Fig. 6C). However, some individual genes within these categories exhibited higher diversity values, such as *psbH* ($\pi = 2.05 \times 10^{-3}$), *ndhJ* ($\pi = 2.32 \times 10^{-3}$), and *ndhF* ($\pi = 2.44 \times 10^{-3}$).

Discussion

We found unprecedented variation in chloroplast genome size, primarily driven by structural rearrangements, gene duplications, and repetitive DNA elements, within the native range of the species *Campanula americana*. This contrasts with the typically conserved structure of chloroplast genomes, especially within species where differences are usually limited to SNPs and short indels [44]. In addition, we found high sequence divergence in genes involved in transcription among lineages. Our findings demonstrate complex plastid genome evolution in *C. americana* and support its potential contribution to reproductive isolation through plastid-nuclear incompatibility (PNI).

Structural and sequence variation in *C. americana*.

The observed variation in chloroplast genome size among *C. americana* lineages exceeds that found in other species. For instance, studies of highly polymorphic *Utricularia amethystina* report minor differences in plastid genome size (312bp between assemblies) [45], and three species in *Gentiana* each have modest intraspecific

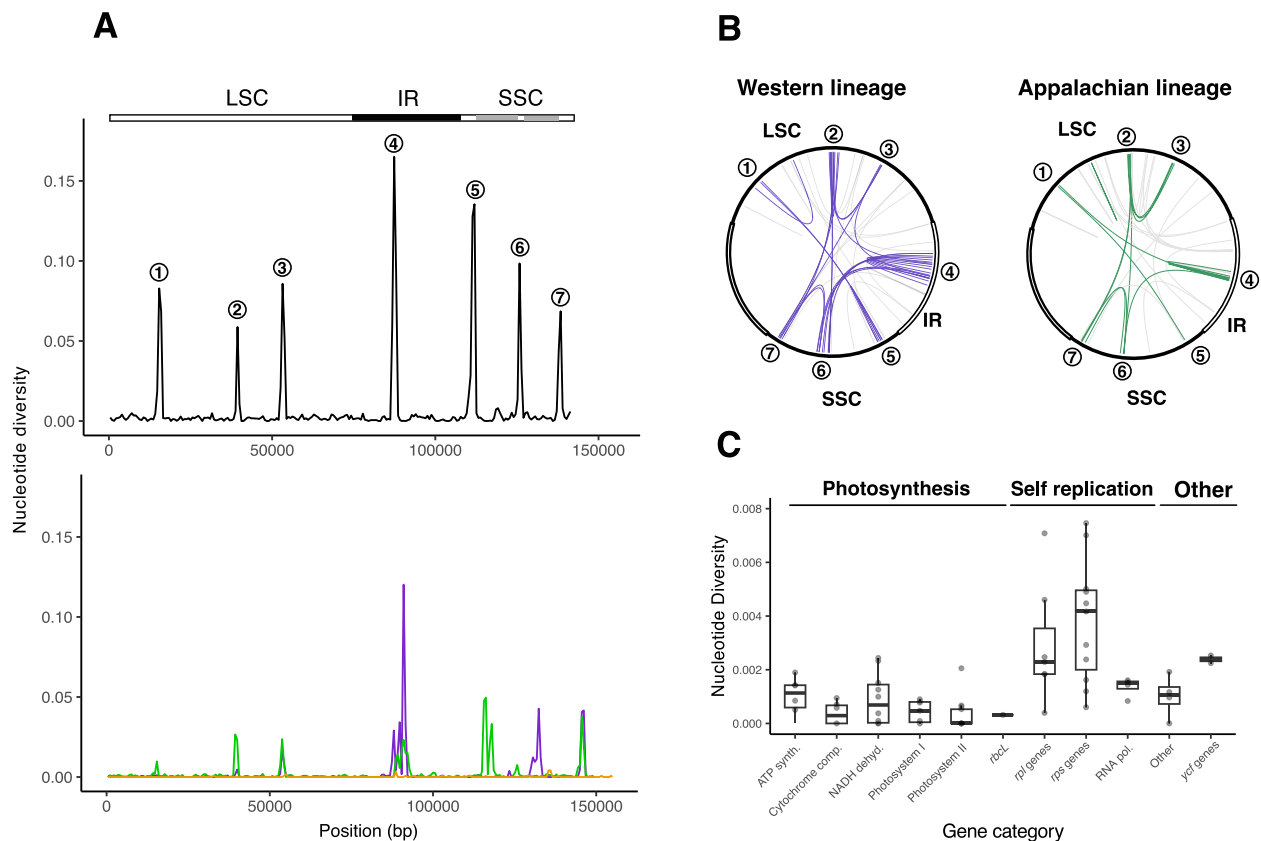


Fig. 6 Nucleotide diversity (π) across the plastid genome of *Campanula americana*. **A**, Upper panel shows the sliding window analysis of π obtained from a whole-genome alignment including all Western, Eastern and Appalachian lineages. Lower panel shows the π analysis within-lineage (Western, purple; Eastern, orange; Appalachian, green). **B**, Dispersed repeats across the Western lineage and Appalachian lineage plastid genome. Gray links represent shared repeats between genomes, while purple and green represent repeats with different nucleotide sequence for Western and Appalachian lineages, respectively. The π hotspots shown in panel A match the regions with divergent repetitive DNA sequences between lineages. **C**, Nucleotide diversity of individual gene alignments. The genes involved in replication show higher π values compared to genes involved in photosynthesis

genome size variation, ranging from 285 to 628 bp [44]. Whereas in *C. americana*, we observed a difference in genome size at least 20-fold larger (~ 13 Kbp), even though all plastid genomes share the same set of genes. Variation in plastid genome size in *C. americana* can be attributed to duplications and sequence divergence, particularly in the IR region, as well as an abundance of repetitive DNA. These types of variation are distinct from the intraspecific variants found in other taxa, e.g. SNPs and short indels [9, 46].

The split between the A and W/E clades of *C. americana* is estimated to have occurred 2 mya [25]. While this timeframe may seem sufficient for genomic variation to accumulate, plastid genomes are known for their high conservation in both coding and non-coding sequences across phylogenetic scales. For example, plastid genome structure is highly conserved among members of Daphniphyllaceae, despite a divergence time of 2.1 mya [47]. Moreover, the plastid genomes of this family remain

