RESEARCH

Open Access

Genomic variation and evolutionary patterns in organelle genomes between annual and perennial *Glycine* species

Xuchen Yang[1†], Jiaxian He[1†], Minghui Zhou[1†], Changwei Bi[2], Jiali Kong[3], Jie Wang[3], Fanjiang Kong[1], Zhiqiang Wu[3*], Zefu Wang[2*] and Meina Li[1*]

Abstract

Background The complexity of structural variations and long stretches of repetitive DNA make the analysis of plant mitochondrial genomes (mitogenomes) exceptionally challenging. A thorough investigation of plant mitogenomes is essential for uncovering the evolutionary processes of plant organelles and optimizing traits related to plant cellular metabolism. The genus *Glycine* includes groups with both perennial and annual life strategies, making it an ideal subject for studying the complexity and variations of plant mitogenomes during evolution across different life strategies.

Results Here, we assembled 20 complete mitochondrial and plastid genomes of *Glycine* accessions, including both annual and perennial species using the latest organelle genome assembly tool. Significant structural variations and differences in tRNA content were observed in the mitogenomes between the two life-history strategy subgenera, while protein-coding genes and rRNAs content were highly conserved. Distinct patterns of nuclear plastid DNAs and nuclear mitochondrial DNAs (NUPTs/NUMTs) were uncovered among annual and perennial species. Genes residing in NUMTs (NUMGs) showed a substantial presence in *Glycine* accessions, with annual soybeans exhibiting a higher proportion of protein-coding genes fully integrated into the nuclear genome. Phylogenetic analysis indicated a closely related evolutionary trajectory between mitochondrial and nuclear genomes in *Glycine*, providing supplementary evidence relevant to the evolutionary history of *Glycine*.

Conclusions This study showed the structural variations and evolutionary patterns of mitochondrial genomes between annual and perennial *Glycine* species. These findings contribute to our understanding of plant organelle complexity, variation and history of intracellular genomic integration.

†Xuchen Yang, Jiaxian He and Minghui Zhou contributed equally to this work.

*Correspondence:
Zhiqiang Wu
wuzhiqiang@caas.cn
Zefu Wang
wangzefu@njfu.edu.cn
Meina Li
limeina@gzhu.edu.cn

Full list of author information is available at the end of the article

Introduction

Mitochondria and plastids each carry their unique organellar genomes, named mitochondrial genome (mitogenome) and plastid genome (plastome), respectively. These genomes play an indispensable role in the evolution and function of plants, regulating key physiological processes such as respiration and photosynthesis, which drive plant biomass accumulation and are closely tied to crop yield [52]. Unlike bilaterian mitogenomes, most land plant mitogenomes exhibit extensive variability in structure, size, and gene set, making it a significant area of study in biology [4]. In contrast to the relatively conserved nature of bilaterian mitochondrial genomes, the size of plant mitogenomes can range from 66 kb to 12,000 kb [17, 52]. Although the mitogenomes of angiosperms are typically assembled into a "master circle" molecule [48], they display a diversity of genomic structures, including linear-mapping molecules (e.g., cytoplasmic male sterility-S line of maize) [2], bi-circular molecules (e.g., *Castanea mollissima*) [18], tri-circular molecules (e.g., *Populus simonii*) [7], and multichromosomal genomes (e.g., *Amborella*, *Sorghum*) [41, 62]. The genetic mechanisms underlying these multi-chromosomal structures remain unclear, as chromosome loss and recombination events make the inheritance and evolution of plant mitogenomes a persistent enigma. These features challenge traditional views on mitochondrial genetic stability and open new avenues for studying plant evolution and organellar genome function. The presence of recombination events and large numbers of repeat sequences further drives the high diversity of plant mitogenomes [37]. Additionally, plant mitogenomes are characterized by low gene density, a large number of introns, and RNA editing sites, all of which contribute to the complexity of gene function [12] The formation of these recombination and dynamic isomers plays a crucial role in maintaining genome stability and functionality [17]. The diversity and complexity of mitogenomes offer unique challenges and opportunities for scientific research. They deepen our understanding of plant cell evolution, genomic stability, and the molecular mechanisms of plant-environment interactions, providing new strategies and insights for crop improvement and genetic breeding, as seen in crops such as wheat [22], rice [21], maize [58], and sorghum [62].

The evolutionary paradox of plant mitochondria has been discovered for nearly 40 years [39]. Researchers have observed that plant mitochondrial DNA (mtDNA) sequences evolve slowly, but structural evolution is rapid [52]. Because of this, studies on plant mitochondria that focus on a single or a few species often yield limited insights. Thus, comparative studies between more closely related species are needed. In addition, previous studies of plant mitogenomes, most of them have focused on a single type of plant life cycle, such as annual or perennial plants [52]. In contrast, comparative studies within a genus that include both annual and perennial plants are less common. The characteristics of perennial plants allow them to survive in adverse environments, whereas annual plants cope with environmental pressures through rapid reproduction [16]. Therefore, understanding these regulatory mechanisms is not only significant for botanical research but also provides valuable insights for applications in agriculture and ecology, especially regarding how to enhance crop adaptability and yield through genetic regulation [35]. Legume *Glycine* comprises two subgenera: the subgenus *Soja* includes annual soybeans, such as cultivated soybean (*G. max*) and its wild ancestor *G. soja*, while the other subgenus encompasses over 30 perennial *Glycine* species [45]. Soybean, the most widely grown oil and protein seed crop in the world, exhibits low genetic diversity in its cultivated gene pool [44]. Perennial *Glycine* species serve as an expanded gene reservoir for improving annual crops, offering traits like higher seed pod number, resistance to nematodes and fungal pathogens, and tolerance to drought and salt stress [64]. Furthermore, in previous studies on the mitochondrial genomes of sorghum, it was found that although there are significant structural differences between the wild and cultivated sorghum, different cultivated varieties show a certain structural similarity to their corresponding wild varieties. The basic genomic framework remains similar, including highly conserved key genes, which helps us to deeply understand how the domestication process affects changes in plant genomes [62]. Hence, *Glycine* serves as an excellent model system for exploring the significance of genetic variations and uncovering the evolutionary mechanisms of plant organelle genomes, as well as their roles in ecological adaptation.

Traditional assembly of plant mitogenomes using second-generation sequencing data has often encountered challenges due to the complexity and repetitiveness of these sequences [8]. However, leveraging the power of third-generation sequencing and the PMAT toolkit, we are now able to delve deeper into these complex genomic structures. Thus, comprehensive studies of plant mitochondrial and plastid genomes are crucial for uncovering fundamental biological processes in plant cells and for enhancing crop productivity to meet global food security challenges. In recent years, many whole-genome studies of both annual and perennial *Glycine* species have yielded high-quality genome data [34, 64]. However, these resources have yet to be extensively applied to the

study of organelle genomes, particularly mitogenome, which are crucial for understanding various agronomic traits such as cytoplasmic male sterility [20], disease resistance [53], and cold tolerance [26], etc. Therefore, comprehensive mitogenome assembly and deep analysis across multiple soybean individuals remain incomplete.

Within the genus *Glycine*, there are limited mitochondrial genomic resources available for systematic study of evolutionary issues and mitogenome variation analysis, with only a few mitogenomes assembled based on next-generation sequencing data reported [11, 20, 33]. To enrich the mitogenome resources in the *Glycine* genus, we conducted complete assemblies of the mitochondrial and plastid genomes for 20 *Glycine* accessions, including 6 *G. max* cultivars, 5 *G. max* landraces, 3 *G. soja*, and 6 perennial *Glycine* species, using third-generation sequencing data. Our systematic comparative analysis revealed structural variations and evolutionary changes in these genomes. Furthermore, we quantified and characterized nuclear plastid DNAs and nuclear mitochondrial DNAs (NUPTs/NUMTs), as well as genes residing in NUMTs (NUMGs). Additionally, we presented the phylogenetic relationships based on conserved genes from both organelles. This study not only enhances our understanding of mitogenome structure and evolution in *Glycine* but also lays a solid foundation for future research aimed at improving agronomic traits, increasing yield, and enhancing cultivar improvement through organelle studies.

Methods

Data acquisition

The Pacbio data for 14 annual soybeans was obtained from the China National Center for Bioinformation (CNCB) with project number PRJCA002030, while the Pacbio data for 6 perennial *Glycine* species was obtained from National Center for Biotechnology Information (NCBI) with project number PRJNA503746. The mitochondrial and plastid genomes of *Vigna angularis*, serving as the outgroup, were retrieved from NCBI with the accession numbers NC_021092.1 and NC_021091.1, respectively. Additional details can be found in Table S1.

Organelle genome assembly and annotation of 20 Glycine accessions

We utilized high-quality sequencing data (PacBio) to perform complete assemblies of the mitochondrial and plastid genomes for 20 *Glycine* accessions. Due to the systematic error rate present in the raw PacBio raw reads, we performed Canu v2.2 to correct the raw sequencing data prior to the initial assembly with default parameter "-correct genomeSize = maxThreads=" [46].

