A complete chloroplast and mitochondrial genome for velvet bean (Mucuna pruriens, Fabaceae), with genome structure and intergenomic sequence transfers analyses

Rizky Dwi Satrio (rizky.satrio@idu.ac.id)
The Republic of Indonesia Defense University
Isna Arofatun Nikmah
The Republic of Indonesia Defense University
Miftahul Huda Fendiyanto
The Republic of Indonesia Defense University
Mentari Putri Pratami
The Republic of Indonesia Defense University
Asih Perwita Dewi
National Research and Innovation Agency
Muhammad Rifqi Hariri
National Research and Innovation Agency
Bernadetta Rina Hastilestari
National Research and Innovation Agency
Kusumadewi Sri Yulita
National Research and Innovation Agency
M Miftahudin
IPB University

Research Article

Keywords: Fabaceae, Genome, Horizontal gene transfer, Mitochondria, Plastid, Underutilized crop

Posted Date: November 20th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-3612837/v1

Additional Declarations: No competing interests reported.

Abstract

Mucuna pruriens, commonly known as the velvet bean, holds considerable economic and cultural importance as a member of the Fabaceae family. The economic value of this plant is derived from its high protein content, medicinal properties, and potential applications in agriculture and industry. Although the entire nuclear genome consisting of 11 chromosomes of *M. pruriens* has been published, the comprehensive assembly of the organelle genome, specifically the mitochondria, has not been previously accomplished. The evaluation of sequence transfer from both the chloroplast and mitochondrial genomes to the nuclear genome has not been conducted. The complete chloroplast and mitochondrial genome of *M. pruriens* var. *utilis* was assembled and annotated through the utilization of a hybrid approach involving Illumina short-reads and Oxford Nanopore long-reads. The genome of chloroplasts was found to be arranged in a singular circular structure measuring 155,617 bp in length. This structure encompassed a total of 75 genes responsible for encoding proteins, along with 4 genes for ribosomal RNA and 28 genes for transfer RNA. The organization of the mitochondrial genome consisted of a singular circular structure measuring 410,745 bp in length, encompassing a total of 39 protein-coding genes, 4 ribosomal RNA genes, and 16 transfer RNA genes. Through the analysis of sequence transfer, it has been determined that a total of 154,498 bp and 158,285 bp of DNA, derived from the chloroplast and mitochondrial genome respectively, have been incorporated into the nuclear genome. These transfer events predominantly occurred in a relatively recent timeframe. The frequency of transfers of chloroplast to nuclear genome sequences was found to be similar to that of transfers of mitochondrial to nuclear genome sequences. This study presents the initial assembly of the entire organelle, specifically the mitochondrial genome of *M. pruriens*, which serves as a valuable tool for comprehending the genomic evolution within the diverse Fabaceae family.

Introduction

Mucuna pruriens var. *utilis*, commonly known as velvet bean, is a species belonging to the Fabaceae family. The plant is known for its high protein content, making it a valuable source of nutrition for both humans and livestock (Siddhuraju et al. 2000). Additionally, *M. pruriens* seeds are rich in levodopa (L-DOPA), a precursor to dopamine, which has led to its use in the pharmaceutical industry for the treatment of Parkinson disease (Nishihara et al. 2005; Lieu et al. 2010, 2012). In traditional medicine systems, different parts of the plant, including seeds, leaves, and roots, are used to treat a wide range of ailments. It is believed to have aphrodisiac, anti-inflammatory, anti-diabetic, and neuroprotective properties (Aditya et al. 2021; Divyashri et al. 2021). The plant also has potential applications in agriculture, as it is a nitrogen-fixing legume that can improve soil fertility and reduce the need for synthetic fertilizers (Blanco-Canqui 2022). Furthermore, *M. pruriens* fibers have been used in the textile industry for making ropes, mats, and other products. On the other hand, the seeds of this crop contain anti-physiological and toxic factors that can lead to an overall decrease in nutritional quality (Singh et al. 2019). These factors include polyphenols, trypsin inhibitors, phytates, cyanogenic glycosides, oligosaccharides, saponins, lectins, and alkaloids (Lampariello et al. 2012). Therefore, the development of genomic resources for this species is essential for developing genome-based genetic improvement to produce varieties that have good nutritional value and safety.

Mucuna pruriens was the first *Mucuna* species whose entire nuclear genome was sequenced and used to study the L-DOPA biosynthesis-related genes (Hao et al. 2022). On this plant, three genome assemblies were published. The African Center of Excellence in Phytomedicine Research published a scaffold-level assembly (ASM337056v1) constructed from Illumina short-reads in 2018, with the following status: fragment number of 18,487 fragments, N50 of 30 kb, and genome size of 397 Mb. Furthermore, Plateau State University researchers published a scaffold-level assembly (ASM2533094v1) in 2022 using Oxford Nanopore long-reads, with better assembly quality: 546

fragments, N50 of 3.5 Mb, and genome size of 489.8 Mb. Recently, BGI-Qingdao researchers successfully assembled the genome of *M. pruriens* var. utilis at the chromosome level (ASM2476070v1) using a combination of shotgun short-reads and Hi-C sequencing. Although 11 nuclear genome chromosome sequences have been published, organelle genome sequence information is still unavailable. The 152,119 bp chloroplast genome sequence of *M. pruriens* has been published from different study (Yuan et al. 2021), but the sequence was not generated from the individual used for the reference genome. The genome of chloroplast and nucleus provide valuable information for a range of utilizations. However, a comprehensive understanding of cellular genomic content is lacking without the sequencing and assembly of the mitochondrial genome.

There is considerable documentation available highlighting the importance of chloroplasts and mitochondria in various metabolic processes, including photosynthesis, cellular respiration, and ATP synthesis (Green 2011). Endosymbiosis, namely the integration of cyanobacteria and alpha-proteobacteria into primordial eukaryotic cells, has resulted in notable alterations in the composition of organelle genomes as well as the translocation of genes to the genome of nucleus (Andersson et al. 2003; Martin et al. 2015). Within the field of plant biology, it is worth mentioning that there exists considerable variability in the size of mitochondrial genomes, a characteristic that distinguishes them from the majority of animal species (Yuan et al. 2014). Furthermore, it is worth noting that these genomes possess the ability to be arranged into numerous circular chromosomes, thereby facilitating a wide array of structural configurations. The variability within plant mitochondrial genomes at the intra-specific level can display notable fluctuations (Terasawa et al. 2006). The occurrence of structural changes in mitochondrial genes may have serious consequences for survival and adaptability. The presence of structural changes in mitochondrial gene order may have profound implications for adaptation (Wang et al. 2016). This is demonstrated by the occurrence of gene chimerism, which can result in cytoplasmic male sterility (Horn et al. 2014). Another noteworthy aspect of the evolutionary process of plant mitochondrial genomes pertains to the transfer of genes or sequences to the nuclear and chloroplast genomes (Kleine et al. 2009; Ala et al. 2023). Throughout the span of approximately one billion years, a notable alteration in the composition of plant mitochondrial genomes has occurred, deviating from their initial alpha-proteobacteria genome (Bhattacharya et al. 2004; Lane and Archibald 2008). Similarly, chloroplast genomes have also diverged from their cyanobacteria genome (Reyes-Prieto et al. 2007). Throughout evolutionary history, a substantial proportion of genes and non-functional fragments originating from chloroplasts and mitochondria have experienced loss or have been transferred to the nuclear genome. This phenomenon persists in contemporary times. Moreover, recent studies have provided evidence that complete chloroplasts can be exchanged between cells and organisms via tissue grafts, alongside the intracellular transfer of genes or gene fragments (Gao et al. 2014).

This finding implies the potential for complete genome transfers between different species, in addition to the transfer of genetic fragments between different cellular compartments (Yue et al. 2012; Wickell and Li 2020). Hence, it is crucial to conduct an inquiry into intergenomic transfer in order to acquire a comprehensive comprehension of the evolutionary dynamics pertaining to plant chloroplast and mitochondrial genomes. In order to obtain a comprehensive comprehension of the evolutionary and divergent processes occurring within plant lineages, it is imperative to meticulously characterize the three genomes involved. The aforementioned characterization allows for an evaluation of how genetic transfers and structural variation in a specific genome have influenced the coordination of important metabolic activities within the cell and across speciation events.

Given the significance of genomes of the chloroplast and mitochondria in plants for comprehending evolutionary mechanisms and the utilization of *M. pruriens*, the subsequent investigations were undertaken: The research employed short reads and long reads to assemble and annotate the entire chloroplast and mitochondrial genome of

M. pruriens. Furthermore, a comparative analysis was undertaken to investigate the chloroplast and mitochondrial genome of *M. pruriens* in comparison to other Fabaceae, particularly Phaseoleae species that have been previously documented in the database. The objective of this analysis was to offer a comprehensive understanding of the modifications that have occurred within the Phaseoleae subfamily. Additionally, a comparative investigation was performed to determine the differences in the transfer of sequences from chloroplasts and mitochondria to the nuclear genome.

Methods

Organellar genome assembly and annotation

The genome sequencing data for the chloroplast and mitochondrial genome assembly of *Mucuna pruriens* var. *utilis* was obtained from the National Centre for Biotechnology Information (NCBI). Specifically, the data from Illumina short-reads (PRJNA862264) and Oxford Nanopore long-reads (PRJNA640972) were collected. In this study, the researchers utilized the mitochondrial genomes of species belonging to the Fabaceae family, with a specific focus on the Phaseoleae group, which were publicly available in the NCBI database (Table 1). These genomes were employed to identify sequences that are specific to the organelles, namely the chloroplast and mitochondrial genomes, that were assembled as part of this research. Uncorrected Nanopore reads were employed in this study due to the potential introduction of artificial consensus sequences through the correction process. Furthermore, the primary datasets employed in the assembly of organellar genomes consist of sequences with higher accuracy, specifically those generated using the Illumina sequencing platform. To optimize the accuracy of Illumina reads during the correction process, Trimmomatic (Bolger et al. 2014) was employed to filter and trim the reads.