highly collinear with those of Crassulaceae and Grossulariaceae, which share a common ancestor from 99.6 mya [47]. Similarly, within Campanulaceae *s.l.*, the three lineages of *Lobelia columnaris* exhibit identical plastid genome structure, with only minimal differences in genome size, despite diverging 1.5 mya [48]—a timeframe comparable to that of *C. americana*. Given these patterns, the observed structural and genome size differences between *C. americana* lineages are unprecedented and underscore the unusually rapid evolution of its plastid genome.

According to previous research, the expansion and contraction of the IR boundary is thought to be a primary cause of plastid genome size variation in flowering plants [15], particularly in Campanulaceae [17, 22]. However, our results indicate that repetitive DNA content, rather than IR boundary expansion or contraction, is the primary driver of genome size variation in *C. americana*. We found that total repetitive DNA content correlates

strongly with genome size across Campanulaceae, with *C. americana* having some of the most repetitive content across plastid genomes. The repeats in *C. americana*, particularly dispersed repeats, are concentrated in specific regions of the genome, especially near inversion endpoints relative to *Triodanis perfoliata*. Research in other taxa has also found a positive association between frequency of repetitive elements and structural variation [15, 44]. Therefore, repetitive elements likely facilitated the structural variation observed in *C. americana* by mechanisms such as illegitimate recombination [49], leading to an increase of repetitive DNA and consequently to an increase of the plastid genome size. This supports the idea that certain regions of the genome, particularly those rich in repetitive elements, are more susceptible to accumulating variation.

The structural variation observed in *C. americana* is part of a broader trend within Campanulaceae, where plastid genomes are characterized by high levels of repetitive DNA and genome size is correlated with repetitive content. For example, repetitive DNA has been shown to facilitate structural rearrangements such as inversions, duplications and translocations [18, 23]. In *C. americana*, the non-random distribution of repeats and their association with structural rearrangements suggest that these elements play a critical role in shaping the evolutionary dynamics of plastid genomes in this family, even at shallow taxonomic levels. Furthermore, differences in nucleotide sequence of these repeats between the *C. americana* lineages suggests independent evolution and lineage-specific evolutionary trajectories, reinforcing the recent and ongoing nature of this evolution. The results of this research add to recent suggestions to reevaluate the drivers of organellar genome evolution, in light of emerging evidence that challenges traditional mutation models in plant organelles [50].

Plastid-nuclear incompatibility and speciation in C. americana.