Subsequently, the error-corrected long reads were utilized as input for the PMAT v1.5.2 assembler to de novo assemble the organelle genomes [8]. Following the draft assembly, manual polishing of the organelle genome was performed using Bandage v0.8.1 based on read depths of the connected contigs [57]. After manually removing certain false links and forks, we obtained premier contigs for the assembly of the organelle genomes.

Genome annotation and codon preference analysis

Mitochondrial protein-coding genes (PCGs), tRNA, and rRNA were annotated using the PMGA online tool (http://www.1kmpg.cn/pmga/) [32]. Meanwhile, plastid PCGs, tRNA, and rRNA were annotated with the CPGAVAS2 web-based platform (http://47.96.249.172:16019/analyzer/home) [47].

Protein-coding sequences were extracted using PhyloSuite v1.2.3 with default parameters [60]. Codon usage bias was assessed and the relative synonymous codon usage (RSCU) value was calculated for the mitogenome's PCGs using MEGA v7.0 [50]. An RSCU value of 1 indicates no codon preference, while values greater than 1 suggest a preference for the codon, and values less than 1 indicate a relative disuse.

Analysis of intracellular gene transfer

We employed Blastn v2.14.0+ [9] for identifying transfer events from organelles to the nuclear genomes with the following screening criteria: matching rate ≥ 80%, E-value ≤ 1e – 6, and length ≥ 100 bp. The results were subsequently categorized into two distinct datasets: one with 80–89% similarity and the other with 90–100% similarity to a recognized organelle sequence. These datasets were used to signify transfers that are older (exhibiting more mutations) and those that are newer (with fewer mutations), respectively. Furthermore, each dataset was subdivided into distinct length categories: 100–199 bp, 200–399 bp, 400–599 bp, 600–799 bp, 800–999 bp, and 1000 bp or longer. Genes derived from organelles were annotated based on the organelle genome annotations.

Phylogenetic analysis

The mitochondrial coding sequences (CDS) from 20 *Glycine* accessions and outgroup *Vigna angularis* were identified using PhyloSuite v1.2.3 [60], with standard settings. Subsequently, these sequences were aligned using MAFFT v7.525 [25], applying the default parameters. The aligned sequences were then refined by employing trimAl v7.525 [10]. Finally, the phylogenetic trees were constructed from the concatenated CDS using IQ-TREE v2.2.0.3 [38] with 5000 bootstrap replicate. The trees were visualized and annotated by online tool tvBOT [59].

Results

Glycine organellar genomes

To achieve a more comprehensive understanding of the organelle genomes in *Glycine*, we assembled a total of 20 organelle genomes from *Glycine* accessions, including both perennial and annual species. In the case of the mitogenomes, the assembled genomes exhibited a notable size variation, with total lengths ranging from 318,377 bp to 409,430 bp (Table S2). This variation underscores the dynamic nature of mitogenomes. The GC content across the 20 mitogenomes displayed a relatively narrow range, fluctuating between 44.8% and 45.5% (Table S2). The graphic genome revealed that the structure of perennials significantly differs from that of annual beans. The mitogenome of perennials (*G. cyrtoloba*, *G. dolichocarpa*, *G. falcata*, *G. stenophita*, *G. syndetica* and *G. tomentella D3*) typically consists of a simple circular structure (Fig. 1A), whereas the mitogenome of annual beans exhibits a highly complex architecture (Fig. 1B, C, D). Within the category of annual soybeans, there is also a noticeable structural difference between wild and cultivated soybeans. The mitogenome of wild soybeans is characterized by a complex large ring (Fig. 1B), while that of cultivated soybeans is composed of a complex large ring plus a simple ring, with the cultivar shown in Fig. 1C and the landrace in Fig. 1D.

The plastomes exhibit remarkable consistency in both structure, total length and GC content. With total lengths ranging from 151,683 bp to 153,047 bp and GC content varying from 35.2 to 35.4% across the 20 plastomes, these observations indicate the highly conserved nature of the plastome (Table S3).

Conservation of gene content and codon preference

Although *G. soja*, *G. max* and perennial *Glycine* species show variations in mitogenome size and structure, there is minimal divergence in their protein-coding genes. All 20 mitogenomes possess a set of 31 conserved PCGs, including components of the respiratory chain complexes: Complex I (NADH Dehydrogenase subunits: *nad1*, *nad2*, *nad3*, *nad4*, *nad4L*, *nad5*, *nad6*, *nad7*, *nad9*), Complex II (Succinate Dehydrogenase subunit: *sdh4*), Complex III (Cytochrome bc1 Complex subunit: *cob*), Complex IV (Cytochrome c Oxidase subunits: *cox1*, *cox3*), ATP Synthase subunits (*atp1*, *atp4*, *atp6*, *atp8*, *atp9*); cytochrome c maturation proteins (*ccmB*, *ccmC*, *ccmFC*, *ccmFN*); ribosomal subunits: the small subunit (*rps1*, *rps3*, *rps4*, *rps10*, *rps12*) and the large subunit (*rpl5*, *rpl16*); the transport membrane protein (*mttB*); maturase R (*matR*)(Fig. 2A). However, two perennial species (*Gcy* and *Gst*) lack the *cox2* gene, and eight annual soybeans (SoyW01, SoyW03, SoyL01, SoyL04, SoyL07, SoyC05, SoyC10) lack the *nad4L* gene. For rRNA genes, all 20 mitogenomes contain *rrn5S*, *rrn18S*, and *rrn26S*.

For tRNA genes, annual soybeans exhibit a high level of conservation, possessing 16 identical tRNA genes, with only minor variations in copy number across different species. However, perennial *Glycine* species display significant variation in tRNA gene content compared with annual soybeans, with some tRNA genes being present only in the perennial species. Additional tRNA genes found in perennial *Glycine* species include *trnL* in *GtoD3*; *trnR* in *Gsy*; *trnT* in *Gcy*, *Gdo*, *Gst*, *Gsy*, and *GtoD3*; and *trnV* in *Gdo*, *Gfa*, *Gsy*, and *GtoD3* (Fig. 2A and Table S2).

The plastomes of all 20 *Glycine* accessions exhibit remarkable consistency in PCG and rRNA, with each accessions containing 75 PCGs and 4 rRNA genes (Fig. 2B and Table S3). For tRNA genes, while all accessions share 26 tRNA in common, *trnV* is exclusively present in perennial species *Gdo*, *Gfa*, *Gsy*, and *GtoD3* (Fig. 2C and Table S3).

We conducted an analysis of amino acid and codon usage in mitochondrial PCGs for both perennial and annual *Glycine* species. The results show that both types of *Glycine* species contain all 20 amino acids, with similar usage frequencies (Fig. 3A, Table S4). However, perennial *Glycine* species exhibit a slightly higher overall amino acid count compared with annual soybeans (Fig. 3B, Table S5). Among the amino acids, leucine (Leu) is the most frequently occurring (10.67–10.92%), followed by serine (Ser) (8.95–9.18%), while cysteine (Cys) (1.41–1.61%) and tryptophan (Trp) (1.42–1.59%) are the least frequent (Fig. 3). The trends in RSCU are consistent across all *Glycine* accessions, with 29 out of 61 codons showing high codon preference (RSCU > 1) and 30 codons showing low codon preference (RSCU < 1). The codons AUG and UGG exhibit no codon preference (RSCU = 1). Except for methionine (Met) and tryptophan (Trp), all amino acids have two or more type of codons (Fig. 3C, Table S6).

