Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Chloroplast Genome-Based Characterization of Cytochrome b_6f Complex Encoding Genes in the Indonesian Strawberry (Fragaria × ananassa 'Mencir')

Ganies Riza Aristya[1]

ganies_riza@ugm.ac.id

Universitas Gadjah Mada

Tika Nur Amini[2]

Universitas Gadjah Mada

Andreas Wibisono[2]

Universitas Gadjah Mada

Tan Rendy[2]

Universitas Gadjah Mada

Denianto Agung Wicaksono[2]

Universitas Gadjah Mada

Ani Widiastuti[3]

Universitas Gadjah Mada

Rina Sri Kasiamdari[4]

Universitas Gadjah Mada

Muhammad Fauzi Arif[5]

Mulawarman University

Research Article

Keywords: Chloroplast genome, Cytochrome b_6f complex, Fragaria × ananassa 'Mencir', Strawberries

Posted Date: September 24th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10793726/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: No competing interests reported.

Abstract

Background

Fragaria × *ananassa* 'Mencir' is a well-adapted strawberry cultivar from Indonesia with desirable agronomic traits, yet its genetic resources remain largely unexplored. In this study, we assembled and characterized the complete chloroplast genome of strawberry 'Mencir' using Oxford Nanopore long-read sequencing, with a specific focus on the cytochrome b_6f complex encoding genes (*petA, petB, petD, petG, petL*, and *petN*).

Results

The chloroplast genome is 155,568 bp in length, exhibiting a typical quadripartite structure with 37.23% GC content. Comparative genomic analysis revealed high synteny among *Fragaria* chloroplast genomes and identified hypervariable intergenic regions (*trnH-psbA, trnK-rps16*, and *petD* cluster) as potential DNA barcoding markers. Phylogenomic reconstruction confirmed the placement of strawberry 'Mencir' within the *Fragaria* clade, although specific inter-cultivar relationships remain unresolved due to high chloroplast genome conservation. Structural analysis of *pet* genes showed that *petA, petG, petL*, and *petN* are intronless, while *petB* and *petD* contain group II introns. Conserved motif and 3D protein structure analyses confirmed high conservation in core subunits PetA, PetB, and PetD (RMSD < 0.23 Å), while small subunits (PetG, PetL, PetN) exhibited greater structural flexibility consistent with their roles as stabilizing components. Genetic diversity analysis revealed low nucleotide diversity across all *pet* genes ($\pi < 0.003$), with *petG* and *petN* showing no variation ($\pi = 0.000$), while Ka/Ks analysis confirmed intensive purifying selection (Ka/Ks < 0.3) maintaining functional integrity across the complex. Codon usage analysis revealed a strong A/T-ending codon bias, reflecting mutational pressure as the primary driver.

Conclusions

These findings provide essential genomic resources for the Indonesian strawberry 'Mencir' and highlight that *pet* genes, due to their extreme conservation, are valuable for studying functional constraints rather than serving as phylogenetic markers. This study contributes to the growing body of chloroplast genomic data in horticultural crops.

Background

The genus *Fragaria* comprises 25 identified species worldwide, including the widely cultivated strawberry (*Fragaria* × *ananassa*), a high-value fruit commodity globally recognized for its exceptional organoleptic characteristics, nutritional value, and high vitamin C content [1–4], In tropical regions like Indonesia, strawberry cultivation possesses high economic value but faces distinct environmental challenges. Although strawberries can adapt to areas with specific rainfall, sunlight, and humidity parameters [5], commercial production in the tropics is largely confined to highlands where cooler temperatures, photoperiods, and genotypes interplay to determine plant performance [6]. Consequently, national

production remains unstable, despite an initial increase recorded from 2021 to earlier periods [7–9], production recently decreased from 28,895 tons in 2022 to 27,721 tons in 2023 [10, 11]. To stabilize yields, the utilization of well-adapted local cultivars is essential. The local cultivar, strawberry 'Mencir', a hybrid resulting from a cross between 'Festival' and 'California' [12, 13] has emerged as a primary target for development. Favored for its sweet flavor, large size, and juicy texture [14], strawberry 'Mencir' exhibits superior phenotypic plasticity, allowing it to adapt better to lower altitudes compared to other cultivars [15]. Such highly adaptable local strawberries possess significant potential to become crucial genetic resources in future breeding programs [16].

To uncover the molecular mechanisms underlying these adaptations without the complications of an octoploid nuclear genome, chloroplast genome analysis provides a robust and reliable alternative. Chloroplasts, which originated from free-living cyanobacteria, play essential roles in photosynthesis and carbon fixation [1, 17]. Unlike the nuclear genome, chloroplast DNA is not affected by changes in ploidy [1, 18]. Recent advances in whole-genome sequencing (WGS) have facilitated the accurate assembly of complete chloroplast genomes, enabling detailed analyses of structural variations, insertions/deletions (InDels), and nucleotide substitutions [19–21],. While this approach has been successfully applied to investigate evolutionary relationships in several *Fragaria* cultivars [22–24], comprehensive chloroplast genomic resources for Indonesian strawberry cultivars remain unreported, severely limiting the development of plastid-based molecular markers.

Investigating the chloroplast genome is particularly vital for understanding how tropical strawberries cope with environmental stresses, such as high light intensity and temperature fluctuations, which primarily affect photosynthetic electron transport. A key component in this physiological process is the cytochrome b_6f complex. Functioning as an electron transport bridge between photosystem II and photosystem I, this complex also contributes significantly to photoprotection through cyclic electron transport [25]. In *Fragaria*, the cytochrome b_6f complex is encoded by six chloroplast genes (*petA*, *petB*, *petD*, *petG*, *petL*, and *petN*) [1]. A comprehensive characterization of these functionally important genes, including their structural prediction, codon usage, and evolutionary relationships, can provide valuable insights into their molecular evolution and functional conservation during environmental adaptation [26–29].