The observed structural variation and sequence divergence in *C. americana* may be linked to PNI, a phenomenon where incompatibilities between the nuclear and plastid genomes lead to reduced fitness in hybrids. Specifically, the Bateson-Dobzhansky-Muller model [51, 52], states that postzygotic isolation results from the accumulation of genetic incompatibilities between loci—or in the case of PNI, between genomes. In this context, the unusually rapidly evolving plastid genome of *C. americana* has accumulated divergent mutations among lineages (e.g., gene duplications, accumulation of repetitive DNA, sequence divergence), likely since the divergence of the A and W/E plastid lineages approximately 2mya [25]. As these potentially maladaptive mutations became fixed in each lineage, compensatory mutations in the nuclear

genome were likely to have been selected for to maintain plastid metabolism. This within lineage co-evolution between genomes is expected to underlie the hybrid breakdown in between-lineages crosses [11] or upon secondary contact. However, the question remains whether gene sequence divergence or structural variation (e.g., duplications, repetitive DNA content) is the direct cause of PNI in *C. americana*.

Previous research found that the magnitude of PNI in *C. americana* is positively correlated with plastid SNP differences between lineages. Crosses between the lineages with the highest plastid genetic distance (WxA) result in albino phenotypes in the F1, accompanied by a drastic fitness reduction. In contrast, crosses between less genetically diverged lineages (WxE) show little to no fitness reduction [26]. In this study, we observed a similar pattern, where the highly incompatible W and A lineages exhibited more structural differences between them compared to the more compatible W and E lineages. However, since the structural variation occurs in non-coding sequences, it is likely not directly responsible for the PNI in *C. americana*. On the other hand, we found that the genes encoding the ribosomal large and small subunits (*rpl* and *rps*), as well as *ycf1* and *ycf2*, and some genes involved in photosynthesis (*psbH*, *ndhJ* and *ndhF*) show elevated sequence divergence among lineages. This divergence, rather than structural variation, may be responsible for PNI in *C. americana*. The plastid ribosome is composed of two subunits, the 30S and the 50S, with their protein components encoded by both plastid and nuclear genes [53]. Additionally, *ycf1* and *ycf2* are integral parts of the translocation complexes in the outer and inner plastid membranes, which are also composed of nuclear and plastid encoded proteins [54, 55]. These complexes mediate the import of most nuclear-encoded proteins targeted to plastids [56]. While our findings suggest that sequence divergence in these genes may contribute to PNI, further analyses are needed to determine how these variations affect plastid-nuclear interactions. Given that both the ribosome and the translocation complexes rely on proteins encoded by both genomes, assessing the impact of plastid sequence divergence on protein interactions requires nuclear genomic data. Additionally, plastid mRNA undergoes RNA editing, which could further modify protein interactions. Future studies incorporating nuclear genomic and transcriptomic data will provide insights into the molecular mechanisms underlying PNI in *C. americana*.

Conclusions

Chloroplast genomes typically exhibit both conserved structure and sequence within a species, with SNPs and short indels being the most common differences

between accessions. However, in this study, we reveal substantial structural diversity in the chloroplast genome of *C. americana*. We demonstrate that large differences in genome size (180 - 200 Kbp) occur within the species, driven by structural variations in the IR region, partial to entire gene duplications, and a concentration of repetitive DNA and duplicated tRNA genes at or near inversion endpoints. Additionally, we identified genes with high nucleotide diversity, which could lead to differences in the protein structure. These findings contribute to our understanding of the evolutionary processes leading to PNI and provide new insights into the genetic architecture of this mechanism of speciation.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12864-025-11525-w>.

Additional file 1. High-molecular weight DNA extraction protocol for *Campanula americana*.

Additional file 2. Gene sequences used as baits in the first round of plastid genome assembly.

Additional file 3.

Acknowledgements

We thank C. Claussen for his help on raising of plants; L. Elhady, A. Perrier, and M. J. Gower-Fici for their assistance during DNA extraction, and the Galloway laboratory for discussion. The authors acknowledge Research Computing at The University of Virginia for providing computational resources and technical support that have contributed to the results reported within this publication. URL: <https://rc.virginia.edu>.

Authors' contributions

All authors contributed to the study design. ALC performed all analyses and wrote the manuscript with input from LG and KBK. ALC and TG performed DNA extractions. TG and KBK performed library prep and DNA sequencing. All authors read and approved the final manuscript.