The landscapes of NUPTs/NUMTs in the Glycine species

Intracellular sequence transfer events among organellar genomes and nuclear genomes happen frequently and continuously, playing a crucial role in understanding the coevolution of genomes within a cell [60]. To explore the distribution of NUPTs/NUMTs within the *Glycine* species, we conducted BLAST analyses against the nuclear and plastid genomes, using mitogenomes as queries for all 20 individuals. The results revealed distinct patterns in the distribution of NUMTs and NUPTs among perennial and annual *Glycine* species (Fig. 4A). In the case of NUMTs, perennial *Glycine* species exhibit a notably higher number of fragments compared with annual soybeans, but the fragments are noticeably shorter. As for NUPTs, perennial *Glycine* species also have a significantly higher number of fragments. Regarding fragment length, apart from *Gfa* which has notably longer
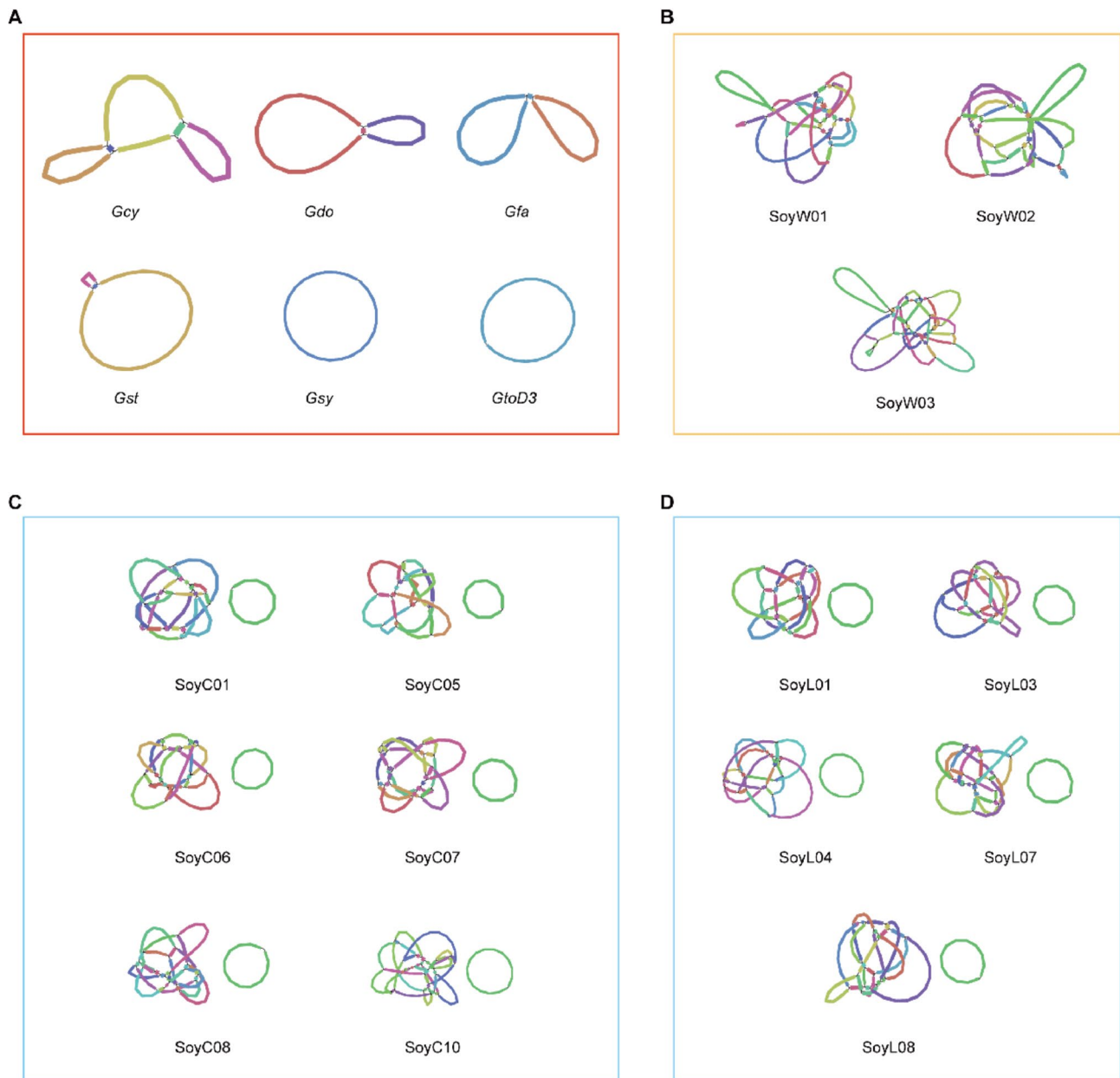
Fig. 1 Mitogenome structure of 20 *Glycine* accessions. **A**, The mitogenome structure of 6 perennial *Glycine* species: Gcy, *Glycine cyrtoloba*; Gdo, *Glycine dolichocarpa*; Gfa, *Glycine falcata*; Gst, *Glycine stenophita*; Gsy, *Glycine syndetika*; GtoD3, *Glycine tomentella D3* (marked with a red box). **B**, The mitogenome structure of 3 annual wild soybeans (*Glycine soja*): SoyW01, SoyW02, SoyW03 (marked with a yellow box). **C**, The mitogenome structure of 6 annual cultivar soybeans (*Glycine max*): SoyC01, SoyC05, SoyC06, SoyC07, SoyC08, SoyC10 (marked with a blue box). **D**, The mitogenome structure of 5 annual landrace soybeans (*Glycine max*): SoyL01, SoyL03, SoyL04, SoyL07, SoyL08 (marked with a blue box). The mitogenomes were visualized using Bandage, with the distinct colors being randomly assigned by the software

fragments, there is no significant difference among the other accessions (Fig. 4A).