To address these gaps, this study aims to sequence and characterize the complete chloroplast genome of the Indonesian strawberry 'Mencir' using Nanopore-based Whole Genome Sequencing. We conducted a comprehensive analysis focusing on overall genome structure (assembly, annotation, and GC content), Simple Sequence Repeats (SSRs), Inverted Repeat (IR) dynamics, and comparative genomic analyses using mVISTA, which included genome-wide RSCU and nucleotide diversity (π) assessments. Phylogenomic reconstruction using complete chloroplast genome data from 29 Rosaceae species was performed to elucidate the evolutionary placement of strawberry 'Mencir' within the family. Furthermore, this study specifically highlights a structural and evolutionary analysis of the cytochrome b_6f complex genes (*petA*, *petB*, *petD*, *petG*, *petL*, and *petN*) through evaluations of exon–intron organization, conserved motifs, three-dimensional protein structure predictions (AlphaFold3) with RMSD comparisons, codon usage bias (RSCU) patterns specific to *pet* genes, genetic diversity (nucleotide diversity, haplotypes, and Tajima's D), and selection pressure (Ka/Ks ratios). The findings of this research are expected to enrich organelle genome

databases and provide fundamental genomic information for Indonesian strawberries. This study also supports future genetic improvement, environmental adaptation studies, and molecular-based breeding programs.

Methods

Sample Collection

Fresh leaf samples of the strawberry 'Mencir' were collected from Karya Farm, located in Gamping Kidul, Sleman, Special Region of Yogyakarta, Indonesia (-7.802164,110.327760) at an altitude of approximately 114 m above sea level, immediately following collection, the leaf tissues were placed in sterile sealable plastic bags and stored in a portable box for transport to the laboratory. For long-term preservation prior to genomic analysis, the samples were transferred to the Laboratory of Genetics and Breeding, Faculty of Biology, Universitas Gadjah Mada, and maintained in a freezer at -20°C [19].

Formal taxonomic identification was performed by Tika Nur Amini. A representative voucher specimen (Voucher Specimen No. LST-UGM.2025.009) has been deposited in the herbarium of Laboratory of Plant Systematics, Faculty of Biology, Universitas Gadjah Mada

High Molecular Weight DNA Isolation and Quality Assessment

High molecular weight (HMW) genomic DNA was extracted from 400 mg of strawberry leaf tissue using a modified cetyltrimethylammonium bromide (CTAB) method. The leaf tissue was pulverized into a fine powder using liquid nitrogen. Cell lysis was performed using a pre-heated (60°C) 2× CTAB buffer (1.4 M NaCl, 0.1 M Tris-HCl, 20 mM EDTA, 2% w/v CTAB, pH 8.0) supplemented with 0.2% β-mercaptoethanol. To remove proteins and secondary metabolites, a chloroform–isoamyl alcohol (CIA; 24:1) extraction was performed. RNA contamination was eliminated by treatment with 3 µL of RNase A at 37°C for 15 min. The DNA was precipitated with cold isopropanol at − 20°C for 1 hour, washed with 70% ethanol, and finally resuspended in TE buffer. The purity and concentration of the extracted DNA were assessed using a NanoDrop Spectrophotometer (NanoDrop Lite Plus Thermo Scientific) and Qubit Fluorometer 4.0 then qualitatively analysed using agarose gel electrophoresis 0.8% and TapeStation (Agilent TapeStation Systems 4200) testing.

Library Preparation and Long Read Sequencing

Long read sequencing was performed using a Whole Genome Sequencing (WGS) approach on the Oxford Nanopore Technologies (ONT) platform. Genomic DNA libraries were prepared using the Ligation Sequencing Kit V14 (SQK-LSK114) starting with 1500 ng of gDNA. The initial steps involved DNA repair and end-preparation using the NEBNext FFPE DNA Repair Mix and Ultra II End-prep Enzyme Mix (New England Biolabs), followed by adapter ligation using the NEBNext Quick Ligation Module. The prepared library was subsequently loaded onto a PromethION Flow Cell (R10.4.1) and sequenced using a PromethION 24 device. Real-time basecalling was executed using the Dorado basecaller v7.11.2

(https://github.com/nanoporetech/dorado) in super-accuracy (SUP) mode. A minimum quality threshold of Q > 10 was applied, with automated adapter and barcode trimming enabled. The quality of the resulting raw data was evaluated using NanoPlot (https://github.com/wdecoster/NanoPlot) to ensure the reads met the standards for downstream assembly and comparative genomic analysis.

Chloroplast Genome Assembly and Annotation

The chloroplast genome of the strawberry 'Mencir' was de novo assembled using Raven (https://github.com/lbcb-sci/raven) from raw FASTQ and subsequently polished with Racon (https://github.com/isovic/racon). The resulting assembly graph (GFA) was visualized and validated in Bandage to confirm the formation of a single, closed circular contig [30]. Genome annotation was performed via GeSeq (CHLOROBOX) using BLAT-based homology searches for gene identification, followed by circular map generation using OGDRAW (https://chlorobox.mpimp-golm.mpg.de/OGDraw.html).

Reference Genomes Selection

Reference chloroplast genomes utilized in the present study were obtained from the NCBI Nucleotide database (https://www.ncbi.nlm.nih.gov/nucleotide/). Genome selection was performed according to the specific requirements of each analysis, with only complete chloroplast genomes being considered to guarantee data integrity. The reference genomes were downloaded in two distinct formats: GenBank format (.gb), which was employed for gene annotation, and FASTA format (.fasta), which was used for sequence alignment and comparative analyses. The complete chloroplast genome of *Fragaria* × *ananassa* Mencir' (GenBank accession: PZ351331.1), sequenced as part of this study, served as the central reference for all subsequent analyses. A comprehensive list of all reference genomes utilized is presented in Table S1 (Supplementary Data).

Comparative Genomic Analyses

Chloroplast simple sequence repeats (SSRs) were identified using MISA-web (https://webblast.ipk-gatersleben.de/misa/) with minimum repetition thresholds of 10 units for mononucleotides, 6 for dinucleotides, and 5 for tri-, tetra-, penta-, and hexanucleotides. Codon usage bias was analysed by extracting protein-coding genes (PCGs) using Python-based SeqIO, followed by Relative Synonymous Codon Usage (RSCU) calculation via RSCUcaller. Sequence divergence was visualized using mVISTA (https://genome.lbl.gov/vista/index.shtml) in Shuffle-LAGAN mode by comparing strawberry 'Mencir' against several *Fragaria* reference genomes. Structural dynamics at the Inverted Repeat (IR) boundaries (LSC/IRb, IRb/SSC, SSC/IRa, and IRa/LSC) were assessed using IRplus (https://irscope.shinyapps.io/IRplus/) to identify potential IR expansions or contractions. Finally, genomic variations, including Single Nucleotide Polymorphisms (SNPs) and Insertion-Deletions (InDels), were identified by aligning sequences with MAFFT v7 (https://mafft.cbrc.jp/) and analysing the polymorphic sites using DnaSP6 [31].