Funding

This work was supported by the National Science Foundation [grant numbers DEB- 2140190, DEB- 2140189] and Oxford Nanopore's Education Beta program.

Data availability

Raw sequence reads are archived in the SRA: Bioproject PRJNA1158461. Plastid genome assemblies have been deposited in the NCBI GenBank. Accession numbers can be found in Table S4. Scripts used for analysis and figure generation are archived in Github: https://www.github.com/Alfredo-LC/Plastid_Campanula.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 23 October 2024 Accepted: 25 March 2025

Published online: 04 April 2025

References

- Wicke S, Schneeweiss GM, dePamphilis CW, Müller KF, Quandt D. The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. *Plant Mol Biol*. 2011;76(3):273–97.
- Sierra J, Escobar-Tovar L, Leon P. Plastids: diving into their diversity, their functions, and their role in plant development. *J Exp Bot*. 2023;74(8):2508–26.
- Sibbald SJ, Archibald JM. Genomic insights into plastid evolution. McFadden G, editor. *Genome Biol Evol*. 2020;12(7):978–90.
- Sloan DB, Warren JM, Williams AM, Wu Z, Abdel-Ghany SE, Chicco AJ, et al. Cytonuclear integration and co-evolution. *Nat Rev Genet*. 2018;19(10):635–48.
- Postel Z, Touzet P. Cytonuclear genetic incompatibilities in plant speciation. *Plants*. 2020;9(4):487.
- Dobrogojski J, Adamiec M, Luciński R. The chloroplast genome: a review. *Acta Physiol Plant*. 2020;42(6):98.
- Raubeson LA, Jansen RK. Chloroplast genomes of plants. *Plant Divers Evol Genotypic Phenotypic Var High Plants*. 2005;45–68.
- Choi IS, Wojciechowski MF, Steele KP, Hopkins A, Ruhlman TA, Jansen RK. Plastid phylogenomics uncovers multiple species in *Medicago truncatula* (Fabaceae) germplasm accessions. *Sci Rep*. 2022;12(1):21172.
- Park J, Xi H, Kim Y. The complete chloroplast genome of *Arabidopsis thaliana* isolated in Korea (Brassicaceae): An investigation of intraspecific variations of the chloroplast genome of Korean *A. thaliana*. *Int J Genomics*. 2020;2020(1):3236461.
- Chou JY, Leu JY. Speciation through cytonuclear incompatibility: Insights from yeast and implications for higher eukaryotes. *BioEssays*. 2010;32(5):401–11.
- Barnard-Kubow KB, So N, Galloway LF. Cytonuclear incompatibility contributes to the early stages of speciation. *Evolution*. 2016;70(12):2752–66.
- Greiner S, Wang X, Rauwolf U, Silber MV, Mayer K, Meurer J, et al. The complete nucleotide sequences of the five genetically distinct plastid genomes of *Oenothera*, subsection *Oenothera*: I. Sequence evaluation and plastome evolution. *Nucleic Acids Res*. 2008;36(7):2366–78.
- Zupok A, Kozul D, Schöttler MA, Niehörster J, Garbsch F, Liere K, et al. A photosynthesis operon in the chloroplast genome drives speciation in evening primroses. *Plant Cell*. 2021;33(8):2583–601.
- Breman FC, Snijder RC, Korver JW, Pelzer S, Sancho-Such M, Schranz ME, et al. Interspecific hybrids between *Pelargonium × hortorum* and species from *P. section Ciconium* reveal biparental plastid inheritance and multi-locus cyto-nuclear incompatibility. *Front Plant Sci*. 2020 [cited 2024 Aug 16];11. Available from: <https://www.frontiersin.org/journals/plant-science/articles/https://doi.org/10.3389/fpls.2020.614871/full>.
- Weng ML, Blazier JC, Govindu M, Jansen RK. Reconstruction of the ancestral plastid genome in Geraniaceae reveals a correlation between genome rearrangements, repeats, and nucleotide substitution rates. *Mol Biol Evol*. 2014;31(3):645–59.
- Gurdon C, Maliga P. Two distinct plastid genome configurations and unprecedented intraspecific length variation in the *accD* coding region in *Medicago truncatula*. *DNA Res*. 2014;21(4):417–27.
- Cosner ME, Raubeson LA, Jansen RK. Chloroplast DNA rearrangements in Campanulaceae: phylogenetic utility of highly rearranged genomes. *BMC Evol Biol*. 2004;4(1):27.
- Knox EB. The dynamic history of plastid genomes in the Campanulaceae sensu lato is unique among angiosperms. *Proc Natl Acad Sci*. 2014;111(30):11097–102.
- Barnard-Kubow KB, McCoy MA, Galloway LF. Biparental chloroplast inheritance leads to rescue from cytonuclear incompatibility. *New Phytol*. 2017;213(3):1466–76.
- Postel Z, Poux C, Gallina S, Varré JS, Godé C, Schmitt E, et al. Reproductive isolation among lineages of *Silene nutans* (Caryophyllaceae): A potential involvement of plastid-nuclear incompatibilities. *Mol Phylogenet Evol*. 2022;1(169):107436.
- Cosner ME, Jansen RK, Moret BME, Raubeson LA, Wang LS, Warnow T, et al. An empirical comparison of phylogenetic methods on chloroplast