Table 1
Features of the organelle genomes of the Fabaceae species sampled

Species	Chloroplast				Mitochondria			
	ID	Length (bp)	Gene Number	GC Content (%)	ID	Length	Gene Number*	GC Content (%)
Apios americana	NC_025909.1	148772	76	35.00	MW448463.1	434145	32	44.98
Glycine max	NC_007942.1	152218	76	35.37	NC_020455.1	402558	32	45.03
Glycine soja	NC_022868.1	152217	76	35.38	NC_039768.1	402545	32	45.03
Indigofera tinctoria	NC_026680.1	158367	77	35.79	MW448462.1	546392	32	44.73
Millettia pinnata	NC_016708.2	152968	77	34.83	NC_016742.1	425718	33	45.00
Mucuna pruriens	this study	155617	75	34,67	this study	410745	39	45.28
Phaseolus vulgaris	NC_009259.1	150285	75	35.44	NC_045135.1	395516	31	45.11
Vigna angularis	NC_021091.1	151683	74	35.19	NC_021092.1	404466	25	45.19
Vigna radiata	NC_013843.1	151271	75	35.23	NC_015121.1	401262	31	45.11
Pisum sativum	NC_039371.1	150285	74	35.44	NC_068896.1	363833	32	45.08

*The number of protein−coding genes after removing duplications and genes only annotated as ORFs is referred to as the gene number in Mitochondria.

In order to isolate reads specific to organelles from the complete sets of reads, we employed known sequences from closely related taxa as "baits". In this study, three coding sequences from *Glycine max* (NC_007942.1) were utilized for the chloroplast genome analysis. These sequences were provided in the fasta nucleotide format. The genes *rbcL*, *matK*, and *ndhF* were selected due to their high probability of being exclusive to the chloroplast and their significant conservation. The *rbcL* and *matK* genes are typically located at opposite ends of the large single-copy (LSC) region, while the *ndhF* gene is typically found in the small single-copy (SSC) region. These genes are strategically distributed throughout the plastid to ensure that long reads can be obtained with relatively uniform coverage (Syme et al. 2021). To account for the substantial difference in size between the mitochondrial genome and the plastome, we incorporated all 72 coding sequences derived from the mitochondrial genome of *Glycine max* (NC_020455.1), including the genes annotated as ORF. The raw Nanopore reads were aligned to the baits using the minimap2 (Li 2018), and subsequently, the Samtools (Li et al. 2009) was employed to extract the reads that were successfully aligned. The chloroplast genome sequence of *G. max* was utilized to exclude the mitochondrial sequence from the chloroplast-specific reads of *M. pruriens*. Similarly, we utilized the entire mitochondrial genome sequence of *G. max* to exclude the presence of the chloroplast sequence from the mitochondria-specific reads of *M. pruriens*.

Subsequently, the Filtlong (https://github.com/rrwick/Filtlong) tool was employed to retain solely the longest extracted reads, with a maximum coverage threshold of 250×. This step was necessary as excessively high coverage levels can lead to increased fragmentation or hinder the assembly process. The nanopore reads that were obtained were subjected to assembly using the Flye (Kolmogorov et al. 2019), followed by two rounds of polishing using Racon (https://github.com/isovic/racon). The subsequent iteration of the assembly process followed the same procedures outlined previously, with the exception of the omission of bait. The initial assembly of the chloroplast genome was employed as bait to obtain the chloroplast-specific sequence of *M. pruriens*. Similarly, the initial assembly of the mitochondrial genome was employed as a bait to obtain the mitochondrial-specific sequence of *M. pruriens* using Bowtie2 (Kolmogorov et al. 2019).

The organelle-specific reads were extracted from the filtered and trimmed Illumina reads, utilizing the second round assembly as a means of bait. The read sets that were extracted was subsequently subjected to random down sampling, utilizing Rasusa (Hall 2022), until a coverage of 250× was achieved. We employed a hybrid assembly approach using both Illumina and Nanopore reads that were exclusively derived from organelles. The Unicycler (Wick et al. 2017) was utilized for this purpose. The initial step in the Unicycler pipeline involves conducting a short-read assembly utilizing the Spades (Bankevich et al. 2012) algorithm, which is subsequently augmented by incorporating long reads to improve the scaffolding process. The circular genomes of chloroplast and mitochondria were fully assembled and subsequently subjected to annotation to determine the precise positions of functional genes.

In order to obtain precise annotations for the chloroplast and mitochondrial genome of *M. pruriens*, the software tool Geseq (Tillich et al. 2017) was utilized. The reference species used for this purpose were *G. max* (NC_007942.1 and NC_020455.1), *G. soja* (NC_022868.1 and NC_039768.1), *Phaseolus vulgaris* (NC_009259.1 and NC_045135.1), *V. radiata* (NC_013843.1 and NC_015121.1), and *V. angularis* (NC_021091.1 and NC_021092.1). The Chloroplot (Zheng et al. 2020) tool was utilized for the visualization of genome structure.

Phylogenetic analysis and comparisons of synteny

In order to examine the phylogenetic connections within the Fabaceae family, the chloroplast and mitochondrial genomes of *M. pruriens* and nine other Phaseoleae species were utilized. *Pisum sativum* was included as an outgroup for comparison (Table 1). Two distinct datasets were created by extracting and aligning the coding sequences of 71 protein-coding genes from chloroplasts and 29 protein-coding genes from mitochondria. The alignment process was performed using MAFFT (Katoh and Toh 2008), and poorly aligned sections were subsequently trimmed using TrimAL (Capella-Gutiérrez et al. 2009). The datasets were subsequently utilized to perform two distinct phylogenetic analyses employing IQ-TREE (Minh et al. 2020), accompanied by 1000 ultrafast bootstrap replicates to evaluate the support for branches. The Bayesian phylogenetic tree was generated using PartitionFinder (Lanfear et al. 2017) and MrBayes (Huelsenbeck and Ronquist 2001) to evaluate the reliability of the gene-based matrix across different methodologies. The resulting trees were visualized using FigTree (https://github.com/rambaut/figtree).

The verification of synteny within the Phaseoleae group was conducted by assessing sequence similarity between the chloroplast and mitochondrial genomes. In this study, a comparative analysis was conducted on the chloroplast and mitochondrial genomes of nine species belonging to the Phaseoleae subfamily. The comparison was performed using the minimap2 algorithm within the D-Genies (Cabanettes and Klopp 2018) software package. To ensure comparability between the circular genome sequences of chloroplast and mitochondrial genomes, the Cyclic DNA Sequence Aligner (Fernandes et al. 2009) software was employed to align the sequences. This alignment

process establishes a consistent starting position for each sequence, as fasta sequence data is typically linear. The genes order within each organellar genome was assessed using Mauve (Darling et al. 2004) software.

Fragment transfer analysis

The identification of transfer events from chloroplast to mitochondria and organelles to nuclear genomes was conducted using the BLASTn (McGinnis and Madden 2004) software, according to the assembly of *M. pruriens* nuclear genome. The data outputs were categorized into two distinct datasets based on the identity scores: one dataset with identity scores ranging from 80–89%, and another dataset with identity scores ranging from 90–100%. These datasets were utilized to represent different transfer timing, as it is commonly observed that more recent transfers tend to exhibit higher identity scores. The two datasets underwent additional division based on alignment lengths, specifically categorized into three ranges: 100–500 bp, 500–1000 bp, and greater than 1000 bp. The Circos (Krzywinski et al. 2009) software was used to visualize these results.

Repeat analysis of organelle genome

The evaluation of repeat types was conducted for the genomes of both organelles. The identification of repeat types was conducted using the censor (Jurka et al. 1996) algorithm on RepBase (Bao et al. 2015) database with default settings. The MISA (Beier et al. 2017) software was employed to detect simple sequence repeats in *M. pruriens*. The minimum thresholds for mono-, di-, tri-, tetra-, penta-, and hexa-motif microsatellite identification were set at 10, 6, 5, 5, 5, and 5 repeat units, respectively.

Results

The M. pruriens chloroplast and mitochondrial genome assembly and annotation

The complete chloroplast and mitochondrial genomes of *M. pruriens* were successfully assembled as a single circular genome through the utilization of reads obtained from short Illumina reads and long Nanopore reads. The chloroplast genome exhibits a size of 155,617 bp, as depicted in Fig. 1. This measurement corresponds to the observed trend in the sequencing of numerous terrestrial plant species. The overall GC content observed in this study was determined to be 34.67%, a value that falls within the range typically observed in other Fabaceae species (33.97–36.86%) (Hong et al. 2021). Through conducting a comparative analysis, we have successfully annotated a total of 104 genes. This set comprises 75 protein-coding genes, 4 rRNA genes, and 28 tRNA genes. The mitochondrial genome of the subject under study has a length of 410,745 bp, as depicted in Fig. 2. This size is consistent with the majority of previously sequenced terrestrial plant species. The overall GC content observed in this study was determined to be 45.28%, a value that falls within the range typically observed in other species belonging to the Fabaceae family (42.75–45.45%) (Hong et al. 2021). Through our comparative analysis, as many as 59 genes were annotated, comprising 39 protein-coding genes, 4 rRNA genes, and 16 tRNA genes. Out of the annotated genes, a total of 10 genes were found to possess introns. It is worth noting that all of these genes were found to contain only a single intron. The specific genes that were identified to possess an intron include *atpF*, *clP1*, *ndhA*, *ndhB*, *pafl*, *petB*, *petD*, *rpl16*, *rps12*, and *rps16*.

The proportion of repeat sequences in the chloroplast and mitochondrial genome accounted for 13.6% (21,094 bp) and 5.6% (22,957 bp) of the total genome, respectively. We divided these repetitive elements into four distinctive groups based on their lengths in order to study the possible influence of longer repeat sequences on the structural variation of plant organelle genomes. The groups were categorized as follows: Group A consisted of sequences that were less than 20 base pairs in length, Group B included sequences ranging from 20 to 100 base pairs, and Group C

encompassed sequences ranging from 101 to 1000 base pairs. Group A was predicted by MISA, while Groups B and C were identified by censor. The cumulative length of repetitive sequences within each group in the chloroplast genome constituted 0.53% (820 bp, 68 fragments) of the entire genome in group A, 5.97% (9286 bp, 150 fragments) in group B, and 7.79% (12,116 bp, 59 fragments) in group C. The cumulative length of repetitive sequences within each group in the mitochondrial genome constituted 0.13% (538 bp, 48 fragments) of the entire genome in group A, 3.36% (13,802 bp, 230 fragments) in group B, and 2.30% (9,428 bp, 43 fragments) in group C. Additionally, it should be emphasized that the existence of longer repetitive elements may induce modifications in the organization of the genomic feature of organelle genome at various stages of cellular development, resulting in more intricate arrangements that surpass the simplified representation of a circular genome presented in this context.