Phylogenomic Analysis

Complete chloroplast genomes of 27 Rosaceae species and two Moraceae species were retrieved from the NCBI database, with the two Moraceae species designated as the outgroup. Multiple sequence alignment was performed using MAFFT [32]. The resulting alignment was manually inspected and adjusted where necessary using AliView [33] to ensure the accuracy of the aligned sites. The best-fit nucleotide substitution model was determined using ModelTest-NG [34], which identified GTR + I+G4 as the optimal model based on the Bayesian Information Criterion (BIC). Phylogenetic reconstruction was conducted via the Maximum Likelihood (ML) method using the RAxML-NG program [35]. Finally, the reliability of the tree topology was assessed with 1,000 bootstrap (BS) replicates.

Gene Structure and Conserved Motif Analysis of the pet Genes

Gene structure information was obtained from genome annotation files in GFF3 format and visualized using TBtools based on CDS and intron arrangements [36]. The structures analyzed were the exons and introns of chloroplast *pet* genes of *Fragaria* × *ananassa* 'Mencir' were analyzed together with their homologs from genus *Fragaria* and *Rubus idaeus* as an outgroup. Conserved motifs within the *F.*× *ananassa* 'Mencir' Pet protein sequences were identified using the MEME Suite 5.5.7 (https://meme-suite.org/meme/tools/meme). Motif analysis was conducted using a maximum number of 10 motifs, whereas all other parameters were maintained at their default values. The Pet proteins from the strawberry 'Mencir' were analyzed alongside their homologs from other *Fragaria* × *ananassa* cultivars, various *Fragaria* species, and *Rubus idaeus* as an outgroup.

3D Structure Prediction of the pet Genes

The protein sequence of *pet* genes strawberry 'Mencir' was extracted from chloroplast genome annotation results using UGENE 53.0 [37]. Three-dimensional protein structure modeling was performed using AlphaFold3 (https://alphafoldserver.com/) [28, 38]. The structural similarity was evaluated by counting root mean square deviation (RMSD) in PyMOL. RMSD values were calculated to compare Pet protein structures between *F.* × *ananassa* 'Mencir' and with 10 reference sequences with 9 from the genus *Fragaria* and 1 from the genus *Rubus* as an outgroup. Complete chloroplast genome sequences of the *F.* × *ananassa* 'Benihoppe', 'Royal Royce', and 'Camarosa' were retrieved from the NCBI Nucleotide database under accession numbers NC_035961.1, ON478194.1, and ON478181.1.

Codon Usage Bias Analysis

To investigate codon usage patterns, codon usage bias analysis was performed using Python 3 within the Anaconda environment. The annotated chloroplast genome of *Fragaria* × *ananassa* 'Mencir' in GenBank format (.gb) was used as the input. Protein-coding sequences (CDS) were extracted using the SeqIO module from the Biopython library [39], with sequences shorter than 60 nucleotides excluded from the analysis to avoid potential artifacts. The target genes belonging to the cytochrome b_6f complex (*petA*, *petB*, *petD*, *petG*, *petL*, and *petN*) were identified through pattern matching using the re library, based on the 'gene' and 'product' feature annotations in the GenBank file.

Relative synonymous codon usage (RSCU) values were calculated using the collections and numpy libraries. RSCU was determined by comparing the observed frequency of each codon to its expected frequency under the null hypothesis of equal usage among synonymous codons. The mean RSCU values were statistically compared between two groups: (1) *pet* genes and (2) all other plastome protein-coding genes using the pandas library. Results were exported as a CSV file for further analysis [40]. A heatmap visualizing RSCU values across all codon families was generated using the matplotlib library [41] to identify patterns of codon preference. Additionally, the complete set of plastid-encoded transfer RNA (tRNA) genes in the *F.* × *ananassa* 'Mencir' chloroplast genome was catalogued and visually inspected using UGENE v53.0 [37].

Genetic Diversity Analysis

SNP and InDel analyses of chloroplast *pet* genes were conducted using DnaSP v.6.12.03 [31]. Only sequences from genus *Fragaria* were included to reduce variation caused by distant evolutionary divergence. Parameters including the number of polymorphic sites (S), haplotypes (h), haplotype diversity (Hd), nucleotide diversity (π), average number of nucleotide differences (k), InDel variation, and Tajima's D were calculated using default settings in DnaSP.

Ka/Ks Ratio Analysis

The codon sequences of six plastid *pet* genes from 19 *Fragaria* species and one *Drymocallis saviczii* species were aligned using the MUSCLE method in MEGA X based own codons. Ambiguous sequence parts or those containing gaps were trimmed to ensure the generated data is of good quality. The Ka (nonsynonymous substitution) and Ks values for each pair of species were estimated using the modified Nei-Gojobori method with DnaSP v.6.12.03 [31], resulting in an Excel output file. Data visualization uses the Python language, the pandas library to convert decimal number formats into standard numeric formats, and then the Ka/Ks values are calculated for each comparison. Only comparisons with Ks > 0 and Ka/Ks < 2 are used to maintain the quality of the obtained data. Descriptive statistics (mean, median, standard deviation, min, max, and n) for each gene are calculated and stored in CSV format. Interactive graphs are created using the Plotly library, combining violin plots and jitter plots. The violin plot shows the density distribution of Ka/Ks values, while the jitter plot displays individual data points. Reference lines at Ka/Ks = 1.0 and 0.5 were added to facilitate interpretation. The graph is equipped with labels, titles, and a legend that includes the sample size. The final results are exported in SVG and PNG formats [40] .

Result

Long Read Sequencing Data

The PromethION sequencing run yielded 1,894,861 reads with a total throughput of 4.57 Gb, featuring a mean read length of 2,412 bp, a median of 916 bp, a read N50 of 5,634 bp, and a maximum length of 423,602 bp (Table S2). The total base accumulation was dominated by fragments under 25 kb (Figure S1). Visual analysis of read quality and length (Figure S2) demonstrated a stable correlation between fragment size and base accuracy, with most reads concentrated within a Quality Score (Q-score) range of 15 to 25.