- gene order data in Campanulaceae. In: Sankoff D, Nadeau JH, editors. Comparative Genomics: Empirical and analytical approaches to gene order dynamics, map alignment and the evolution of gene families. Dordrecht: Springer Netherlands; 2000 [cited 2024 Aug 13]. p. 99–121. Available from: https://doi.org/10.1007/978-94-011-4309-7_11.
22. Cosner ME, Jansen RK, Palmer JD, Downie SR. The highly rearranged chloroplast genome of *Trachelium caeruleum* (Campanulaceae): multiple inversions, inverted repeat expansion and contraction, transposition, insertions/deletions, and several repeat families. *Curr Genet*. 1997;31(5):419–29.
 23. Haberle RC, Fourcade HM, Boore JL, Jansen RK. Extensive rearrangements in the chloroplast genome of *Trachelium caeruleum* are associated with repeats and tRNA genes. *J Mol Evol*. 2008;66(4):350–61.
 24. Li CJ, Wang RN, Li DZ. Comparative analysis of plastid genomes within the Campanulaceae and phylogenetic implications. *PLoS ONE*. 2020;15(5):e0233167.
 25. Barnard-Kubow KB, Debban CL, Galloway LF. Multiple glacial refugia lead to genetic structuring and the potential for reproductive isolation in a herbaceous plant. *Am J Bot*. 2015;102(11):1842–53.
 26. Barnard-Kubow KB, Galloway LF. Variation in reproductive isolation across a species range. *Ecol Evol*. 2017;7(22):9347–57.
 27. Barnard-Kubow KB, Sloan DB, Galloway LF. Correlation between sequence divergence and polymorphism reveals similar evolutionary mechanisms acting across multiple timescales in a rapidly evolving plastid genome. *BMC Evol Biol*. 2014;14(1):268.
 28. Mansion G, Parolly G, Crowl AA, Mavrodiev E, Cellinese N, Oganessian M, et al. How to Handle Speciose Clades? Mass Taxon-Sampling as a Strategy towards Illuminating the Natural History of Campanula (Campanulaceae). *PLoS ONE*. 2012;7(11):e50076.
 29. Li H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*. 2018;34(18):3094–100.
 30. Kolmogorov M, Yuan J, Lin Y, Pevzner PA. Assembly of long, error-prone reads using repeat graphs. *Nat Biotechnol*. 2019;37(5):540–6.
 31. Jin JJ, Yu WB, Yang JB, Song Y, dePamphilis CW, Yi TS, et al. GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. *Genome Biol*. 2020;21(1):241.
 32. Katoh K, Standley DM. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Mol Biol Evol*. 2013;30(4):772–80.
 33. Trifinopoulos J, Nguyen LT, von Haeseler A, Minh BQ. W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. *Nucleic Acids Res*. 2016;44(W1):W232–5.
 34. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods*. 2017;14(6):587–9.
 35. Díez Menéndez C, Poczar P, Williams B, Myllys L, Amirouche A. IRplus: An Augmented Tool to Detect Inverted Repeats in Plastid Genomes. *Genome Biol Evol*. 2023;15(10):evad177.
 36. Marçais G, Delcher AL, Phillippy AM, Coston R, Salzberg SL, Zimin A. MUMmer4: A fast and versatile genome alignment system. *PLOS Comput Biol*. 2018;14(1):e1005944.
 37. Goel M, Sun H, Jiao WB, Schneeberger K. SyRl: finding genomic rearrangements and local sequence differences from whole-genome assemblies. *Genome Biol*. 2019;20(1):277.
 38. Goel M, Schneeberger K. plotsr: visualizing structural similarities and rearrangements between multiple genomes. *Bioinformatics*. 2022;38(10):2922–6.
 39. Rozas J, Ferrer-Mata A, Sánchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE, et al. DnaSP 6: DNA Sequence Polymorphism Analysis of Large Data Sets. *Mol Biol Evol*. 2017;34(12):3299–302.