Structure of organellar genomes among Fabaceae

To provide a comprehensive understanding of the *M. pruriens* organellar genome, we obtained all accessible chloroplast and mitochondrial genomes of Phaseoleae species from the National Centre for Biotechnology Information (NCBI) database. A total of nine species belonging to the Phaseoleae were identified as meeting the specified criteria. The taxon *Pisum sativum*, belonging to the Fabaceae family, was selected as the outgroup species of Phaseoleae.

In this study, we conducted a comparative analysis of fundamental metrics pertaining to genome size, variability, and content. Our objective was to evaluate the divergences in structural evolution within the Phaseoleae group, as well as between the two organelles under investigation (Table 1). Regarding variations in mitochondrial genome length, the *Phaseolus vulgaris* species exhibited the smallest genome size, measuring 395,516 bp, while the *Indigofera tinctoria* species possessed the largest genome size, measuring 546,392 bp. The variation in mitochondrial genome size exceeds 100 kb, while the disparity in chloroplast genome length among the studied species is relatively constrained. Specifically, *Apios americana* exhibits the smallest chloroplast genome at 148,772 bp, whereas *Indigofera tinctoria* possesses the largest chloroplast genome at 158,367 bp. The length of the *M. pruriens* chloroplast and mitochondrial genomes, which were recently assembled, was found to be intermediate when compared to the other Phaseoleae species included in this research. Specifically, the *M. pruriens* genome was 155,617 bp in length, while the chloroplast and mitochondrial genomes were 410,745 bp in length.

After removing duplicate protein genes and ORFs, the number of genes in the mitochondrial genomes of V. angularis and M. pruriens was found to range from 25 to 39 (Table 1). When conducting a comparative analysis of GC content in the organelles of Phaseoleae, it was observed that chloroplast genomes exhibited a range of 34.83–35.79%, while mitochondria displayed a range of 44.73–45.28%. Based on the obtained results, it is evident that mitochondrial genomes within the Phaseoleae group exhibit a higher degree of structural and content diversity compared to chloroplasts.

In this study, we conducted a comparative analysis of the synteny of complete organellar genomes across all species belonging to the Phaseoleae group. The objective was to evaluate the extent of structural rearrangement that exists between various lineages. The observation of the dynamic nature of plant organellar genome structure, both within and across species, in previous studies has motivated the present analyses aimed at identifying potential patterns.

Dot-plot analyses were conducted to examine the chloroplast genome of *M. pruriens* in comparison to species belonging to the Phaseoleae species. The results revealed the occurrence of a significant inversion within the central region of the sequence (Fig. 3). The central region of the genome, known as the large single copy (LSC), exhibits a

reverse orientation compared to the prevailing arrangement observed in the chloroplast genome of Phaseoleae. The chloroplast of *M. pruriens* exhibits genomic fragments that possess a gene arrangement akin to that observed in all members of the Phaseoleae tribe. However, it also contains a substantial fragment with an orientation that is opposite to the aforementioned arrangement. The confirmation of this result was obtained through the utilization of the gene arrangement map in the genome alignment (**Fig. S1**). The Phaseoleae chloroplast genome alignment, utilizing the publicly available *M. pruriens* accession as RefSeq (NC_054176.1), also demonstrated comparable findings. Nevertheless, a distinction can be observed between the chloroplast genome sequences and the sequences examined in this particular study. The orientation of the short single copy (SSC) fragment in sequence NC_054176.1 was consistent with that of *G. max* and *G. soja* (data not shown). However, the chloroplast genome sequence analyzed in this study exhibits a reverse orientation. In this study, it was observed that the SSC fragment of *M. pruriens* exhibits a similar orientation direction to that of the majority of species within the Phaseoleae genus. Considering that the chloroplast genome sequence in this study was assembled from a combination of short- and long-reads, with nanopore long-reads serving as scaffolding, this study confirmed the structure of the gene arrangement in the *M. pruriens chloroplast* genome.

An examination of the mitochondrial genome through dot-plot analysis revealed that there were only short segments (measuring less than 5 kilobases) of synteny observed between *M. pruriens* and *P. sativum* (Fig. 4). Nevertheless, when comparing Phasesoleae species, particularly *G. max* and *G. soja*, it was observed that there were longer stretches of synteny, although the gene arrangement between them did not exhibit a discernible pattern (**Fig. S2**). The observed conservation and variability patterns among species in the Phaseoleae group suggest that the decay of mitochondrial genomic synteny is correlated with the amount of time that has passed since their divergence. However, it is worth noting that functional genes within the genome, despite undergoing rearrangements, maintain internal synteny and therefore exhibit a high level of conservation in terms of protein coding.

Organellar phylogenetic relationships in Fabaceae

The utilization of intact organellar genomes has served as a fundamental indicator in elucidating the phylogenetic relationships among different species. In the present investigation, we combined the collective CDS from each chloroplast and mitochondrial genome and employed these datasets to deduce phylogenetic trees (Fig. 5). The chloroplast and mitochondrial trees exhibit identical topological structures and branching patterns. The topological structure of the chloroplast and mitochondrial phylogenetic trees was identical. When the branch lengths of the trees are compared, it is clear that the rate of evolution of mitochondrial coding genes was substantially lower than that of chloroplast genes. The genomes of both chloroplast and mitochondria provide evidence supporting the notion that *M. pruriens* is the sister group of the cultivated Phaseoleae, commonly known as "edible beans." This group includes *G. max*, *G. soja*, *Phaseolus vulgaris*, *V. radiata*, and *V. angularis*.

Sequence transfer between genomes

The essentiality of comprehending intergenomic evolution necessitates the characterization of sequence transfer within organellar and nuclear genomes. Through the acquisition of comprehensive and refined assemblies for the three genomes of *M. pruriens*, by considering the lengths and characteristics of the organelles and the nuclear genome, we were able to evaluate the sequence transfer that occurs between them.

In this study, we conducted a comparative analysis of the nuclear, mitochondrial, and chloroplast genome assemblies obtained from *M. pruriens*. We aimed to identify and quantify any instances of sequence transfers that may have occurred between these genomes. In each category of fragment length, the number of transferred sequences was found to be greater in chloroplasts compared to mitochondria at the 80–100% sequence similarity

groups. The study revealed that a significant proportion of sequence transfers, specifically 67.0% from chloroplast and 82.0% from mitochondria, were categorized within the 90–100% sequence similarity range (Fig. 6). An interesting observation can be made regarding the number of transferred sequences in the 500–600, 600–700, and 700–800 base pair (bp) categories. Specifically, it is noteworthy that the percentage of sequence similarity in the chloroplast is higher in the 80–90% range compared to the 90–100% range. The elevated occurrence of transferred sequences displaying a sequence similarity ranging from 90–100% indicates an ongoing mechanism in which sequences are transferred and subsequently lost within relatively short time periods.

Additionally, an evaluation was carried out to determine the quantity of transfers, and the following mapping of the chromosomal locations of transfers within the nuclear genome (Fig. 6). All nuclear chromosomes showed a similar trend in the total number of sequence transfers into the nuclear genome. The frequency of chloroplast genome transfers on nuclear chromosomes was three times higher than that of mitochondrial genome transfers in both 90–100% and 80–90% fragments similarity (Fig. 6 and Fig. 7). Fragments with 90–100% similarity are more than twice as numerous as those with 80–90% similarity in each chromosome of the nuclear genome, particularly in the chloroplast. Meanwhile, on each chromosome of the nuclear genome, the ratio of mitochondrial fragments with similarity of 90–100% and 80–90% was relatively diverse. Mitochondrial genome fragments with similarity of 80–90% to 90–100% in chromosome 9 of the nuclear genome have the same ratio, which is significantly higher than the ratio of 5% in chromosome 11.

Horizontal gene transfer from the organelle genome to the nuclear genome occurs in relatively large numbers, approximately 3000 for chloroplasts and 1000 for mitochondria, respectively. Transfer of fragments between organelle genomes in *M. pruriens* is possible, but rare. Only nine fragment transfers were discovered between the chloroplast and mitochondria genomes (Fig. 7), with the vast majority occurring in protein-coding sequence regions. These sequences, interestingly, were also found in the repeat region.

Discussion

The assembly of plant chloroplast and mitochondrial genomes

Fragment inversions in the chloroplast genome have been observed and studied in various plant species (Li and Zheng 2018). The aforementioned inversions are regarded as structural variations that offer significant insights into the evolutionary processes of plant genomes. In this investigation, the chloroplast genome of *M. pruriens* was comprehensively sequenced using a combination of short Illumina reads and long Nanopore reads. This method was chosen to overcome the difficulties involved with assembling such complex genomes. The chloroplast genome of *M. pruriens* was found to be circular and had a total length of 155,617 bp. It was divided into four distinct regions, namely the long single-copy (LSC), short single-copy (SSC), and a pair of inverted repeats (IR), as indicated in Fig. 1. The genome of *M. pruriens* exhibits an inversion mutation in its large single copy (LSC) region, in contrast to the chloroplast genome of most species within the Phaseoleae species (Fig. 4). The confirmation of this result was achieved by utilizing the chloroplast genome sequence of *M. pruriens* obtained in this study, as well as the previous assembly that solely relied on Illumina short reads. Although the two chloroplast genome assemblies exhibited the same pattern in the LSC region, there were notable variations in the SSC position. The orientation of the *M. pruriens* chloroplast SSC region, as observed in the previous study, exhibited a pattern that was comparable to that of the *Glycine* species. However, the present study revealed a contrasting orientation. The utilization of extended Nanopore reads in the scaffolding procedure provided validation for the enhancement of the *M. pruriens* chloroplast genome sequence and assembly in this investigation.

Observations of inversions in the chloroplast LSC region have been documented in diverse plant species. The LSC region constitutes one of the three regions comprising the quadripartite structure of the chloroplast genome, alongside the SSC region and the IR regions (Xu et al. 2015). As an illustration, the length of the LSC region in the chloroplast genome of *G. max* was reported to be 83,175 bp (Saski et al. 2005). This region is flanked by the IR regions and separated from the SSC region by the IR regions as well. Inversions within the LSC region can occur, leading to changes in gene order and potentially affecting the functioning of chloroplast genes. Prior research has shown that the chloroplast genomes of herbaceous plants in the Zingiberaceae family have undergone rapid evolutionary changes, including structural changes such as inversions (Yang et al. 2022). The identification of inversions within chloroplast genomes suggests the inherent dynamism of these genomes and their potential significance in facilitating adaptation and evolutionary processes (Cui et al. 2006). The aforementioned inversions play a significant role in the ongoing evolutionary process of chloroplast genomes, offering valuable insights into the mechanisms of genome evolution, adaptation, and the diversification of species. The presence of inversions within the LSC region can give rise to various consequences, including impacts on gene expression, maintenance of genome integrity, and the evolutionary patterns observed in chloroplast genomes across diverse plant lineages (Yagi and Shiina 2014). Conducting additional research on the mechanisms and functional implications associated with fragment inversions in the chloroplast genome will contribute to the advancement of knowledge regarding plant evolution and the significance of chloroplasts in endosymbiosis.