Based on the similarity of the transferred fragments, we categorize them into two types: 80–90% similarity (relatively ancient) and 90–100% similarity (relatively recent). The majority of NUMTs and NUPTs are of the recent type. Among these, all NUMTs and recent NUPTs are predominantly concentrated within the 0–200 bp range, whereas ancient NUPTs primarily range from 0 to 400 bp in length (Fig. 4B, C). Recent NUMTs and NUPTs exhibit a marked predominance of short fragments in perennial *Glycine* species compared with annual legumes (Fig. 4B). This distinction is not observed in the ancient NUMTs and NUPTs (Fig. 4C).
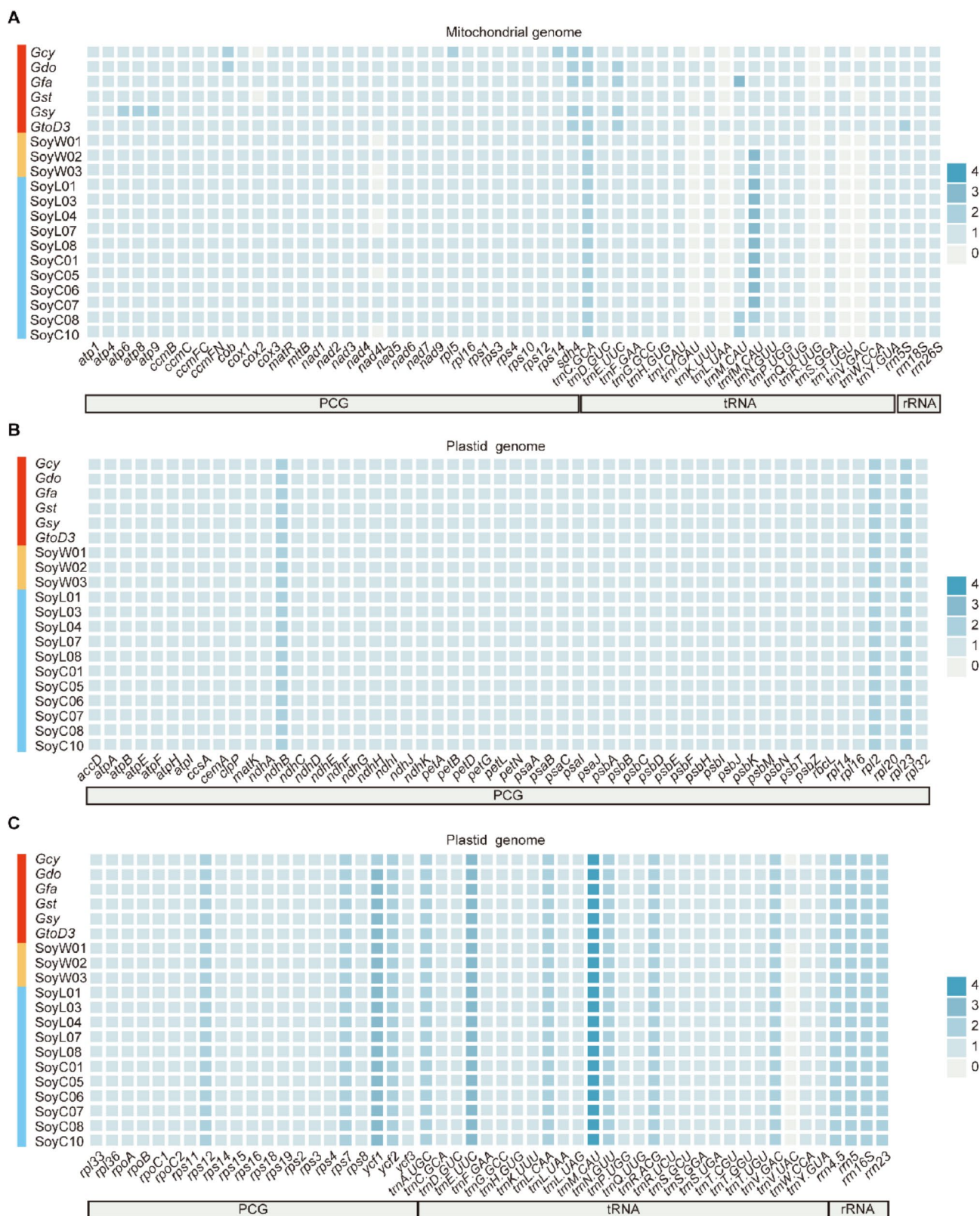
Fig. 2 Gene content of 20 *Glycine* organelle genomes. **A**, Gene content of 20 *Glycine* mitogenomes. **B** and **C**, Gene content of 20 *Glycine* plastomes. Different colors represent the copy number of genes present in each genome
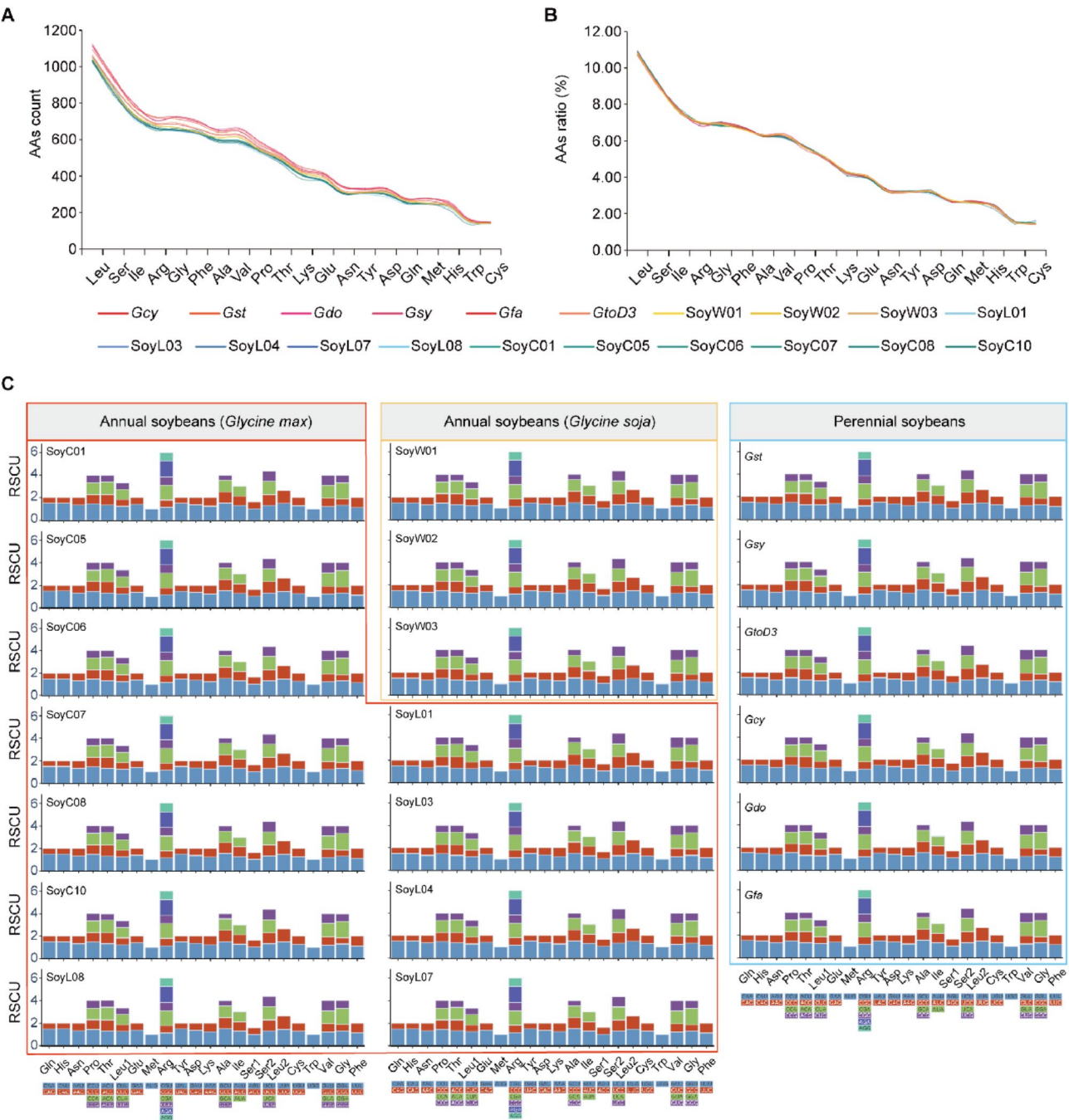
Fig. 3 Codon preference of 20 *Glycine* mitogenomes. **A**, The total count of each amino acid residue in all mitochondrial proteins is shown on the y-axis. **B**, The relative percentage of each amino acid residue in all mitochondrial proteins is shown on the y-axis. **C**, Relative synonymous codon usage (RSCU) in 20 *Glycine* mitogenomes (*Glycine max* in red frame, *Glycine soja* in yellow frame, perennial *Glycine* species in blue frame)

Intracellular gene transfer from organelles to nuclear

Based on gene annotation information and NUMTs information, we examined the distribution of NUMGs across all *Glycine* accessions. Our findings reveal a substantial presence of NUMGs in *Glycine* species (Fig. 5A). Specifically, for PCGs and rRNAs within NUMGs, annual soybeans showed a high proportion of genes fully integrated into the nuclear genome, whereas perennial species showed a higher incidence of partial gene transfer (Fig. 5B, C). Conversely, no notable differences were observed between annual and perennial *Glycine* species regarding tRNA within NUMGs. Further investigation identified three tRNAs (trn-L, trnR, and trnS) with no transfer events across all individuals, while two tRNAs (trnV and trnW) displayed complete transfer events in all individuals (Fig. 5B). These results collectively highlight
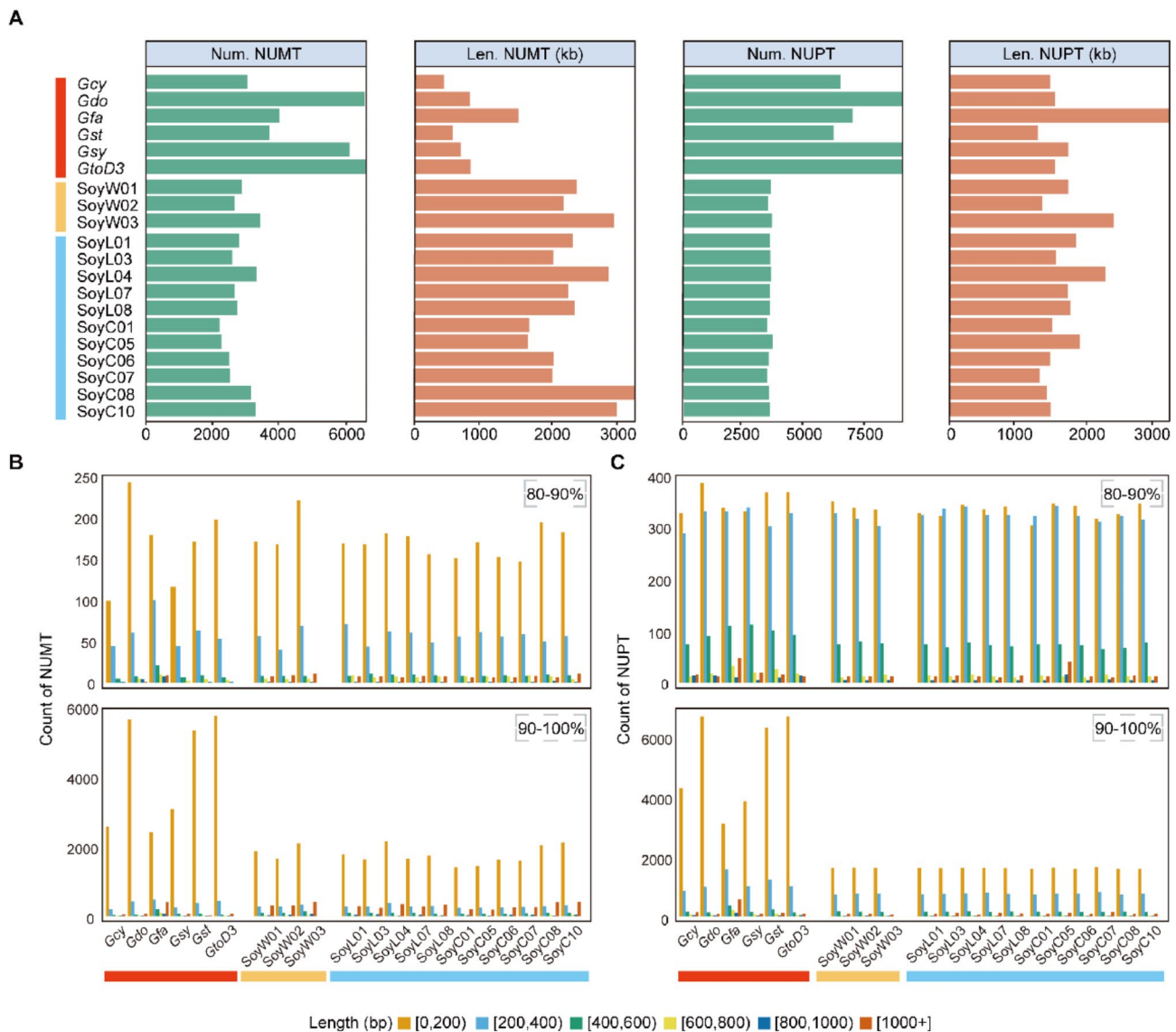
Fig. 4 Distribution of inserts by size and type of *Glycine* organelle sequences transferred to the nucleus. **A**, Total number and total length of NUMTs (left two panels) and NUPTs (right two panels) are demonstrated by accession. **B**, Number and length of NUMTs are demonstrated by accession, with panels separated by percent similarity to origin. **C**, Number and length of NUPTs are demonstrated by accession, with panels separated by percent similarity to origin

distinct patterns of NUMGs distribution between annual and perennial *Glycine* species, underscoring their evolutionary divergence. Additionally, in the gene annotation analyses, we identified a few individuals lacking the cox2 (*Gcy* and *Gst*) or nad4L (SoyW01, SoyW03, SoyL01, SoyL04, SoyL07, SoyC05, SoyC10) genes in their mitogenomes, with both genes detected fully transferred into the nuclear genome.