General Characteristic of Strawberry ‘Mencir Chloroplast Genome

The chloroplast genome of strawberry ‘Mencir’ was assembled and annotated (Fig. 1). The genome size of ‘Mencir’ was 155,568 bp in length and exhibiting a typical quadripartite structure characteristic of angiosperms. It comprises a Large Single Copy region (LSC: 85,551 bp), a Small Single Copy region (SSC: 18,145 bp), and a pair of Inverted Repeats (IR: 25,936 bp each). The overall GC content averaged 37.23%, with substantial variation observed across the individual regions. Specifically, the GC content of the IR regions (42.85%) was markedly higher than that of the LSC (35.12%) and SSC (31.15%) regions. This elevated GC content within the IR regions is a common feature driven primarily by the high density of ribosomal RNA (rRNA) genes, which reached a GC content of 53.53% in this genome (Table S6).

Comparative Chloroplast Genome Analysis

Analysis of Simple Sequence Repeats (SSRs) in the *F.* × *ananassa* 'Mencir' chloroplast genome revealed a strong dominance of mononucleotide repeats, which accounted for 37 of the 42 identified loci (Fig. 2a). These repeats were driven almost entirely by A/T motifs, comprising 20 T, 16 A, 1 G, and 0 C loci (Fig. 2b, Table S2). Furthermore, the total SSR abundance and spatial arrangement in strawberry 'Mencir' remain highly conservative when compared to reference cultivars ('Benihoppe', and 'Royal Royce'), as well as related *Fragaria* species (*F. chiloensis, F. virginiana, F. vesca*) and the outgroup *Malus domestica* (Fig. 4b).

The Relative Synonymous Codon Usage (RSCU) values was estimated to characterize codon preference in the chloroplast genome of strawberry ‘Mencir’. In these chloroplast genome, analysis of 87 protein-coding genes (80,877 bp, 26,959 codons) revealed Leucine (10.65%, 2,872 codons) and Cysteine (1.15%, 312 codons) as the most and least frequent amino acids, respectively. RSCU values ranged from 0.39 (GAC) to 1.95 (TTA). A strong preference for A/U-ending codons was evident: 29 of 30 high-frequency codons (RSCU > 1) ended in A or U (with TTG as the sole exception), whereas 29 of 32 low-frequency codons (RSCU < 1) ended in C or G. Furthermore, UAA was the dominant stop codon (RSCU = 1.72), while ATG (Met) and UGG (Trp) showed no bias (RSCU = 1.0), consistent with established plant chloroplast trends [42].

Comparative analysis showed that the CUB profile is highly conserved among 'Mencir', 'Benihoppe', 'Camarosa', and 'Royal Royce' (Fig. 3), indicating that codon bias in *F.* × *ananassa* is governed by species-level evolutionary constraints rather than cultivar-specific breeding pressures. All cultivars displayed a marked preference for A and U at the third codon (wobble) position, such as GCT (Ala) and AGA (Arg), a hallmark of dicotyledonous genomes characterized by low GC3 content [43].

To elucidate the genomic characteristics and evolutionary divergence of the Indonesian strawberry 'Mencir', a comparative analysis of its chloroplast genome was performed against two global cultivars ('Benihoppe' and 'Royal Royce'), three wild *Fragaria* species (*F. chiloensis*, *F. virginiana*, and *F. vesca*), and one outgroup (*Malus domestica*). Global alignment visualized using mVISTA demonstrated a high degree of sequence conservation across all *Fragaria* taxa, exhibiting sequence identities near 100% throughout most regions

(Figure S3). In contrast, *Malus domestica* displayed extensive sequence divergence across both coding and non-coding regions, confirming its evolutionary distance as a Rosaceae outgroup [44].

Genetic variability within the *Fragaria* plastomes is predominantly concentrated within non-coding intergenic spacer (IGS) regions, represented by distinct visual troughs in the alignment plot. High levels of polymorphism were observed in intergenic regions such as *trnH-GUG* to *psbA*, *trnK-UUU* to *rps16*, and the *petD* cluster. While exonic regions generally displayed high conservation, certain protein coding genes exhibited pronounced divergence, most notably *ycf1* and *ycf2*. The high frequency of SNPs and InDels within *ycf1* and *ycf2* positions them as mutational hotspots in the *Fragaria* plastome, driven by their large sequence length and relaxed selection pressure near the IR boundaries [1, 45].

Our results demonstrate a high degree of structural conservation at the IR boundaries among *F. × ananassa* cultivars and across the *Fragaria* genus (Fig. 4). At the JLB junction, all analysed strawberry cultivars, including strawberry 'Mencir', exhibited the *rps19* gene entirely within the LSC region without expansion into the IRb. This pattern was consistent in *F. chiloensis*, *F. vesca*, and *F. virginiana*, indicating a lack of IR expansion toward the LSC within this genus. In contrast, the outgroup *Malus domestica* showed *rps19* partially located within the IRb, reflecting a genus-specific IR expansion within the Rosaceae family [46].

Similarly, the JSB and JSA junctions displayed nearly identical configurations across all strawberry genomes. The *ycf1* gene consistently crossed the SSC/IR boundaries with uniform partial fragment lengths with approximately 12 bp at the JSB and 1,092–1,104 bp at the JSA. No significant variation in *ycf1* fragment length was detected among *F. × ananassa* cultivars, suggesting that structural dynamics in this region are extremely limited at the intraspecific level. Furthermore, at the JLA junction, the *trnH* gene was consistently located within the LSC region with a stable distance from the junction across all *Fragaria* taxa.

Analysis of nucleotide diversity π across 79 chloroplast protein-coding genes among 11 Rosaceae species (*Fragaria × ananassa* 'Mencir' as the reference sequence) revealed an overall mean value of 0.0272, ranging from 0.0 to 0.0751 (Fig. 5). The highest nucleotide diversity was observed in the *psbI* gene ($\pi = 0.0751$), followed by the ribosomal protein gene *rpl22* ($\pi \approx 0.062$), the phylogenetic marker gene *matK* ($\pi \approx 0.058$), and the NADH dehydrogenase subunit gene *ndhF* ($\pi \approx 0.058$). The mean diversity of $\pi = 0.0272$ across the intergeneric Rosaceae scale is considered moderate to high compared to intraspecific or infrageneric variation within *Fragaria* ($\pi < 0.002$). The distribution of these hypervariable genes is predominantly concentrated in the Large Single Copy (LSC) and Small Single Copy (SSC) regions.