The mitochondrial genomes of plants have experienced significant and expeditious alterations in their structural composition subsequent to the initial endosymbiotic event (Knoop et al. 2011). As a result of this particular mode of evolution, the complexity of mitochondrial genome composition limits the effectiveness of conventional sequencing and assembly methods. The conventional circular model of genome organization, which is generally applicable to the majority of animal species, proves to be insufficient in comprehending the intricate genomic structure of plant mitochondria (Christensen 2013). Plant mitochondria exhibit a diverse range of genomic structures, including multiple circular forms, branched configurations, linear arrangements, and mixed variations (Mower et al. 2012). The main objective of this research endeavor was to comprehensively sequence the mitochondrial genome of *M. pruriens*. To address the difficulties associated with the assembly of complex genomes of this nature, we employed a hybrid approach that incorporated both short and long reads. The mitochondrial genome, which was assembled in a circular form, had a length of 410,745 bp. It contained a total of 39 protein-coding genes, as indicated in Fig. 2. Circular arrangements seem to be the prevailing pattern observed in the assembled mitochondrial genomes of Fabaceae (Table 1). Based on the size and structure of the mitochondrial genome, some sites showed the potential to generate branching or sub-circular structures (Gualberto et al. 2014). This phenomenon implies that the mitochondrial genome of *M. pruriens* might exhibit distinct variations during various developmental stages or in different types of tissues, similar to what has been observed in other plant species (Kitazaki and Kubo 2010; Knoop et al. 2011). Further investigation is required to gain a comprehensive understanding of this phenomenon in the genome of plant mitochondria. To achieve these objectives, the acquisition of the initial comprehensive mitochondrial genome in *Mucuna* will serve as a valuable point of reference for subsequent investigations aimed at comprehending the functionality and organization of plant mitochondria.

The evolution of chloroplast and mitochondrial genomes

The evolution of the organellar genome in Fabaceae has been the subject of several studies. These studies have provided insights into the structural changes, gene content, and phylogenetic relationships within this diverse plant family (Palmer 1992; Antunes et al. 2020; Hong et al. 2021; Yang et al. 2022). One study focused on the chloroplast genome evolution in legumes, including Fabaceae, and identified localized hypermutation, gene losses, and

fragment inversions as dynamic features of the chloroplast genomes in the Papilionoideae subfamily (Liu et al. 2016). These findings suggest that the chloroplast genomes in Fabaceae are subject to evolutionary changes, potentially driven by adaptive processes and genetic drift. It has been observed that there is a rapid evolution of plant mitochondria, even within a single species. Significant differences between individuals define this evolution, which extends to hundreds of bases placed outside of functioning genes. Consequently, this phenomenon contributes to a wide range of genome sizes in plant mitochondria. Furthermore, it has been shown that heteroplasmy, extensive genome recombination, and gene chimerism may be detected in mitochondrial genomes at both the species and individual levels (McCauley 2013). These phenomena have been found to be linked to certain phenotypes, although they may not necessarily be the sole causative factors. When comparing lineages that are more distantly related, there is a greater disparity in genome size and structure. However, it is noteworthy that functional genes are conserved despite these differences.

Because of their great anthropogenic relevance, the mitochondrial genomes of numerous crop species have been intensively sequenced and investigated (Arrieta-Montiel et al. 2009; Mandel and McCauley 2015; Wang et al. 2016; Ala et al. 2023). However, the sequencing of mitochondrial genomes from underutilized species, such as *M. pruriens*, has been comparatively limited. Considering the fact that the Fabaceae family encompasses a wide range of significant crop species, as well as lesser-known crop species and various other forms, it presents an excellent opportunity for investigating the evolutionary patterns of organelle genomes. Moreover, studying the evolutionary alterations in relation to gene expression and adaptation processes, such as speciation, can provide valuable insights (Horn et al. 2014; Yagi and Shiina 2014). The sequencing of all three genomic compartments of the *M. pruriens* marks a significant milestone as it is the first species within the *Mucuna* genus to have achieved this feat. In this study, we conducted a comparative analysis of the organelles found in the *Mucuna* lineage and those present in other significant lineages within the Phaseoleae family (Table 1). In the Phaseoleae clade, the chloroplast genome exhibited restricted structural variation, similar to the observed inversion in the LSC region of the chloroplast (Fig. 3 & S1). The gene content within chloroplasts exhibited limited variation, indicating the presence of functional conservation observed in other lineages. Furthermore, to ensure the reliability of the datasets for phylogenetic reconstruction, Bayesian analyses were performed using MrBayes in addition to the ML approach used in IQTREE. No significant variations in topology were observed; however, certain disparities in the length of phylogenetic branch (Fig. 3). The phylogenetic study of organellar genomes in the Fabaceae revealed that tree branch lengths in chloroplast genes were much longer than in mitochondrial genes. This study supports the findings of Hong et al. (2021), who discovered that the rate of evolution of the protein-coding sequence of mitochondria is slower than that of the chloroplast in the Fabaceae family, by studying the organelles genome of *Dalbergia odorifera*.

The feature of rapidly evolution has been found in the mitochondrial genome as a whole. In mitochondrial genomes, the rate of genomic structural alteration is greatly increased. When comparing Fabaceae lineages above the generic level, there is no synteny beyond sections of 5 Kb (Fig. 4). The observed differences in the rate of sequence and structural evolution are thought to be caused by mitochondrial break-induced repair mechanisms (Gualberto et al. 2014).

Regarding phylogenetic topology, it is expected that the chloroplast and mitochondrial trees would exhibit a high degree of similarity due to their uniparentally non-recombinant genomes (Fig. 5). Nevertheless, it is important to recognize the presence of various cases of phylogenetic discordance. These instances could potentially arise from a variety of factors, such as inefficient lineage sorting after a biparental organellar inheritance event, differences in evolutionary rates, and insufficient species representation throughout the sampling procedure (Lynch et al. 2006; Zhang and Sodmergen 2010; Sloan and Taylor 2011; Greiner et al. 2015). It is worth noting that such inheritance

events have been documented in the Phaseoleae group. In order to enhance comprehension of the evolutionary patterns exhibited by mitochondrial genomes, it is imperative to undertake a more extensive sampling approach encompassing a wider range of individuals. The Fabaceae family provides a very ideal lineage for the complete exploration of many theories relevant to genome evolution in plants due to its vast variety of habits, phenology, and species richness (Choi et al. 2019; Shatskaya et al. 2023). The first step toward understanding the diversity of the mitochondrial genome in this specific family is the sequencing and assembly of the whole mitochondrial genome in *M. pruriens*.

Intergenomic sequence transfers

The transfer of sequences between genomic compartments is a prominent and constantly changing characteristic of the endosymbiotic mechanism, as evidenced by the presence of identifiable sequences shared among various plant genomes (Yue et al. 2012; Gao et al. 2014; Wickell and Li 2020). Numerous investigations have been undertaken to elucidate the underlying processes of sequence transference. These include the inadvertent integration resulting from the improper mending of double-stranded breaks, as well as the deliberate induction of functional genetic variability in the recipient genome as a response to stress or other stimuli (Gualberto et al. 2014). Regardless of the various mechanisms involved, intensive investigations are required to comprehensively document the range of transfers occurring among plants. This research aimed to find trends in the distribution, frequency, timing, and possible biases of these transfers.

In this study, we conducted a comprehensive analysis of whole sequence interchange from all three genomic compartments in *M. pruriens*, marking the first instance of such an assessment. Based on our analyses, it is evident that the occurrence of chloroplast transfers surpassed that of mitochondrial transfers. Specifically, approximately 99.28% (154,498 bp) of the nuclear genome contained fragments originating from the chloroplast genome, whereas only 38.54% (158,285 bp) of the nuclear genome exhibited mitochondrial origins. However, it should be noted that the cumulative length transferred by each organelle genome in the genome of the nucleus was comparable. In contrast to the 83% chloroplast transfer observed in the nuclear genome of *Oryza* (Matsuo et al. 2005), and the 16% observed in *Arabidopsis* (Leister 2005), the level of chloroplast transfer in *M. pruriens*, at 99%, is comparatively large. Nevertheless, it was observed that *Dalbergia odorifera* and closely related species also exhibited numerous cases of abundant chloroplast sequence transfer (Hong et al. 2021), indicating that the Fabaceae family may have a propensity for experiencing elevated rates of chloroplast sequence transfer. Further investigation is required to elucidate the underlying factors that influence the occurrence of transfers, as well as the subsequent retention and elimination processes in the recipient genome. This will help to ascertain the reasons behind the observed variations in transfer rates among the plant species that have been examined to date.

While the nuclear and chloroplast genomes of *M. pruriens* have been previously documented (Yuan et al. 2021; Hao et al. 2022), The inclusion of the full mitochondrial genome allowed for detailed comparisons across various genomic compartments within an individual as well as with other lineages. This broader perspective enables a more comprehensive examination of evolutionary changes. The accurate assembly of plant mitochondrial genomes was made possible by the combination of short-read and long-read sequencing, which is restricted due to the limited synteny that even closely related lineages exhibit. This study was also able to detect inversions and rearrangements of genes in the chloroplast genome by combining these sequencing technologies. In subsequent investigations, it is plausible to extend the utilization of this methodology to encompass a wider range of tissues, populations, and species for the purpose of elucidating the impact of structural and functional evolution on the considerable diversity observed in plant organellar genomes. The information possesses the capacity to be implemented in a multitude of contexts and endeavors, encompassing predictive breeding investigations, organellar transplantation, and a wide

range of constructs and targets for gene editing. The Fabaceae family and *Mucuna* genus present themselves as promising subjects for further academic investigation, owing to their extensive diversity and wide range of practical applications for human benefit. The potential reason behind the extensive variety observed within the Fabaceae family could be attributed to the capacity of specific lineages to swiftly adapt in response to periods of rapid environmental transformation. This adaptability may be facilitated by the transfer of genetic material between different genomes.