 40. Paradis E. pegas: an R package for population genetics with an integrated-modular approach. *Bioinformatics*. 2010;26(3):419–20.
 41. Benson G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res*. 1999;27(2):573–80.
 42. Doyle JJ, Doyle JL. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem Bull*. 1987;19(1):11–5.
 43. Sugiura M. RNA Editing in Chloroplasts. In: Göringer HU, editor. RNA Editing [Internet]. Berlin, Heidelberg: Springer; 2008 [cited 2024 Aug 28]. p. 123–42. Available from: https://doi.org/10.1007/978-3-540-73787-2_6.
 44. Sun SS, Pan ZY, Fu Y, Wang SJ, Fu PC. Rampant intraspecific variation of plastid genomes in *Gentiana* section Chondrophyllae. *Ecol Evol*. 2024;14(9):e70239.
 45. Silva SR, Pinheiro DG, Penha HA, Plachno BJ, Michael TP, Meer EJ, et al. Intraspecific variation within the *Utricularia amethystina* species morphotypes based on chloroplast genomes. *Int J Mol Sci*. 2019;20(24):6130.
 46. Choi IS, Jansen R, Ruhlman T. Caught in the Act: Variation in plastid genome inverted repeat expansion within and between populations of *Medicago minima*. *Ecol Evol*. 2020;10(21):12129–37.
 47. Zhang R, Liu Y, Liu S, Zhao Y, Xiang N, Gao X, et al. Comparative organelle genomics in Daphniphyllaceae reveal phylogenetic position and organelle structure evolution. *BMC Genomics*. 2025;26(1):40.
 48. Pérez-Pérez MA, Yu W. Pleistocene origin and colonization history of *Lobelia columnaris* Hook. f. (Campanulaceae: Lobelioideae) across sky islands of West Central Africa. *Ecol Evol*. 2021;11(22):15860–73.
 49. Charbonneau JLM, Cronn RC, Liston A, Wojciechowski MF, Sanderson MJ. Plastome Structural Evolution and Homoplastic Inversions in Neo-Astragalus (Fabaceae). *Genome Biol Evol*. 2021;13(10):evab215.
 50. Wang J, Zou Y, Mower JP, Reeve W, Wu Z, Wang J, et al. Rethinking the mutation hypotheses of plant organellar DNA. *Genomics Commun* [Internet]. 2024 Sep 30 [cited 2025 Feb 22];1(1). Available from: <https://www.maxapress.com/article/doi/https://doi.org/10.48130/gcomm-0024-0003>.
 51. Dobzhansky Th. Studies on hybrid sterility. II. Localization of sterility factors in *Drosophila pseudoobscura* hybrids. *Genetics*. 1936;21(2):113–35.
 52. Muller HJ. Isolating mechanisms, evolution, and temperature. *Biol Symp*. 1942;6:71.
 53. Pulido P, Zagari N, Manavski N, Gawronski P, Matthes A, Scharff LB, et al. CHLOROPLAST RIBOSOME ASSOCIATED supports translation under stress and interacts with the ribosomal 30S subunit1. *Plant Physiol*. 2018;177(4):1539–54.
 54. Kikuchi S, Bédard J, Hirano M, Hirabayashi Y, Oishi M, Imai M, et al. Uncovering the protein translocator at the chloroplast inner envelope membrane. *Science*. 2013;339(6119):571–4.
 55. Kikuchi S, Asakura Y, Imai M, Nakahira Y, Kotani Y, Hashiguchi Y, et al. A Ycf2-FtsHi heteromeric AAA-ATPase complex is required for chloroplast protein import. *Plant Cell*. 2018;30(11):2677–703.
 56. Gao LL, Hong ZH, Wang Y, Wu GZ. Chloroplast proteostasis: A story of birth, life, and death. *Plant Commun* [Internet]. 2023 Jan 9 [cited 2024 Aug 17];4(1). Available from: [https://www.cell.com/plant-communications/abstract/S2590-3462\(22\)00256-5](https://www.cell.com/plant-communications/abstract/S2590-3462(22)00256-5).

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.