Phylogenomic Analysis

Based on the generated Maximum Likelihood (ML) phylogenetic tree (Fig. 6), the Rosaceae family is clearly partitioned into three major subfamilies: Amygdaloideae, Dryadoideae, and Rosoideae, all supported by maximum bootstrap values (100%). The topology reveals that the genus *Fragaria* is more closely related to *Dasiphora fruticosa* and *Potentilla micrantha* than to the *Rosa* or *Rubus* lineages, providing additional support for the circumscription of the tribe Potentilleae within Rosoideae [47]. The Amygdaloideae clade (green block), comprising *Pyrus*, *Malus*, and *Prunus*, serves as a sister group to the Dryadoideae and Rosoideae clades, reinforcing the distinct evolutionary trajectories of pome and stone fruits.

Structural and Conserved Motif Analysis of pet Genes

Gene structure of *pet* genes chloroplast genomes were analyzed using TBtools to determine their exon-intron organization. In the resulting visualization (Fig. 7A), exons are represented as green boxes, while introns are depicted as black lines connecting these boxes. Based on this representation, two distinct structural patterns were observed among the *pet* genes. The genes *petA*, *petG*, *petL*, and *petN* were found to be intronless, containing only a coding DNA sequence (CDS). The composition of conserved motifs is stable of each *pet* genes across *Fragaria* species, including *F. × ananassa* 'Mencir', indicating that these motifs are highly conserved throughout evolution shown in Fig. 7B, respectively.

Predicted Three-Dimensional Structures of the Cytochrome b_6f Proteins

Three-dimensional (3D) structure prediction of six cytochrome b_6f complex proteins (PetA, PetB, PetD, PetG, PetL, and PetN) from *F. × ananassa* 'Mencir' using AlphaFold3 (Fig. 8). The predicted structures were highly similar to those of other strawberry cultivars ('Benihoppe', 'Royal Royce', and 'Camarosa'). The core subunits PetA, PetB, and PetD showed high pTM scores ranging from 0.83 to 0.93. In contrast, the smaller subunits PetG, PetL, and PetN had lower pTM scores ranging from 0.49 to 0.55. Root mean square deviation (RMSD) analysis was performed to measure structural similarity (Table 1). The core proteins PetA, PetB, and PetD exhibited low RMSD values, with averages of 0.2047 Å, 0.1005 Å, and 0.2285 Å, respectively, across all comparisons. PetG, PetL, and PetN, which are small subunits of the cytochrome b_6f complex, have higher average RMSD values than the large subunit. PetL shows the highest average RMSD value. The high predicted TM-score (pTM) values and low root mean square deviation (RMSD < 1 Å) of the core subunits, PetA, PetB, and PetD, indicate that their three-dimensional structures are well conserved.

Table 1
RMSD Comparison of *pet* Genes in *F.* × *ananassa* 'Mencir' Against Other Cultivars and Species within the Rosaceae Family.

Fragaria* × *ananassa **'Mencir' Indonesia PZ351331.1**	**PetA**	**PetB**	**PetD**	**PetG**	**PetL**	**PetN**
Fragaria × *ananassa* 'Benihoppe' China NC_035961.1	0,204	0,082	0,173	0,815	1,329	0,488
Fragaria × *ananassa* 'Royal Royce' Unknown ON478194.1	0,186	0,090	0,177	0,239	1,269	0,529
Fragaria × *ananassa* 'Camarosa' Unknown ON478181.1	0,226	0,069	0,242	0,515	0,503	0,575
Fragaria chiloensis Unknown MW537844.1	0,172	0,098	0,199	0,401	0,677	0,468
Fragaria virginiana USA NC019602.1	0,230	0,154	0,222	0,290	0,943	0,370
Fragaria vesca subsp. *vesca* Unknown NC_015206.1	0,154	0,103	0,381	0,541	1,130	0,468
Fragaria moschata Unknown MZ702811.1	0,218	0,113	0,240	0,341	0,939	0,536
Fragaria pentaphylla Unknown KY434061.1	0,341	0,114	0,208	0,250	1,434	0,369
Fragaria nilgerrensis Unknown MK560340.1	0,131	0,100	0,196	0.443	1,368	0,409
Rubus idaeus China OR972717.1	0,185	0,082	0,247	0,493	0,966	0,494
RMSD Mean	**0,2047**	**0,1005**	**0,2285**	**0,4328**	**1,0558**	**0,4706**
Codon Usage Analysis of the pet Genes						

Relative Synonymous Codon Usage (RSCU) analysis was performed to identify the codons preferred by the *pet* genes. In the resulting heatmap (Fig. 9), the *pet* genes depict a distinct codon usage pattern compared to other chloroplast genes in the *F.* × *ananassa* 'Mencir' chloroplast genome. Strong codon usage bias is indicated by codons ending in adenine (A) or thymine (T).

Frequently used codons with RSCU values greater than 1 in the *pet* genes include TTA (Leu), ATT (Ile), ATA (Ile), TCT (Ser), CCT (Pro), ACT (Thr), and GCT (Ala). In contrast, codons ending in guanine (G) or cytosine (C), such as TTC (Phe), CTG (Leu), ATC (Ile), and GGC (Gly), were under-represented with RSCU values lower

than 1 in the *pet* genes and classified as rare codons. The "Difference (PET – Other)" column in the analysis identified codons with significantly higher usage in *pet* genes relative to other chloroplast genes, with notable positive difference values for codons including CCT (Pro), GCT (Ala), TCA (Ser), and CGT (Arg). Non-synonymous codons for methionine (ATG) and tryptophan (TGG), as well as stop codons, were excluded from the biological interpretation of usage bias.

Genetic Diversity Analysis

Analysis of genetic diversity in chloroplast *pet* genes revealed different levels of sequence variation among the analyzed loci (Table 2). Among all genes, *petD* showed the highest nucleotide diversity value (π = 0.00253) and the highest average number of nucleotide differences (k = 1.222). In contrast, *petG* and *petN* showed no nucleotide variation, with π and k values equal to 0.000.

The highest nucleotide diversity and average number of nucleotide differences of *petD* indicating that this gene possessed the greatest sequence variation within the analyzed *Fragaria* taxa meanwhile in *petG* and *petN* showed no nucleotide diversity indicating highly conserved sequences. The *petB* gene also exhibited relatively high genetic variation, with three polymorphic sites, four haplotypes, haplotype diversity (Hd) of 0.711, and one detected insertion/deletion (InDel). Similarly, *petD* contained four polymorphic sites and four haplotypes, further supporting its relatively informative nature compared to other *pet* genes.