Declarations

Declaration of competing interest

The authors have no conflicts of interest to declare.

Author Contribution

R.D.S and I.A.N wrote the main manuscript text. All authors reviewed the manuscript.

Acknowledgments

The research was funded by the Research and Innovation for Advanced Indonesia (RIIM) Batch 2 Program from the National Research and Innovation Agency (BRIN) and the Indonesia Endowment Fund for Education (LPDP), Ministry of Finance of Indonesia.

References

1. Aditya B, Arivarasu L, Rajeshkumar S, Thangavelu L (2021) Antioxidant and anti-inflammatory activity of titanium dioxide nanoparticles synthesized using Mucuna pruriens. J Pharm Res Int 33:414–422. https://doi.org/10.9734/jpri/2021/v33i62a35616
2. Ala KG, Zhao Z, Ni L, Wang Z (2023) Comparative analysis of mitochondrial genomes of two alpine medicinal plants of Gentiana (Gentianaceae). PLoS One 18:1–21. https://doi.org/10.1371/journal.pone.0281134
3. Andersson SGE, Karlberg O, Canbäck B, Kurland CG (2003) On the origin of mitochondria: A genomics perspective. Philos Trans R Soc B Biol Sci 358:165–179. https://doi.org/10.1098/rstb.2002.1193
4. Antunes AM, Soares TN, Targueta CP, et al (2020) The chloroplast genome sequence of Dipteryx alata Vog. (Fabaceae: Papilionoideae): genomic features and comparative analysis with other legume genomes. Brazilian J Bot 43:271–282. https://doi.org/10.1007/s40415-020-00599-3
5. Arrieta-Montiel MP, Shedge V, Davila J, et al (2009) Diversity of the Arabidopsis mitochondrial genome occurs via nuclear-controlled recombination activity. Genetics 183:1261–1268. https://doi.org/10.1534/genetics.109.108514
6. Bankevich A, Nurk S, Antipov D, et al (2012) SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19:455–477. https://doi.org/10.1089/cmb.2012.0021
7. Bao W, Kojima KK, Kohany O (2015) Repbase Update, a database of repetitive elements in eukaryotic genomes. Mob DNA 6:1–6. https://doi.org/10.1186/s13100-015-0041-9

8. Beier S, Thiel T, Münch T, et al (2017) MISA-web: a web server for microsatellite prediction. Bioinformatics 33:2583–2585. https://doi.org/10.1093/bioinformatics/btx198
9. Bhattacharya D, Yoon HS, Hackett JD (2004) Photosynthetic eukaryotes unite: endosymbiosis connects the dots. Bioessays 26:50–60. https://doi.org/10.1002/bies.10376
10. Blanco-Canqui H (2022) Cover crops and soil ecosystem engineers. Agron J 114:3096–3117. https://doi.org/10.1002/agj2.21160
11. Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120. https://doi.org/10.1093/bioinformatics/btu170
12. Cabanettes F, Klopp C (2018) D-GENIES: dot plot large genomes in an interactive, efficient and simple way. PeerJ 6:e4958. https://doi.org/10.7717/peerj.4958
13. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T (2009) trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25:1972–1973. https://doi.org/10.1093/bioinformatics/btp348
14. Choi I-S, Schwarz EN, Ruhlman TA, et al (2019) Fluctuations in Fabaceae mitochondrial genome size and content are both ancient and recent. BMC Plant Biol 19:1–15. https://doi.org/10.1186/s12870-019-2064-8
15. Christensen AC (2013) Plant mitochondrial genome evolution can be explained by DNA repair mechanisms. Genome Biol Evol 5:1079–1086
16. Cui L, Leebens-Mack J, Wang LS, et al (2006) Adaptive evolution of chloroplast genome structure inferred using a parametric bootstrap approach. BMC Evol Biol 6:1–12. https://doi.org/10.1186/1471-2148-6-13
17. Darling ACE, Mau B, Blattner FR, Perna NT (2004) Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Res 14:1394–1403. https://doi.org/10.1101/gr.2289704
18. Divyashri S, Priyadharshini R, Rajeshkumar S, Sinduja P (2021) Comparative anti-inflammatory activity of hexane and aqueous extract of Mucuna pruriens. J Pharm Res Int 33:1156–1166. https://doi.org/10.9734/jpri/2021/v33i60b34727
19. Fernandes F, Pereira L, Freitas AT (2009) CSA: an efficient algorithm to improve circular DNA multiple alignment. BMC Bioinformatics 10:1–13. https://doi.org/10.1186/1471-2105-10-230
20. Gao C, Ren X, Mason AS, et al (2014) Horizontal gene transfer in plants. Funct Integr Genomics 14:23–29. https://doi.org/10.1007/s10142-013-0345-0
21. Green BR (2011) Chloroplast genomes of photosynthetic eukaryotes. plant J 66:34–44. https://doi.org/10.1111/j.1365-313X.2011.04541.x
22. Greiner S, Sobanski J, Bock R (2015) Why are most organelle genomes transmitted maternally? Bioessays 37:80–94. https://doi.org/10.1002/bies.201400110
23. Gualberto JM, Mileshina D, Wallet C, et al (2014) The plant mitochondrial genome: Dynamics and maintenance. Biochimie 100:107–120. https://doi.org/10.1016/j.biochi.2013.09.016
24. Hall MB (2022) Rasusa: Randomly subsample sequencing reads to a specified coverage. J Open Source Softw 7:3941. https://doi.org/10.21105/joss.03941
25. Hao S, Ge Q, Shao Y, et al (2022) Chromosomal-level genome of velvet bean (Mucuna pruriens) provides resources for L-DOPA synthetic research and development. DNA Res 29:dsac031. https://doi.org/10.1093/dnares/dsac031
26. Hong Z, Liao X, Ye Y, et al (2021) A complete mitochondrial genome for fragrant Chinese rosewood (Dalbergia odorifera, Fabaceae) with comparative analyses of genome structure and intergenomic sequence transfers.

BMC Genomics 22:1–13. https://doi.org/10.1186/s12864-021-07967-7

27. Horn R, Gupta KJ, Colombo N (2014) Mitochondrion role in molecular basis of cytoplasmic male sterility. Mitochondrion 19:198–205. https://doi.org/10.1016/j.mito.2014.04.004
28. Huelsenbeck JP, Ronquist F (2001) MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics 17:754–755. https://doi.org/10.1093/bioinformatics/17.8.754
29. Jurka J, Klonowski P, Dagman V, Pelton P (1996) CENSOR—a program for identification and elimination of repetitive elements from DNA sequences. Comput Chem 20:119–121. https://doi.org/10.1016/s0097-8485(96)80013-1
30. Katoh K, Toh H (2008) Recent developments in the MAFFT multiple sequence alignment program. Brief Bioinform 9:286–298. https://doi.org/10.1093/bib/bbn013
31. Kitazaki K, Kubo T (2010) Cost of having the largest mitochondrial genome: evolutionary mechanism of plant mitochondrial genome. J Bot 2010:1–12. https://doi.org/10.1155/2010/620137
32. Kleine T, Maier UG, Leister D (2009) DNA transfer from organelles to the nucleus: the idiosyncratic genetics of endosymbiosis. Annu Rev Plant Biol 60:115–38. https://doi.org/10.1146/annurev.arplant.043008.092119
33. Knoop V, Volkmar U, Hecht J, Grewe F (2011) Mitochondrial genome evolution in the plant lineage. In: Kempken F (ed) Plant Mitochondria. Springer, New York, pp 3–29
34. Kolmogorov M, Yuan J, Lin Y, Pevzner PA (2019) Assembly of long, error-prone reads using repeat graphs. Nat Biotechnol 37:540–546. https://doi.org/10.1038/s41587-019-0072-8
35. Krzywinski M, Schein J, Birol I, et al (2009) Circos: an information aesthetic for comparative genomics. Genome Res 19:1639–1645. https://doi.org/10.1101/gr.092759.109
36. Lampariello LR, Cortelazzo A, Guerranti R, et al (2012) The magic velvet bean of Mucuna pruriens. J Tradit Complement Med 2:331–339. https://doi.org/10.1016/s2225-4110(16)30119-5
37. Lane CE, Archibald JM (2008) The eukaryotic tree of life: endosymbiosis takes its TOL. Trends Ecol Evol 23:268–275. https://doi.org/10.1016/j.tree.2008.02.004
38. Lanfear R, Frandsen PB, Wright AM, et al (2017) PartitionFinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. Mol Biol Evol 34:772–773. https://doi.org/10.1093/molbev/msw260
39. Leister D (2005) Genomics-based dissection of the cross-talk of chloroplasts with the nucleus and mitochondria in Arabidopsis. Gene 18:110–116. https://doi.org/10.1016/j.gene.2005.03.039
40. Li B, Zheng Y (2018) Dynamic evolution and phylogenomic analysis of the chloroplast genome in Schisandraceae. Sci Rep 8:9285. https://doi.org/10.1038/s41598-018-27453-7
41. Li H (2018) Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics 34:3094–3100. https://doi.org/10.1093/bioinformatics/bty191
42. Li H, Handsaker B, Wysoker A, et al (2009) The sequence alignment/map format and SAMtools. Bioinformatics 25:2078–2079. https://doi.org/10.1093/bioinformatics/btp352
43. Lieu CA, Kunselman AR, Manyam B V, et al (2010) A water extract of Mucuna pruriens provides long-term amelioration of parkinsonism with reduced risk for dyskinesias. Parkinsonism Relat Disord 16:458–465. https://doi.org/10.1016/j.parkreldis.2010.04.015
44. Lieu CA, Venkiteswaran K, Gilmour TP, et al (2012) The antiparkinsonian and antidyskinetic mechanisms of Mucuna pruriens in the MPTP-treated non-human primate. Evid Based Complement Altern Med 2012:1–10. https://doi.org/10.1155/2012/840247