Phylogenetic analysis

To determine the phylogenetic relationships among organelles, we constructed two maximum likelihood trees using conserved PCGs from the mitogenomes

and plastomes, respectively. In both phylogenetic trees, perennial *Glycine* species consistently occupy the basal positions (Fig. 6). However, discrepancies are observed among the annual soybeans. The mitochondrial PCGs-based phylogeny corroborates previous findings based on nuclear genomes [64](Fig. 6A), positioning wild soybeans at the base of the annual soybean clade. Differently, the plastid PCGs-based phylogeny shows that the wild soybean accessions (SoyW01, SoyW02 and SoyW03) cluster with the cultivated soybeans (Fig. 6B).

These differences may stem from the varying evolutionary rates and genetic variations of mitochondria and plastid. Mitochondria and plastid possess distinct
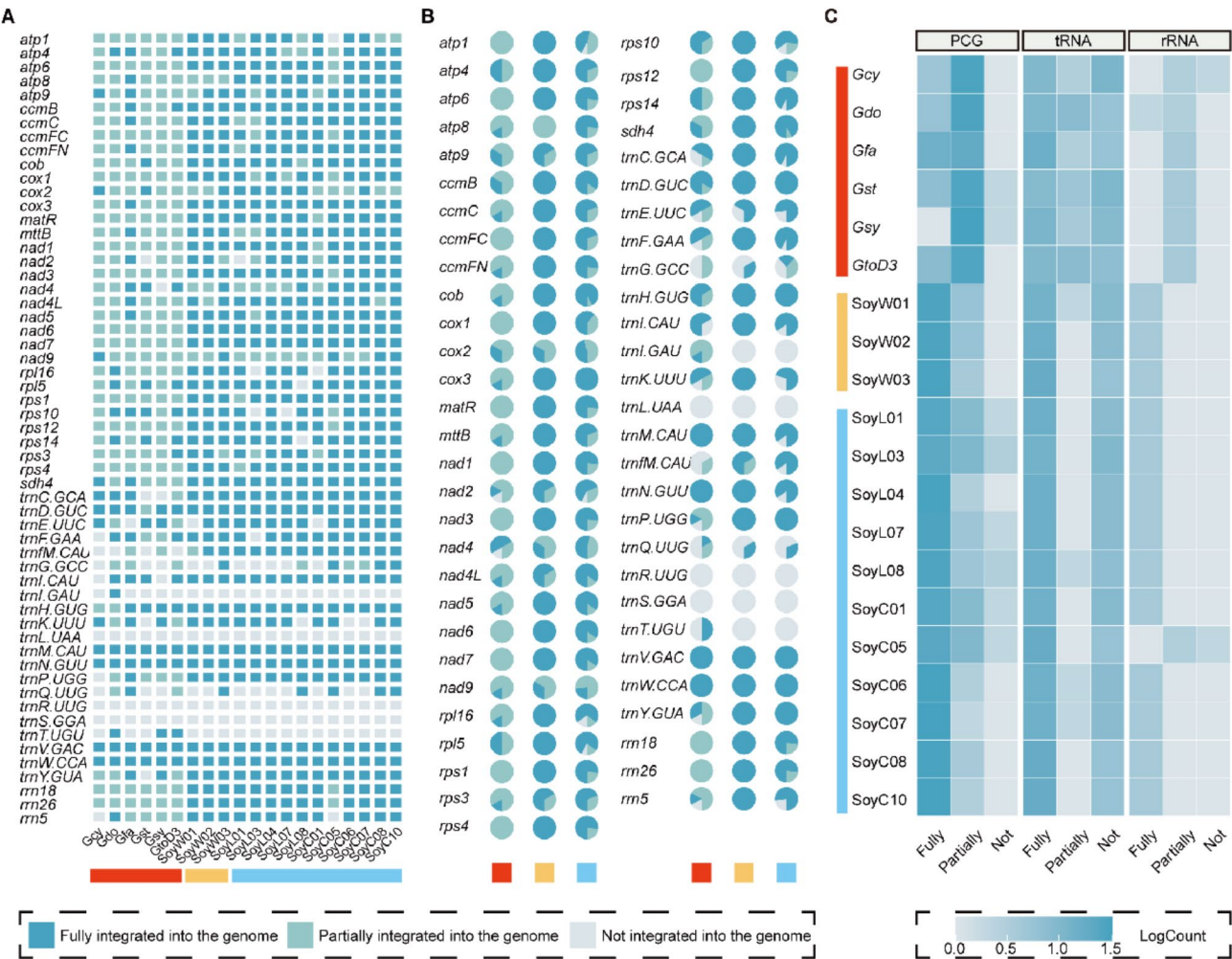
Fig. 5 Gene transfer events in *Glycine* mitogenomes. **A**, Overview of NUMGs among 20 *Glycine* accessions. **B**, Proportion of genes that are fully, partially, and not integrated into nuclear genome for each gene. **C**, Heatmap of genes that are fully, partially, and not integrated into nuclear genome in PCGs, tRNAs, and rRNAs across 20 *Glycine* accessions. In (**A**) and (**B**), dark blue, light blue, and gray represent fully, partially, and not integrated into nuclear genome, respectively. In (**C**), the scale represents the log-transformed number of genes with fully, partially, and not integrated into nuclear genome

genetic features and evolutionary histories, potentially influenced by different selection pressures and environmental factors. Additionally, evolutionary processes such as gene transfer and recombination could impact the phylogenies of mitochondria and plastid. Furthermore, the phylogenetic relationships of leguminous plants may also be influenced by other factors such as gene flow, hybridization, and polyploidization events. These factors could lead to genome mixing and recombination, thereby affecting the topology of phylogenetic trees.

Discussion
Complex structures in the annual soybean mitogenomes
The plant mitogenome was initially described as a simplified circular molecule, but advancements in research reveal a far more complex in mitogenome structure across various plants [48, 54]. Based on previous studies of soybean mitochondria and our findings, the

mitochondrial structure of soybean is highly complex among individuals.

Early studies on structure of soybean mitochondria, reliant on electron microscopy, highlighted the supercoiled form of mtDNA and proposed dispersion among circular molecules of varying sizes [36, 51]. Biochemical methods further detected recombination activity in soybean mitochondria, including ATP and Mg^{2+}-dependent strand invasion, suggesting complex DNA recombination in soybean [36].