Table 2
Genetic diversity and InDel variation of chloroplast *pet* genes in *Fragaria* species

Gene	Length (bp)	S (SNP sites)	h (haplotype)	Hd	π (Pi)	K	InDels (I)	Indel haplotype	Tajima's D
petA	963	3	3	0.378	0.00078	0.756	0	–	–
petB	654	3	4	0.711	0.00147	0.956	1	2	0.819 (ns)
petD	483	4	4	0.711	**0.00253**	**1.222**	0	–	–
petG	114	0	1	0.000	0.00000	0.000	0	–	–
petL	96	1	2	0.200	0.00208	0.200	0	–	–
petN	96	0	1	0.000	0.00000	0.000	1	2	-1.111 (ns)

Ka/Ks Analysis

The analysis of Ka/Ks ratios for six *pet* genes (*petA*, *petB*, *petD*, *petG*, *petL*, and *petN*) across 19 *Fragaria* species and one *Drymocallis* species revealed a highly conservative evolutionary pattern among genes encoding components of the cytochrome b_6f complex (Fig. 10). Overall, the Ka/Ks values detected for genes with valid data (*petA*, *petB*, *petD*, and *petL*) were substantially below the 0.5 threshold, with the majority of values concentrated below 0.3 and mean values approaching zero. The absence of valid data for two genes, *petG* and *petN* (indicated as "No data"), represents a significant finding in this study. The inability to calculate

Ka/Ks ratios for these genes was attributable to Ks = 0 across all pairwise comparisons, indicating that no synonymous substitutions were detected among the analyzed species.

Discussion

Chloroplast genome structure

Although moderate relative to the platform's theoretical capacity [48], the throughput of chloroplast genome strawberry ‘Mencir’ still provided sufficient coverage for medium-scale genomes, with an N50 exceeding 5 kb supports assembly contiguity in complex regions better than short-read technologies [49]. The disparity between mean and median lengths indicates a right-skewed distribution caused by short fragment shearing or sample degradation during library preparation [50]. The weighted histogram reveals a persistent "long tail" of reads exceeding 400 kb [48], which is critical for resolving genomic gaps and structural variations in repetitive regions [49].

Unlike short-read technologies that suffer from end-of-read quality decay, Nanopore sequencing maintains consistent accuracy regardless of fragment length [51]. The high density of reads consistently above the Q10 threshold (> 90% accuracy) ensures high reliability for downstream basecalling and mapping. Furthermore, the presence of ultra-long reads reaching near Q20 (~ 99% accuracy) reflects successful optimization of flow cell chemical parameters, preserving signal integrity during DNA translocation [51].

Assembly and Annotation of the Chloroplast Genome

This genome size falls well within the conserved range reported for land plant chloroplasts, which typically ranges between 120 and 170 kb [17]. The high GC content of rRNA genes and IR regions contributes to the structural stability of the chloroplast genome via stronger hydrogen bonding, acting as a mechanism to minimize mutational rates in these highly conserved functional regions [52, 53].

Comparative Chloroplast Genome Analysis

This pronounced A/T bias is a widespread feature of angiosperm plastomes, primarily attributed to the naturally low GC content and elevated mutation rates prevailing within non-coding chloroplast regions [54]. Spatially, over three-quarters of the identified SSRs are concentrated within the Large Single Copy (LSC) region and intergenic spacers (IGS). This distribution highlights the evolutionarily dynamic nature of LSC/IGS regions, which tolerate higher rates of short-repeat accumulation due to lower selective pressure compared to Inverted Repeats or protein-coding genes [55]. The total SSR abundance and spatial arrangement in strawberry 'Mencir' reflecting a deeply conserved preference for A/T-based SSRs across these lineages.

Codon usage bias (CUB) plays a pivotal role in translational efficiency, protein production, and folding [56]. The overall CUB pattern in 'Mencir' is strongly biased toward A/U-ending codons, aligning with the general nucleotide composition of plastid genomes. The pronounced preference for the TTA codon and the UAA terminator suggests that mutational pressure driving higher A/U content is the primary factor shaping codon usage in this cultivar [57, 58]. In land plant plastomes, this A/U enrichment reflects an evolutionary

strategy to preserve gene expression stability and optimize translational efficiency, where mutational bias typically exerts a stronger influence than translational selection [42]. This strong profile similarity indicates that the core genomic mechanisms maintaining protein synthesis efficiency in strawberry 'Mencir' remain remarkably stable relative to introduced cultivars despite differences in agro-climatic adaptation.

Across all evaluated taxa, the Inverted Repeat (IR) regions exhibited a markedly higher level of conservation compared to the Large Single Copy (LSC) and Small Single Copy (SSC) regions. This structural stability aligns with the general features of angiosperm plastomes, where IR regions undergo a copy correction mechanism that protects nucleotide sequences against rapid mutation [59, 60].

Non-coding intergenic spacer (IGS) regions are characterized by an accumulation of single nucleotide polymorphisms (SNPs) and insertions or deletions (InDels), reflecting elevated nucleotide substitution rates [61]. Notably, the divergence within the *petD* cluster (primarily localized in its non-coding intronic and intergenic regions) highlights unique sequence variations in strawberry 'Mencir' compared to reference taxa [1]. Because these non-coding variations are often linked to structural dynamics near the LSC boundary, these non-coding loci provide a potential genetic targets for future DNA barcoding, which require further wet-lab validation.

Overall, the mVISTA alignment confirms that strawberry 'Mencir' shares strong genomic affinity with cultivated and wild *Fragaria* species while maintaining distinct polymorphic signatures. Identifying these hypervariable IGS loci and exonic hotspots enriches our understanding of strawberry plastome architecture and facilitates the development of high resolution molecular markers for marker assisted breeding and germplasm conservation.

The analysis of chloroplast genome architecture, specifically the junctions between Inverted Repeats (IR) and Single Copy (LSC/SSC) regions, is essential for understanding plastome structural evolution. In this study, we utilized IRscope to comparatively visualize the four primary junction sites: JLB (LSC/IRb), JSB (IRb/SSC), JSA (SSC/IRa), and JLA (IRa/LSC) across *Fragaria* cultivars [62]. While angiosperm plastomes are generally conservative, the expansion or contraction of IR boundaries can lead to gene duplication, the creation of pseudogenes, or significant changes in total genome length [17].