45. Liu W, DaPeng N, Wang Y, et al (2016) Intraspecific and heteroplasmic variations, gene losses and inversions in the chloroplast genome of Astragalus membranaceus. Sci Rep 22:21669. https://doi.org/10.1038/srep21669
46. Lynch M, Koskella B, Schaack S (2006) Mutation pressure and the evolution of organelle genomic architecture. Science (80-) 311:1727–1730. https://doi.org/10.1126/science.1118884
47. Mandel JR, McCauley DE (2015) Pervasive Mitochondrial Sequence Heteroplasmy in Natural Populations of Wild Carrot, Daucus Carota Spp. Carota L. PLoS One 10:e0136303. https://doi.org/10.1371/journal.pone.0136303
48. Martin W, Garg SG, Zimorski V (2015) Endosymbiotic theories for eukaryote origin. Philos Trans R Soc B Biol Sci 370:20140330. https://doi.org/10.1098/rstb.2014.0330
49. Matsuo M, Ito Y, Yamauchi R, Obokata J (2005) The rice nuclear genome continuously integrates, shuffles, and eliminates the chloroplast genome to cause chloroplast-nuclear DNA flux. Plant Cell 17:665–675. https://doi.org/10.1105/tpc.104.027706
50. McCauley DE (2013) Paternal leakage, heteroplasmy, and the evolution of plant mitochondrial genomes. New Phytol 200:966–977. https://doi.org/10.1111/nph.12431
51. McGinnis S, Madden TL (2004) BLAST: at the core of a powerful and diverse set of sequence analysis tools. Nucleic Acids Res 32:W20–W25. https://doi.org/10.1093/nar/gkh435
52. Minh BQ, Schmidt HA, Chernomor O, et al (2020) IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. Mol Biol Evol 37:1530–1534. https://doi.org/10.1093/molbev/msaa015
53. Mower JP, Sloan DB, Alverson AJ (2012) Plant mitochondrial genome diversity: the genomics revolution. In: Wendel J, Greilhuber J, Dolezel J, Leitch I (eds) Plant Genome Diversity. Springer, Viena, pp 123–144
54. Nishihara E, Parvez MM, Araya H, et al (2005) L-3-(3,4-Dihydroxyphenyl)alanine (L-Dopa), an allelochemical exuded from velvet bean (Mucuna pruriens) roots. Plant Growth Regul 45:113–120. https://doi.org/10.1007/s10725-005-0610-x
55. Palmer JD (1992) Comparison of chloroplast and mitochondrial genome evolution in plants. In: Herrmann RG (ed) Cell Organelles. Springer, Vienna, pp 99–133
56. Reyes-Prieto A, Weber APM, Bhattacharya D (2007) The origin and establishment of the plastid in algae and plants. Annu Rev Genet 41:147–168. https://doi.org/10.1146/annurev.genet.41.110306.130134
57. Saski C, Lee SB, Daniell H, et al (2005) Complete chloroplast genome sequence of Gycine max and comparative analyses with other legume genomes. Plant Mol Biol 59:309–322. https://doi.org/10.1007/s11103-005-8882-0
58. Shatskaya N V., Bogdanova VS, Kosterin OE, Vasiliev G V. (2023) New Insights into Plastid and Mitochondria Evolution in Wild Peas (Pisum L.). Diversity 15:1–25. https://doi.org/10.3390/d15020216
59. Siddhuraju P, Becker K, Makkar HPS (2000) Studies on the nutritional composition and antinutritional factors of three different germplasm seed materials of an under-utilized tropical legume, Mucuna pruriens var. utilis. J Agric Food Chem 48:6048–6060. https://doi.org/10.1021/jf0006630
60. Singh S, Gupta PS, Gupta RK (2019) Evaluation of anti-anxiety activity of Mucuna pruriens. J Drug Deliv Ther 9:104–107. https://doi.org/10.22270/jddt.v9i4-a.3420
61. Sloan DB, Taylor DR (2011) Evolutionary rate variation in organelle genomes: the role of mutational processes. In: Bullerwell C (ed) Organelle Genetics: Evolution of Organelle Genomes and Gene Expression. Springer, Berlin, pp 123–146
62. Syme AE, McLay TGB, Udovicic F, et al (2021) Long-read assemblies reveal structural diversity in genomes of organelles–an example with Acacia pycnantha. Gigabyte 2021:1–23. https://doi.org/10.46471/gigabyte.36

63. Terasawa K, Odahara M, Kabeya Y, et al (2006) The mitochondrial genome of the moss Physcomitrella patens sheds new light on mitochondrial evolution in land plants. Mol Biol Evol 24:699–709. https://doi.org/10.1093/molbev/msl198
64. Tillich M, Lehwark P, Pellizzer T, et al (2017) GeSeq–versatile and accurate annotation of organelle genomes. Nucleic Acids Res 45:W6–W11. https://doi.org/10.1093/nar/gkx391
65. Wang X, Bi C, Xu Y, et al (2016) The whole genome assembly and comparative genomic research of Thellungiella parvula (extremophile crucifer) mitochondrion. Int J Genomics 2016:5283628. https://doi.org/10.1155/2016/5283628
66. Wick RR, Judd LM, Gorrie CL, Holt KE (2017) Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. PLoS Comput Biol 13:e1005595. https://doi.org/10.1371/journal.pcbi.1005595
67. Wickell DA, Li F (2020) On the evolutionary significance of horizontal gene transfers in plants. New Phytol 225:113–117. https://doi.org/10.1111/nph.16022
68. Xu J-H, Liu Q, Hu W, et al (2015) Dynamics of chloroplast genomes in green plants. Genomics 106:221–231. https://doi.org/10.1016/j.ygeno.2015.07.004
69. Yagi Y, Shiina T (2014) Recent advances in the study of chloroplast gene expression and its evolution. Front Plant Sci 5:1–7. https://doi.org/10.3389/fpls.2014.00061
70. Yang Q, Fu G-F, Wu Z-Q, et al (2022) Chloroplast genome evolution in four montane Zingiberaceae taxa in China. Front Plant Sci 12:1–9. https://doi.org/10.3389/fpls.2021.774482
71. Yuan J, Zhu Q, Liu B (2014) Phylogenetic and biological significance of evolutionary elements from Metazoan mitochondrial genomes. PLoS One 9:e84330. https://doi.org/10.1371/journal.pone.0084330
72. Yuan X-L, Zha D-X, Xue L, et al (2021) Complete chloroplast genome sequence of Mucuna prurien and its phylogenetic position. Mitochondrial DNA Part B 6:2480–2481. https://doi.org/10.1080/23802359.2021.1915199
73. Yue J, Hu X, Sun H, et al (2012) Widespread impact of horizontal gene transfer on plant colonization of land. Nat Commun 3:1152. https://doi.org/10.1038/ncomms2148
74. Zhang Q, Sodmergen (2010) Why does biparental plastid inheritance revive in angiosperms? J Plant Res 123:201–206. https://doi.org/10.1007/s10265-009-0291-z
75. Zheng S, Poczai P, Hyvönen J, et al (2020) Chloroplot: an online program for the versatile plotting of organelle genomes. Front Genet 11:576124. https://doi.org/10.3389/fgene.2020.576124

Figures

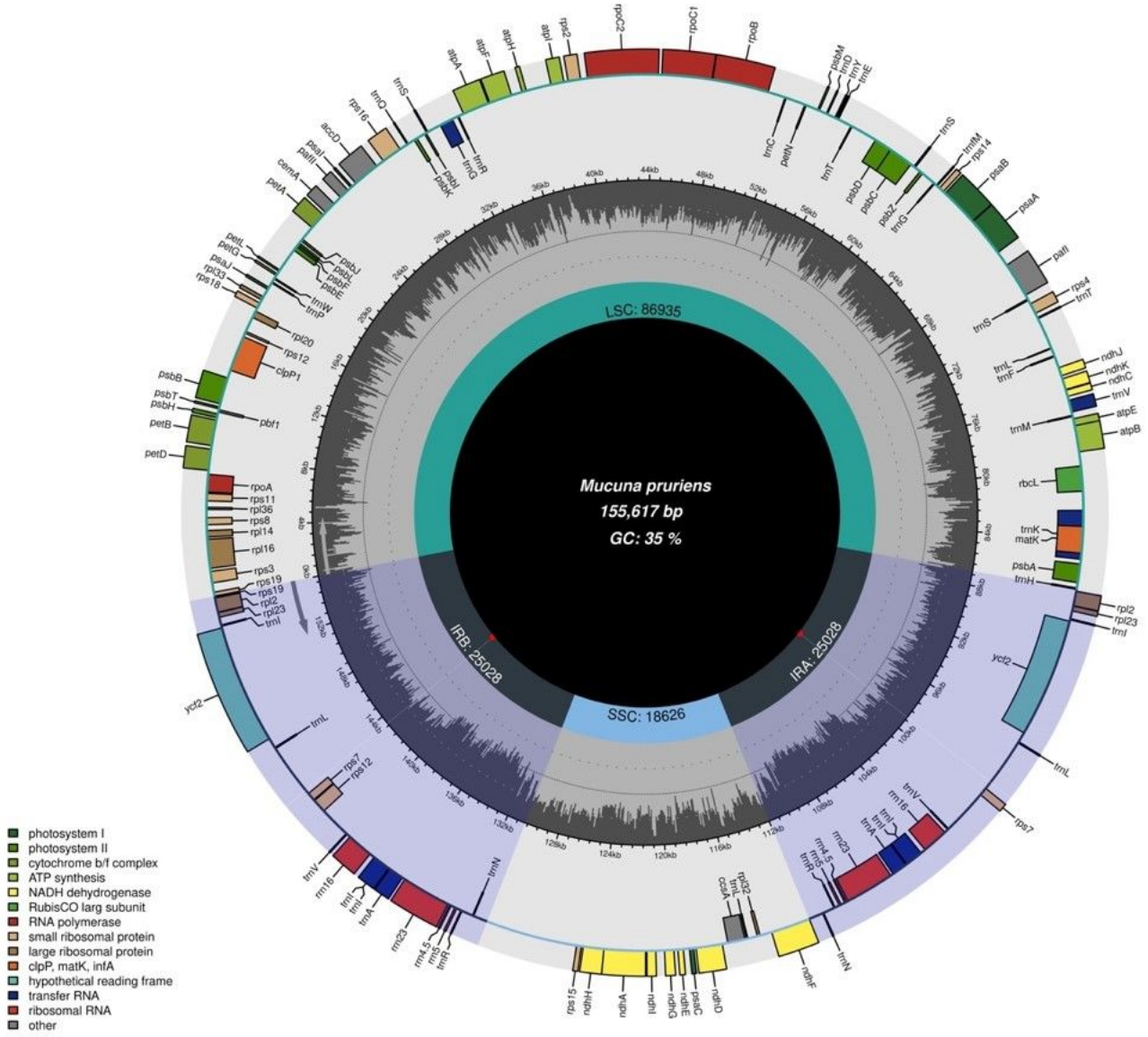

Figure 1

The provided diagram illustrates the structure of the chloroplast genome of *Mucuna pruriens*.