Advancements in sequencing technologies, notably second-generation sequencing, enabled deeper exploration of soybean's mitogenome structure. Recent studies confirmed multiple sub-genomic circles in soybean through assembly, aligning with electron microscopy findings [11, 33]. Furthermore, large repetitive sequences in soybeans may mediate the formation of multiple
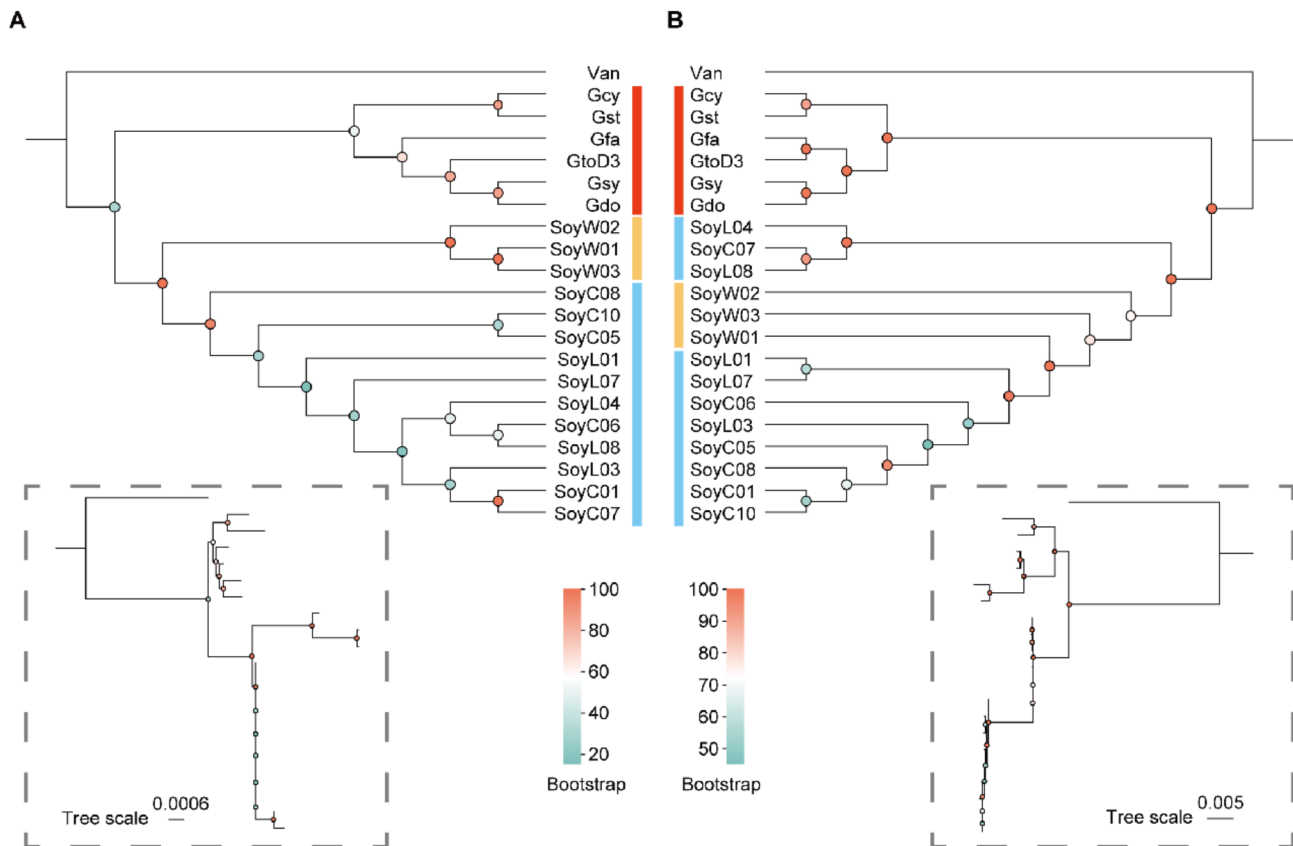
Fig. 6 Phylogenetic relationships of 20 *Glycine accessions based on* organelle genomes. **A**, Maximum likelihood phylogenetic tree based on PCGs of mitogenomes. **B**, Maximum likelihood phylogenetic tree based on PCGs of plastomes

circular structures, a theory that has also been verified in other plants [15, 23].

Third-generation sequencing technologies have provided deeper insights into the structure of the soybean mitogenome. From the 20 *Glycine* accessions' mitogenome assemblies, we implied that perennials typically feature a simple circular structure, whereas annual soybean exhibits a highly complex reticular structure. This includes a complex large reticular circle in wild soybean and a combination of a complex large reticular circle and a simple circle in cultivated soybean. Notably, the structural complexity of plant mitochondrial genomes is often closely associated with the presence of repetitive sequences [4]. Repetitive sequences are not only the primary drivers of genomic rearrangements but can also lead to the formation of various subgenomic circular or linear molecules [52]. For instance, in the *Glycine* genus, frequent recombination of repetitive sequences may have contributed to the formation of reticular structures, similar to the multichromosomal mitochondrial genomes observed in species such as *Cucumis sativus* and *Silene* [3, 48]. Additionally, asymmetric recombination of repetitive sequences may also influence the mechanisms of

genome replication and inheritance, further increasing genomic structural diversity [37], Davila et al. 2011).

These insights shed new light on the structural evolution of plant mitogenomes and underscore the critical role of repetitive sequences in shaping genomic complexity. The complex reticular structures observed in the *Glycine* genus may be directly linked to the high-frequency recombination of repetitive sequences within their genomes. Such recombination events could drive dynamic structural changes by generating subgenomic molecules or facilitating chromosomal separation. Consequently, future research should prioritize investigating the functional roles, recombination mechanisms, and broader impacts of repetitive sequences in the *Glycine* mitogenomes, particularly their influence on genomic stability and evolutionary potential. This focus will deepen our understanding of the structural evolutionary patterns governing plant mitogenomes.

Gene content variation among Glycine

In the Fabaceae family, mitogenomes exhibit considerable gene content variation, including multiple independent losses of mitochondrial genes [13]. These losses are influenced by various evolutionary mechanisms, such

as intracellular gene transfer and horizontal gene transfer [13]. Previous studies indicated that four ribosomal protein genes (*rps2*, *rps7*, *rps11*, *rps13*) were lost in the common ancestor of all Fabaceae species, which aligns with our findings (Fig. 2). Our study demonstrated that the *cox2* gene is lost in two perennial *Glycine* species (*Gcy* and *Gst*) but exists in all other individuals examined (Fig. 2). This pattern of loss is consistent with previous observations that *cox2* can be lost in specific lineages within the Fabaceae family [13]. In contrast to previous studies that found the *nad4L* gene present in all Fabaceae individuals they examined, our results revealed that the *nad4L* gene is lost in several annual soybeans, including SoyW01, SoyW03, SoyL01, SoyL04, SoyL07, SoyC05, SoyC10 (Fig. 2). This discrepancy highlights the dynamic nature of gene content in mitogenomes, even within closely related species.

Cox2 and *nad4L* are both essential for the proper functioning of the mitochondrial electron transport chain and are therefore critical for cellular energy production. Even if we find that in some accessions, there are deletions in the mitogenome of either cox2 or nad4L, we find that these genes are fully integrated into the nuclear genome. This suggests that the loss of these mitochondrial genes may be compensated by their nuclear counterparts, ensuring that their functional roles are maintained within the cell. The integration of mitochondrial genes into the nuclear genome is a well-documented phenomenon, often resulting in the transfer of functional genes to the nuclear genome where they can continue to be expressed and contribute to cellular processes [1, 24]. Overall, our findings on gene loss in *Glycine* species provide valuable insights into the evolutionary dynamics of mitogenomes in the Fabaceae family. The observed patterns of gene loss and nuclear integration underscore the complex interplay between mitochondrial and nuclear genomes and highlight the importance of studying these processes to understand plant genome evolution comprehensively.

The tRNA system in plant mitochondria exhibits a high level of complexity, primarily characterized by gene loss, replacement, and the ability to import tRNAs from other sources [7, 56]. During evolution, plant mitochondria have lost some tRNA genes. To maintain functionality, these lost genes are replaced by importing nuclear-encoded tRNAs from the cytoplasm or through horizontal gene transfer. This replacement and import mechanism highlights the flexibility of the mitochondrial tRNA system, ensuring that mitochondria can adapt to various environmental changes [56]. The tRNA import mechanism involves complex bioenergetic processes that selectively transport tRNAs into the mitochondria, allowing them to maintain their normal functions [6].

There are significant differences in mitochondrial tRNA import capabilities among different plant species, suggesting that this ability may have been independently acquired during evolution. These differences provide important clues for understanding the mitochondrial tRNA import mechanisms [29]. Furthermore, in some plants, such as *Silene*, mitochondrial tRNA genes have experienced rapid loss and replacement, with nuclear-encoded tRNAs substituting the original mitochondrial tRNAs [55]. This indicates that nuclear-encoded tRNAs can quickly become functional even in systems where gene loss has occurred recently [55]. Additionally, studies have shown that voltage-dependent anion channels (VDACs) play a crucial role in the tRNA import process in plant mitochondria, revealing the complexity of the tRNA import pathway [43]. In this study, we found that the tRNA system of annual soybean exhibits a high degree of conservation, with 16 identical tRNA genes and only minor differences in copy number among species (Fig. 2A). However, in perennial species, both the types and copy numbers of tRNAs show significant variation and certain tRNA genes are found only in perennial species (Fig. 2A). These demonstrate the remarkable evolutionary flexibility of tRNAs and their ability to cope with gene loss and functional replacement through various mechanisms.