In summary, the strawberry plastome exhibits a highly conservative IR boundary structure, with no evidence of significant expansion or contraction among cultivars. The minimal variation observed in total chloroplast genome length does not correlate with junction shifts. These findings suggest that macro-structural plastome changes are not the primary driver of differentiation at the strawberry cultivar level. Consequently, molecular discrimination of these cultivars is more effectively achieved through sequence-level variations, such as SNPs and InDels, rather than IR boundary characteristics [60]. The degree of nucleotide diversity reflects taxonomic divergence at the intergeneric and inter-subfamily levels within the family Rosaceae (subfamilies Rosoideae and Amygdaloideae). Conversely, several core genes exhibited a value of $\pi = 0.0$, indicating strong purifying selection acting to maintain vital chloroplast functions.

The *matK* and *rpl22* genes in the LSC region are well known for their rapid rate of nucleotide substitution accumulation across flowering plant genera [63]. Meanwhile, the elevated π values observed in *ndh* genes

(*ndhF, ndhA, ndhH*) within the SSC region reflect a tendency for Rosaceae plastomes to undergo relaxed selective pressure and susceptibility to micro-recombination events [17, 52]. The emergence of an extreme variation peak in psbI ($\pi = 0.0751$) suggests that although the Photosystem II complex is generally highly conserved, the small transmembrane subunit psbI tolerates high sequence flexibility across Rosaceae genera without disrupting Photosystem II dimerization stability.

A specific focus on the cytochrome b_6f complex (the *pet* gene family) reveals a differential pattern of conservation among its subunits. The *petG* ($\pi \approx 0.034$), *petA* ($\pi \approx 0.032$), and *petB* ($\pi \approx 0.031$) subunits exhibited diversity levels above the genomic average, indicating an accumulation of mutations tolerated on the outer electron-capturing domains. In contrast, the coding sequences (CDS) of the *petL* ($\pi \approx 0.021$), *petN* ($\pi \approx 0.018$), and *petD* ($\pi \approx 0.018$) subunits showed significantly tighter sequence constraints to preserve hydrophobic inter-protein interactions [64]. While the non-coding regions flanking *petD* accumulate polymorphic variations suitable for barcode marker development (as observed in our mVISTA analysis), its protein-coding exon remains strictly conserved under purifying selection to preserve Cytochrome b_6f structural integrity. This differential diversity pattern underscores that chloroplast conservation levels in Rosaceae are governed primarily by functional protein constraints rather than genomic coordinate positions, while providing a foundation for selective pressure analysis (dN/dS) and protein structure modeling of *pet* genes in the strawberry 'Mencir' chloroplast.

Phylogenomic Analysis

The results of phylogenomic analysis are consistent with the established Rosaceae classification by Potter *et al.* (2007) and the APG IV system (2016), which recognize these clades as monophyletic groups. Within this framework, the sample strawberry 'Mencir' is nested within the Rosoideae clade (red block), showing the closest phylogenetic affinity to 'Royal Royce' (ON478194.1). Although the bootstrap support at this specific node is relatively low (33%), their clustering within the broader *Fragaria* clade (*F. chiloensis*, *F. virginiana*, and *F. vesca*) firmly validates the molecular identity of strawberry 'Mencir' as a member of the strawberry hybrid complex.

The use of *Artocarpus heterophyllus* and *Morus alba* (Moraceae) as an outgroup with 100% bootstrap support successfully anchors the phylogeny, confirming that Rosaceae is a monophyletic group within the order Rosales [19]. Overall, this phylogenetic analysis clarifies the taxonomic positioning of *F. × ananassa* 'Mencir' and highlights its evolutionary proximity to other key genera within the family.

Structural and Conserved Motif Analysis of pet Genes

It is normal for those genes as photosynthetic core genes in plants to not have introns. This high level of structural homogeneity suggests that genes that are engaged in photosynthesis are under a lot of purifying selection pressure, which keeps their function and stability [17, 65].

On the other hand, *petB* and *petD* differ because they both have group II introns in their coding areas. The chloroplast operon *psbB–psbT–psbH–petB–petD* is made up of these genes. Introns in *petB* and *petD* are linked to post-transcriptional RNA splicing and are an important way to control gene expression in chloroplasts that has been passed down through plant evolution. It is believed that the presence of introns

in *petB* and *petD* comes from the evolutionary history of the prokaryotic progenitors of chloroplasts. Specifically, group II introns are responsible for post-transcriptional splicing, gene expression regulation, and transcript variety. It is interesting to note that these introns are not found in all photosynthetic species. They are usually not found in many algal lineages, but they are always found in land plants. This pattern shows that *petB* and *petD* introns were probably picked up during evolution and show how mobile they are in plastid genomes [66].

It exposed that closely related species usually shared a similar motif composition. This is because the cytochrome b_6f complex is a pivotal component of the photosynthetic electron transport chain, hence changes in its core functional domain have the potential to impede photosynthetic performance [67]. The absence of unique motifs between the *F.* × *ananassa* 'Mencir' and other *Fragaria* species indicates that species differences do not affect the basic architecture of these essential photosynthetic genes. The only place where motif variation was seen was in the outgroup, *Rubus idaeus*. This shows that the basic function of the cytochrome b_6f complex is very stable, but little changes in its structure can build up over extended periods of evolution without affecting its main function [68].

Predicted Three-Dimensional Structures of the Cytochrome b_6f Proteins

The RMSD values of *F.* × *ananassa* 'Mencir' showed no significant differences when compared with other *Fragaria* cultivars, indicating that genetic variation among strawberry cultivars is insufficient to alter the architecture of the cytochrome b_6f complex. Interestingly, comparison with *Rubus idaeus* revealed RMSD values that were consistently below 1 Å and largely within the same range as those observed among *Fragaria* species. This confirms that structural conservation of cytochrome b_6f complex extends beyond the genus *Fragaria* to the broader Rosaceae family. This indicates that the structure of this complex in *Fragaria* genus is highly conserved due to strict functional selection during the evolution of the Rosaceae family. Significant structural changes in the core subunits can inhibit efficient electron transport, thus maintaining their structures, characterized by high structural uniformity, as evidence of strong purifying selection [27, 65]. The results showed that the Pet proteins were highly conserved in structure. In contrast to the large subunit, PetG, PetL, and PetN, the small subunits exhibited lower predicted TM-score and higher root mean square deviation (RMSD). The difficulty in predicting these small subunits is related to their short protein structure and the predominance of transmembrane domains in most of their regions. Rather than functional differences, the structural variations are likely due to the flexibility of the protein itself or small shifts in its terminal region. Functionally, PetG, PetL, and PetN are more involved in maintaining stability and regulating the activity of the cytochrome b_6f complex, rather than directly driving in electron transfer [69–71].