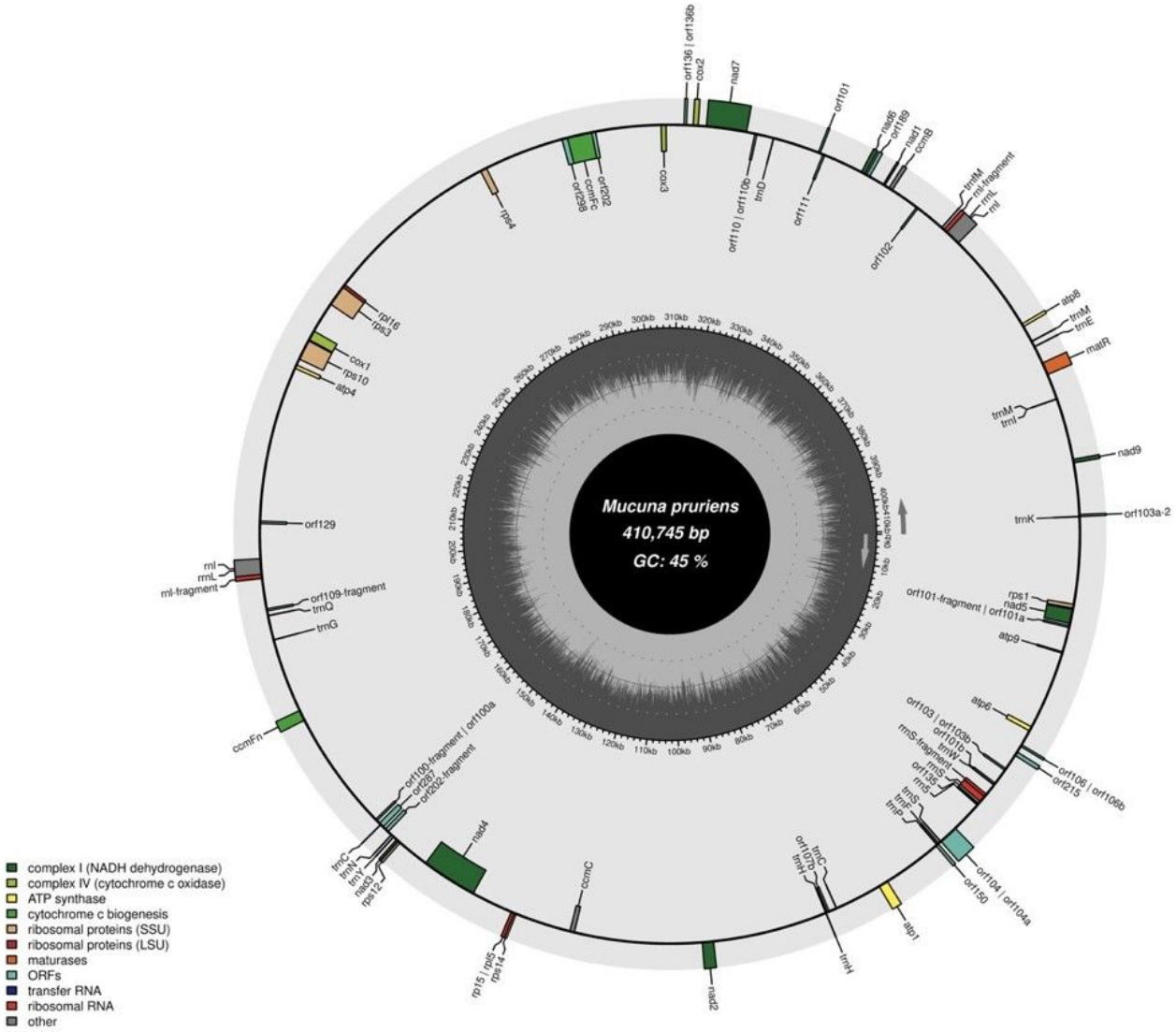

Figure 2

The provided diagram illustrates the structure of the mitochondrial genome of *Mucuna pruriens*.

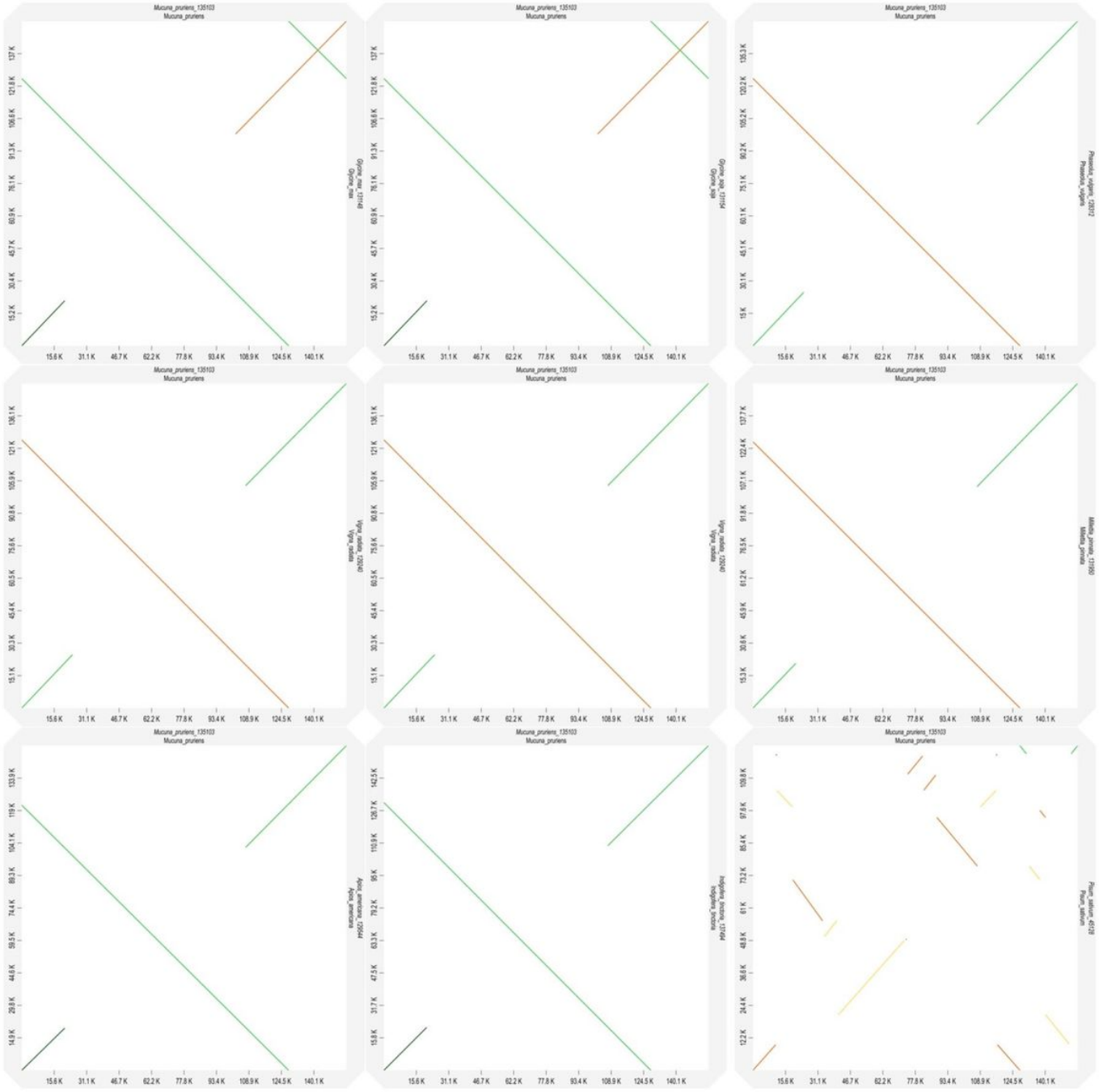

Figure 3

Regions of synteny between chloroplast genomes in Phaseoleae are represented by dot-plot graphs, with *Mucuna pruriens* serving as the reference.

Figure 4

Regions of synteny between mitochondrial genomes in Phaseoleae are represented by dot-plot graphs, with *Mucuna pruriens* serving as the reference.