Recent studies have further illuminated the evolutionary context of tRNA loss and compensatory mechanisms. The demands of photosynthesis are considered a key factor in maintaining the high conservation of the chloroplast tRNA system [5, 14]. In photosynthetic plants, chloroplasts require high levels of translational activity to support the rapid turnover of photosynthetic proteins, which may impose strong functional constraints on tRNAs and their associated enzymes [14]. In contrast, the translational demands of mitochondria are relatively lower, leading to greater flexibility in the evolution of the mitochondrial tRNA system [56]. Particularly in non-photosynthetic parasitic plants, the loss of mitochondrial tRNA genes is more pronounced, and in some cases, mitochondria rely entirely on the import of nuclear-encoded tRNAs [5]. This divergence suggests that the demands of photosynthesis may indirectly influence the evolutionary trajectory of the mitochondrial tRNA system, especially in dual-targeted systems where chloroplasts and mitochondria share the same tRNA processing enzymes [40].

Moreover, research has shown that the loss of mitochondrial tRNA genes in parasitic plants is often accompanied by the retargeting of nuclear-encoded tRNAs, where cytosolic tRNAs are imported into mitochondria to replace the lost mitochondrial tRNAs [14]. This compensatory mechanism is particularly notable in parasitic plants, especially in species that have completely lost photosynthesis, where the reconstruction of the mitochondrial tRNA system may be more radical [14].

For example, some parasitic plants have completely lost mitochondrial tRNA genes but still maintain mitochondrial translation through the import of nuclear-encoded tRNAs [14]. This evolutionary trend suggests that the mitochondrial tRNA system is highly plastic in responding to gene loss, capable of achieving functional compensation through various mechanisms. However, the mechanisms underlying gene variation among different strategies' *Glycine* were not explored in depth, which could provide further context for understanding the observed variations.

The role of nupts and NUMTs in shaping genomic diversity and evolution in Glycine species

The integration of NUPTs and NUMTs into the nuclear genome plays a significant role in the evolution and structural complexity of plant genomes [27, 30, 49]. NUMTs are formed through the transfer of mitochondrial DNA into the nuclear genome, often facilitated by mechanisms such as double-strand breaks [31]. There is a positive correlation between the abundance of NUMTs and genome size, with larger genomes typically harboring more NUMTs [19]. Intracellular transfer events occur frequently and continuously, predominantly integrating into genomic regions abundant in transposable elements, thereby contributing to the variation and structural evolution of the nuclear genome [61].

Following their transfer, NUMTs/NUMGs are typically inactivated by epigenetic mechanisms, such as high levels of DNA methylation, to prevent interference with normal nuclear gene functions [63]. Nevertheless, the accumulation of NUMTs is closely associated with speciation, particularly during plant polyploidization, where they play a significant role in genome remodeling and species differentiation, providing new insights into plant genome evolution and speciation [42]. Additionally, the distribution of NUMTs and NUPTs in the nuclear genome is non-random, predominantly located in gene-rich regions, while being sparse or absent near chromosome centromeres [28].

Recent studies, such as those conducted on the *Triticum/Aegilops* complex, provide insight into the dynamic evolutionary fate of NUPTs and NUMTs, showing that NUPTs and NUMTs are often associated with sequence variations and contribute to genomic diversity and adaptation [63]. Our research on *Glycine* species corroborates these findings, revealing that the transfer patterns of NUPTs and NUMTs differ significantly between annual and perennial *Glycine* species. Perennial species exhibit a more fragmented pattern, characterized by a greater number of shorter sequences, in contrast to annual species. This divergence could be attributed to the perennial's requirement for swift adaptation to domestication pressures, suggesting that such transfer patterns may play a role in shaping their unique evolutionary paths.

In summary, the integration of NUPTs and NUMTs into the nuclear genome has significant implications for the structural and functional evolution of the soybean mitogenome. Our findings underscore the importance of these transfer events in shaping the genomic landscape of *Glycine* species. Despite the insights gained, our study has limitations that should be acknowledged. First, the focus on *Glycine* species may not fully capture the broader implications of NUPTs and NUMTs across diverse plant taxa. Future research could benefit from a comparative analysis involving a wider range of species to better understand the evolutionary dynamics at play. Also, the functional implications of these transferred sequences remain largely unexplored, warranting further investigation into their roles in gene regulation and adaptation in different environmental contexts.

Abbreviations

CDS	Coding sequences
Gcy	*Glycine cyrtoloba*
Gdo	*Glycine dolichocarpa*
Gfa	*Glycine falcata*
Gst	*Glycine stenophita*
Gsy	*Glycine syndetika*
GtoD3	*Glycine tomentella D3*
Mitogenome	Mitochondrial genome
mtDNA	Mitochondrial DNA
NUMGs	Genes residing in NUMTs
NUMT	Nuclear mitochondrial DNA
NUPT	Nuclear plastid DNA
PCG	Protein-coding gene
Plastome	Plastid genome
rRNA	Ribosomal RNA
RSCU	Relative synonymous codon usage
tRNA	Transfer RNA

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12870-025-06312-4.

> Supplementary Material 1

Author contributions
XCY: Data curation, Formal analysis, Resources, Writing - original draft, Writing - review & editing. JXH: Data curation, Formal analysis, Writing - review & editing. MHZ: Data curation, Writing - review & editing. CWB: Writing - review & editing, Validation. JLK: Supervision, Writing - review & editing. JW: Supervision, Writing - review & editing. FJK: Project administration, Supervision, Writing - review & editing. ZQW: Conceptualization, Project administration, Writing - review & editing. ZFW: Conceptualization, Project administration, Writing - review & editing. MNL: Conceptualization, Project administration, Writing - review & editing. All authors read and approved the final manuscript.

Yang *et al. BMC Plant Biology*   (2025) 25:353

Page 13 of 14

Data availability

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
¹Guangdong Provincial Key Laboratory of Plant Adaptation and Molecular Design, Innovative Center of Molecular Genetics and Evolution, School of Life Sciences, Guangzhou University, Guangzhou 510006, China
²State Key Laboratory of Tree Genetics and Breeding, Co-Innovation Center for Sustainable Forestry in Southern China, College of Ecology and Environment, Nanjing Forestry University, Nanjing 210037, China
³Shenzhen Branch, Guangdong Laboratory of Lingnan Modern Agriculture, Key Laboratory of Synthetic Biology, Ministry of Agriculture and Rural Affairs, Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen 518120, China