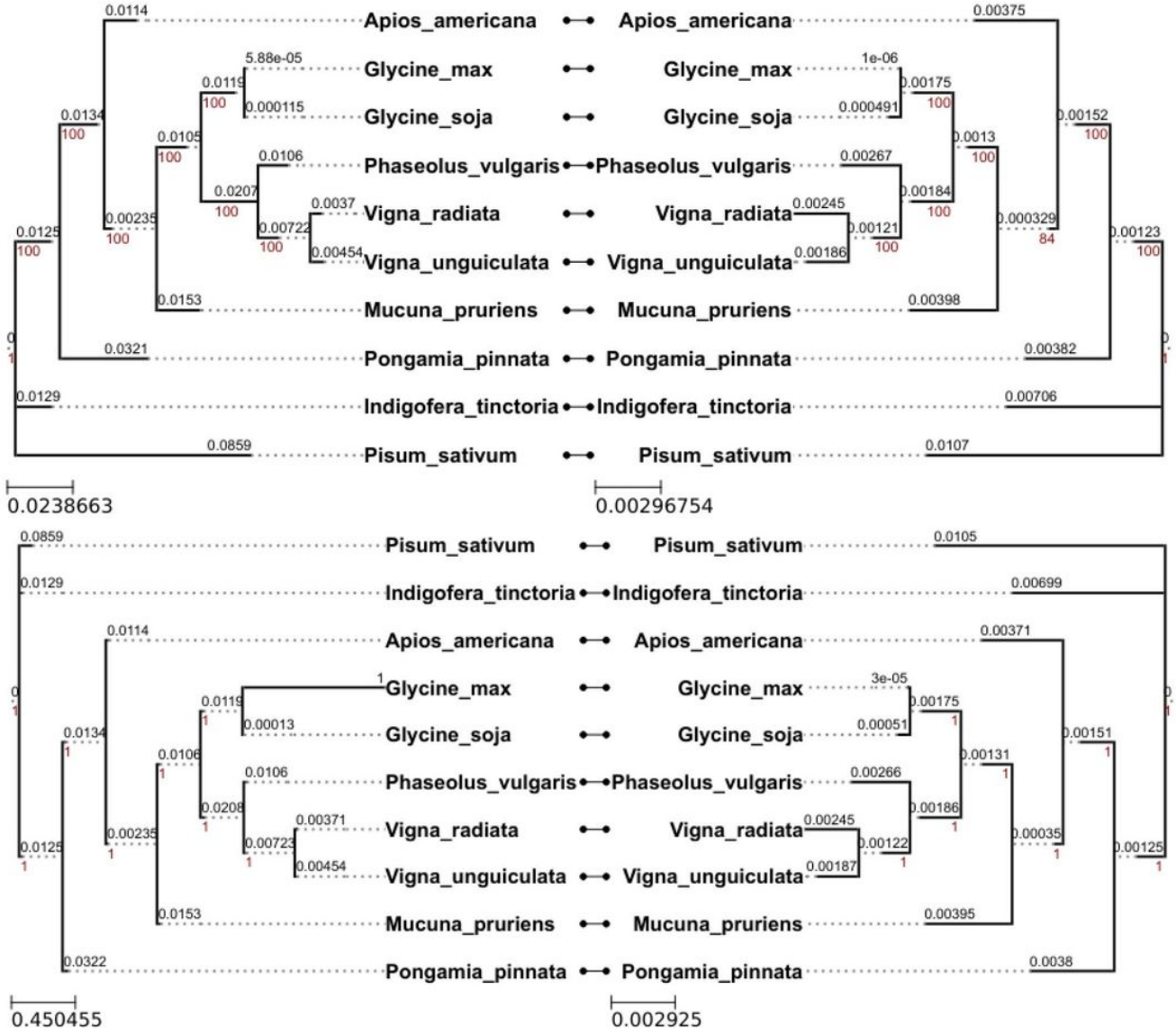

Figure 5

The phylogenetic relationships of the Fabaceae family were determined using chloroplast genes (on the left) and mitochondrial genes (on the right) through the application of Maximum Likelihood (top) and Bayesian Inference (bottom) methods.

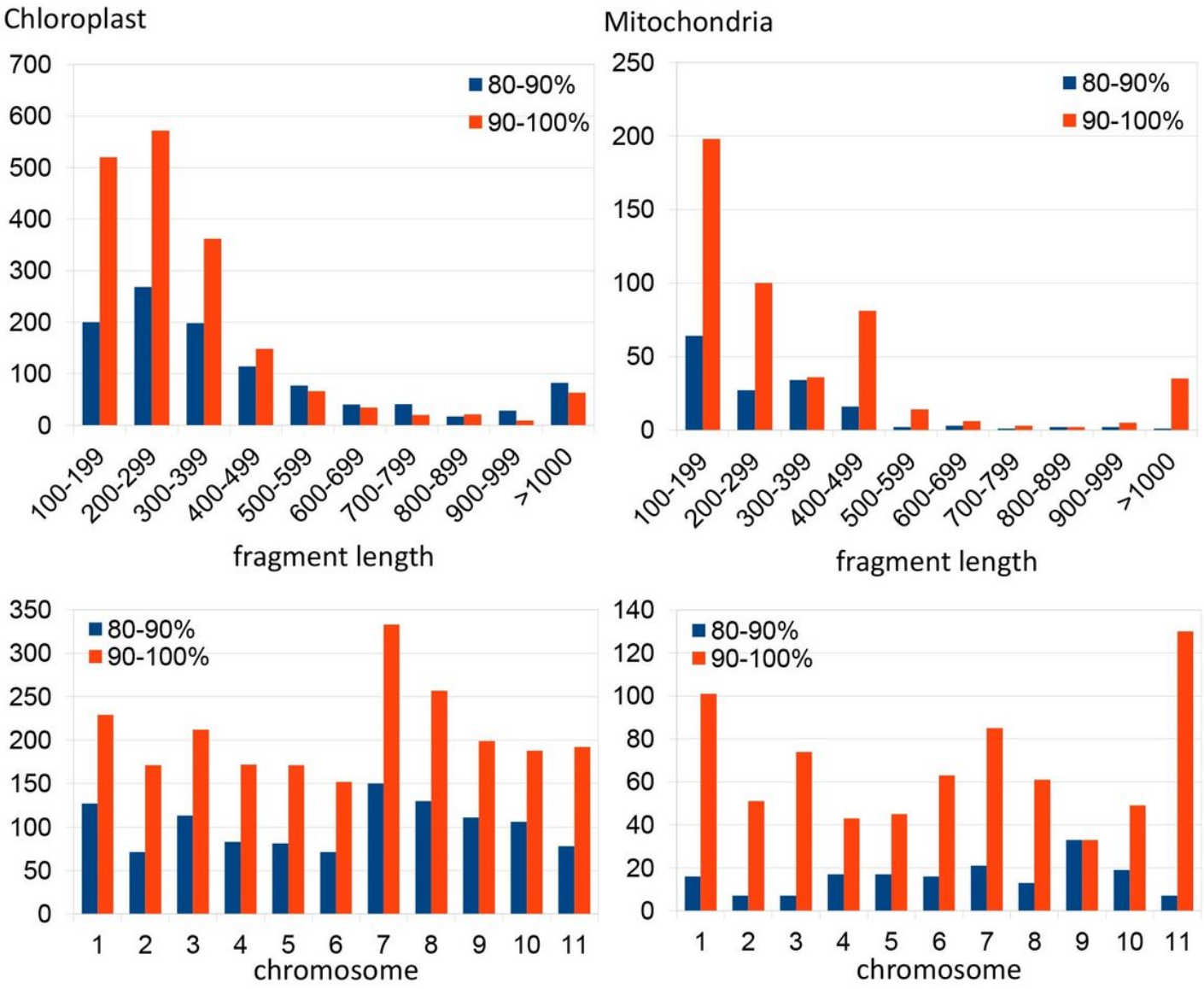

Figure 6

The number and distribution of transferred sequences from the chloroplast and mitochondrial genome to the nuclear genome by fragment size and chromosomal location.

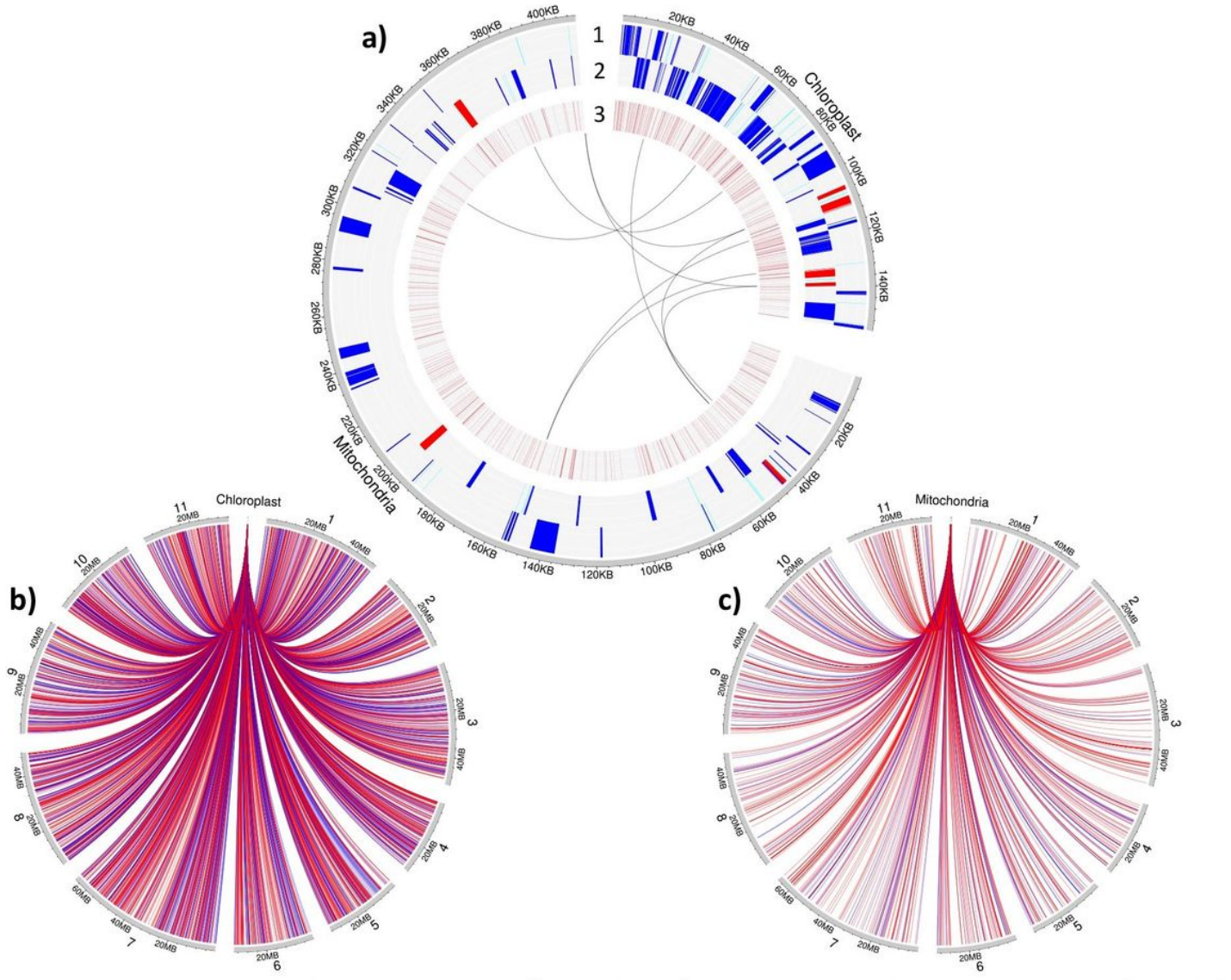

Figure 7

Sequence transfer between organellar and nuclear genomes. a) Sequences transferred from chloroplasts to mitochondria. The 1st, 2nd, and 3rd circles were the genes present in the positive strand, negative strand, and repeat sequences, respectively. The blue, red, and cyan colors in the 1st and 2nd circles represent protein, rRNA, and tRNA coding genes, respectively. b) Chloroplast 11 chromosomes of the nuclear genome. c) Mitochondrial sequences transferred to the 11 chromosomes of the nuclear genome. Blue and red colors represent fragments that have similarities of 80-90% and 90-100%, respectively.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- MucunapruriensMitogenomessupplementary.docx