References

1. Adams KL, Qiu YL, Stoutemyer M et al. Punctuated evolution of mitochondrial gene content: High and variable rates of mitochondrial gene loss and transfer to the nucleus during angiosperm evolution. Proc. Natl. Acad. Sci. U.S.A. 2002;99(15): 9905–9912.
2. Allen JO, Fauron CM, Minx P, et al. Comparisons among two fertile and three male-sterile mitochondrial genomes of maize. Genetics. 2007;177(2):1173–92.
3. Alverson AJ, Rice DW, Dickinson S, et al. Origins and recombination of the bacterial-sized multichromosomal mitochondrial genome of cucumber. Plant Cell. 2011;23(7):2499–513.
4. Backert S, Nielsen BL, Börner T. The mystery of the rings: structure and replication of mitochondrial genomes from higher plants. Trends Plant Sci. 1997;2(12):477–83.
5. Berrissou C, Cognat V, Koechler S, et al. Extensive import of nucleus-encoded tRNAs into chloroplasts of the photosynthetic lycophyte, Selaginella kraussiana. Proc Natl Acad Sci U S A. 2024;121(46):e2412221121.
6. Bhattacharyya SN, Adhya S. The complexity of mitochondrial tRNA import. RNA Biol. 2004;1(2):84–8.
7. Bi C, Qu Y, Hou J, et al. Deciphering the multi-chromosomal mitochondrial genome of *Populus simonii*. Front Plant Sci. 2022;13:914635.
8. Bi C, Shen F, Han F, et al. PMAT: an efficient plant mitogenome assembly toolkit using low-coverage HiFi sequencing data. Hortic Res. 2024;11(3):uhae023.
9. Camacho C, Coulouris G, Avagyan V, et al. BLAST+: architecture and applications. BMC Bioinf. 2009;10(1):421.
10. Capella-Gutierrez S, Silla-Martinez JM, Gabaldon T. TrimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics. 2009;25(15):1972–3.
11. Chang S, Wang Y, Lu J, et al. The mitochondrial genome of soybean reveals complex genome structures and gene evolution at intercellular and phylogenetic levels. PLoS ONE. 2013;8(2):e56502.
12. Chevigny N, Schatz-Daas D, Lotfi F, et al. DNA repair and the stability of the plant mitochondrial genome. Int J Mol Sci. 2020;21(1):328.
13. Choi IS, Schwarz EN, Ruhlman TA, et al. Fluctuations in Fabaceae mitochondrial genome size and content are both ancient and recent. BMC Plant Biol. 2019;19:1–15.
14. DeTar RA, Chustecki JM, Martinez-Hottovy A, et al. Photosynthetic demands on translational machinery drive retention of redundant tRNA metabolism in plant organelles. Proc Natl Acad Sci U S A. 2024;121(52):e2421485121.
15. Fauron C, Casper M, Gao Y, et al. The maize mitochondrial genome: dynamic, yet functional. Trends Genet. 1995;11(6):228–35.
16. Friedman J. The evolution of annual and perennial plant life histories: ecological correlates and genetic mechanisms. Annu Rev Ecol Evol Syst. 2020;51(1):461–81.
17. Gualberto JM, Newton KJ. Plant mitochondrial genomes: dynamics and mechanisms of mutation. Annu Rev Plant Biol. 2017;68(1):225–52.
18. Guo H, Liu Q, Chen Y, et al. Comprehensive assembly and comparative examination of the full mitochondrial genome in *Castanea mollissima* Blume. Genomics. 2023;115(6):110740.
19. Hazkani-Covo E, Zeller RM, Martin WF. Molecular poltergeists: mitochondrial DNA copies (numts) in sequenced nuclear genomes. PLos Genet. 2010;6(2):e1000834.
20. He T, Ding X, Zhang H, et al. Comparative analysis of mitochondrial genomes of soybean cytoplasmic male-sterile lines and their maintainer lines. Funct Integr Genomic. 2021;21:43–57.
21. He W, Xiang K, Chen C, et al. Master graph: an essential integrated assembly model for the plant mitogenome based on a graph-based framework. Briefings Bioinf. 2023;24(1):bbac522.
22. Hu H, Zhang F, Wang P, et al. Evolutionary genetics of wheat mitochondrial genomes. Crop J. 2023;11(6):1774–81.
23. Jeffrey D, Palmer, et al. Tripartite structure of the *Brassica campestris* mitochondrial genome. Nature. 1984;307(5950):437–40.
24. Jeremy N, Timmis, Michael A, et al. Endosymbiotic gene transfer: organelle genomes Forge eukaryotic chromosomes. Nat Rev Genet. 2004;5(2):123–35.
25. Kazutaka K. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013;30(4):772–80.
26. Kerbler SM, Taylor NL, Millar AH. Cold sensitivity of mitochondrial ATP synthase restricts oxidative phosphorylation in *Arabidopsis thaliana*. New Phytol. 2019;221(4):1776–88.
27. Kleine T, Maier UG, Leister D. DNA transfer from organelles to the nucleus: the idiosyncratic genetics of endosymbiosis. Annu Rev Plant Biol. 2009;60(1):115.
28. Kong J, Wang J, Nie L, et al. Evolutionary dynamics of mitochondrial genomes and intracellular transfers among diploid and allopolyploid cotton species. BMC Biol. 2025;23(1):1–20.
29. Kumar R, Maréchal-Drouard L, Akama K, et al. Striking differences in mitochondrial tRNA import between different plant species. Mol Genet Genomics. 1996;252:404–11.
30. Leister D. Origin, evolution and genetic effects of nuclear insertions of organelle DNA. Trends Genet. 2006;21(12):655–63.
31. Lenglez S, Hermand D, Decottignies A. Genome-wide mapping of nuclear mitochondrial DNA sequences links DNA replication origins to chromosomal double-strand break formation in *Schizosaccharomyces Pombe*. Genome Res. 2010;20(9):1250–61.
32. Li J, Ni Y, Lu Q et al. PMGA: A plant mitochondrial genome annotator. Plant Commun. 2024;101191.
33. Liu H, Yu J, Yu X, et al. Structural variation of mitochondrial genomes sheds light on evolutionary history of soybeans. Plant J. 2021;108(5):1456–72.
34. Liu Y, Du H, Li P, et al. Pan-genome of wild and cultivated soybeans. Cell. 2020;182(1):162–76. e113.
35. Madrid E, Severing EI, de Ansorena E, et al. Transposition and duplication of MADS-domain transcription factor genes in annual and perennial Arabis species modulates flowering. Proc Natl Acad Sci U S A. 2021;118(39):e2109204118.
36. Manchekar M, Scissum-Gunn K, Song D, et al. DNA recombination activity in soybean mitochondria. J Mol Biol. 2006;356(2):288–99.
37. Maréchal A, Brisson N. Recombination and the maintenance of plant organelle genome stability. New Phytol. 2010;186(2):299–317.
38. Minh BQ, Schmidt HA, Chernomor O, et al. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. Mol Biol Evol. 2020;37(5):1530–4.
39. Palmer JD, Herbon LA. Plant mitochondrial DNA evolved rapidly in structure, but slowly in sequence. J Mol Evol. 1988;28:87–97.
40. Pujol C, Bailly M, Kern D, et al. Dual-targeted tRNA-dependent amidotransferase ensures both mitochondrial and chloroplastic Gln-tRNAGln synthesis in plants. Proc Natl Acad Sci U S A. 2008;105(17):6481–5.
41. Rice DW, Alverson AJ, Richardson AO, et al. Horizontal transfer of entire genomes via mitochondrial fusion in the angiosperm *Amborella*. Science. 2013;342(6165):1468–73.
42. Richly E, Leister D. NUMTs in sequenced eukaryotic genomes. Mol Biol Evol. 2004;21(6):1081–4.

43 Salinas T, Duchêne A-M, Delage L et al. The voltage-dependent anion channel, a major component of the tRNA import machinery in plant mitochondria. Proc. Natl. Acad. Sci. U.S.A. 2006;103(48): 18362–18367.

44 Schmutz J, Cannon SB, Schlueter J, et al. Genome sequence of the palaeo-polyploid soybean. Nature. 2010;463(7278):178–83.

45 Sedivy EJ, Wu F, Hanzawa Y. Soybean domestication: the origin, genetic architecture and molecular bases. New Phytol. 2017;214(2):539–53.

46 Sergey K, Brian P, et al. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Res. 2017;27(5):722–36.

47 Shi L, Chen H, Jiang M, et al. CPGAVAS2, an integrated plastome sequence annotator and analyzer. Nucleic Acids Res. 2019;47(W1):W65–73.

48 Sloan DB. One ring to rule them all? Genome sequencing provides new insights into the 'master circle'model of plant mitochondrial DNA structure. New Phytol. 2013;200(4):978–85.

49 Sloan DB, Warren JM, Williams AM, et al. Cytonuclear integration and co-evolution. Nat Rev Genet. 2018;19(10):635–48.

50 Sudhir K, Glen S, Koichiro T. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol Biol Evol. 2016;33(7):1870–4.

51 Synenki RM, Levings III, C. S. and, Shah DM. Physicochemical characterization of mitochondrial DNA from soybean. Plant Physiol. 1978;61(3):460–4.

52 Wang J, Kan S, Liao X, et al. Plant organellar genomes: much done, much more to do. Trends Plant Sci. 2024;29(7):754–69.

53 Wang J, Xu G, Ning Y, et al. Mitochondrial functions in plant immunity. Trends Plant Sci. 2022;27(10):1063–76.

54 Wang J, Zou Y, Mower JP et al. Rethinking the mutation hypotheses of plant organellar DNA. gComm. 2024;1(1).

55 Warren JM, Salinas-Giegé T, Triant DA, et al. Rapid shifts in mitochondrial tRNA import in a plant lineage with extensive mitochondrial tRNA gene loss. Mol Biol Evol. 2021;38(12):5735–51.

56 Warren JM, Sloan DB. Interchangeable parts: the evolutionarily dynamic tRNA population in plant mitochondria. Mitochondrion. 2020;52:144.

57 Wick RR, Schultz MB, Justin Z, et al. Bandage: interactive visualization of *de Novo* genome assemblies. Bioinformatics. 2015;31(20):3350–2.

58 Xiao S, Xing J, Nie T, et al. Comparative analysis of mitochondrial genomes of maize CMS-S subtypes provides new insights into male sterility stability. BMC Plant Biol. 2022;22(1):469.

59 Xie J, Chen Y, Cai G, et al. Tree visualization by one table (tvBOT): a web application for visualizing, modifying and annotating phylogenetic trees. Nucleic Acids Res. 2023;51(W1):W587–92.

60 Zhang D, Gao F, Jakovli I, et al. PhyloSuite: an integrated and scalable desktop platform for streamlined molecular sequence data management and evolutionary phylogenetics studies. Mol Ecol Resour. 2020;20(1):348–55.

61 Zhang GJ, Dong R, Lan L-N, et al. Nuclear integrants of organellar DNA contribute to genome structure and evolution in plants. Int J Mol Sci. 2020;21(3):707.

62 Zhang S, Wang J, He W, et al. Variation in mitogenome structural conformation in wild and cultivated lineages of sorghum corresponds with domestication history and plastome evolution. BMC Plant Biol. 2023;23(1):91.

63 Zhang Z, Zhao J, Li J, et al. Evolutionary trajectory of organelle-derived nuclear DNAs in the *Triticum/Aegilops* complex species. Plant Physiol. 2024;194(2):918–35.

64 Zhuang Y, Wang X, Li X, et al. Phylogenomics of the genus *Glycine* sheds light on polyploid evolution and life-strategy transition. Nat Plants. 2022;8(3):233–44.

Publisher's note