Codon Usage Analysis of the pet Genes

This is in line with the fact that the chloroplast genome is rich in AT and has to do with how well translation works [1]. It is interesting that the codon preference pattern in the *pet* genes is not exactly the same as that of other chloroplast genes. The strong preference for the CCT and GCT codons, for instance, shows that this bias can't be explained by the passive effects of mutations alone. Rather, this pattern indicates potential

translational preferences, provides a baseline codon reference for future plastid engineering. For genes involved in photosynthesis, this is important because it can increase the rate and accuracy of protein synthesis [72, 73]. The notable positive discrepancies in values for specific codons bolster the argument that the *pet* genes have undergone adaptation regarding codon utilisation. This adaptation is believed to be an evolutionary mechanism for sustaining the efficient and stable synthesis of the cytochrome b_6f complex. This codon usage pattern can be applied in chloroplast transformation. Adjustment to the codons used by rearranging the composition of the selected codons allows for more optimal expression of foreign genes [74].

Genetic Diversity Analysis

The higher polymorphism in *petB* and *petD*, with three and four polymorphic sites respectively, indicates relatively more sequence variability in these loci compared to other *pet* genes. However, this variability is still moderate ($\pi < 0.003$), suggesting that despite being the most variable among the *pet* genes, these loci remain functionally constrained. These results reveal differential selective constraints among *pet* gene subunits, *petB* and *petD* exhibit slightly relaxed constraints compared to the highly conserved *petG* and *petN* ($\pi = 0.000$), yet all subunits maintain low overall nucleotide diversity [1, 64]. Baniulis *et al.* (2008) demonstrated that the four small subunits (PetG, PetL, PetM, PetN) serve as peripheral 'hydrophobic sticks' that maintain complex stability, while the core subunits (PetA, PetB, PetD) contain redox prosthetic groups essential for electron transfer. This functional differentiation explains the extreme conservation of *petG* and *petN* observed in our study, where no nucleotide variation was detected ($\pi = 0.000$). Thus, the observed variation patterns offer important insights into the molecular evolution of photosynthetic genes [61].

Ka/Ks Analysis

This condition reflects an extreme level of conservation, even higher than that observed in other *pet* genes. Similar phenomena have been reported in chloroplast evolutionary studies of other plant groups [75], where genes with essential functions often exhibit extremely low or zero synonymous substitution rates. The absence of synonymous nucleotide variation suggests that not only the encoded proteins but also the gene structure itself may be under stringent functional constraints, resulting in a highly restricted overall mutation rate.

Although variation among genes was observed, with *petA* and *petD* displaying narrower value distributions compared to *petB* and *petL*, these differences are primarily quantitative and do not alter the main conclusion that all *pet* genes are predominantly under negative selection. Such variation likely reflects subtle differences in the degree of selective constraint among subunits of the cytochrome b_6f complex; however, overall, these remain within a highly restricted evolutionary framework. The absence of Ka/Ks values exceeding 1.0 across all analyzed genes confirms that no evidence of positive selection driving adaptive changes was detected in the *Fragaria* group examined.

The Ka/Ks values below 0.3 across all *pet* genes confirm intensive purifying selection, consistent with the functional roles described by Baniulis *et al.* (2008). The core subunits PetA, PetB, and PetD contain redox prosthetic groups (two b-type hemes, one c-type heme, and a [2Fe-2S] cluster) essential for electron transfer, explaining the strong selective constraints observed [1]. In contrast, the small subunits PetG and PetN,

which function as structural 'hydrophobic sticks' stabilizing the complex, exhibit even more extreme conservation (π = 0.000; Ks = 0). This differential conservation pattern reflects the hierarchy of functional importance within the cytochrome b_6f complex, where any alteration to core electron transfer components would be catastrophic for photosynthetic efficiency [64, 68].

This result is consistent with previous studies on *Fragaria* chloroplast genomes [1], as well as comparative analyses in the Rosaceae and Poaceae families [76, 77], which consistently reported that genes involved in photosynthetic pathways, particularly subunits of the cytochrome b_6f complex, exhibit the highest levels of conservation and the lowest Ka/Ks values. The extreme conservation of *petG* and *petN*, where Ks = 0, has also been observed in chloroplast evolutionary studies of other plant groups [75], where genes with essential functions often exhibit extremely low or zero synonymous substitution rates. The absence of synonymous nucleotide variation suggests that not only the encoded proteins but also the gene structure itself may be under stringent functional constraints, resulting in a highly restricted overall mutation rate.

Functionally, these results are consistent with the biological roles of *pet* genes as encoders of core components of the photosynthetic electron transport chain. Because any alteration in protein structure has the potential to disrupt electron transfer efficiency and plant viability, natural selection tends to maintain these gene sequences in a nearly identical state across species. The differential conservation patterns observed in this study also provide insights into the evolutionary strategies of photosynthetic genes in tropical strawberries. The strong purifying selection acting on the cytochrome b_6f complex reflects the universal conservation of essential photosynthetic functions across the Rosaceae family. [68].

The implication of these findings is that *pet* genes, particularly *petG* and *petN*, are unsuitable as molecular markers for phylogenetic studies at the species level due to their excessively low variation. Instead, they are more appropriately employed as models for studying purifying selection mechanisms and functional conservation in chloroplast genomes [61, 78].

Conclusions

In this study, we successfully assembled and annotated the complete chloroplast genome of *Fragaria* × *ananassa* 'Mencir' (155,568 bp) using high-throughput long-read sequencing via the Oxford Nanopore PromethION platform. The genome exhibits a highly conserved quadripartite structure typical of angiosperms, comprising LSC (85,551 bp), SSC (18,145 bp), and a pair of IR regions (25,936 bp each), with an overall GC content of 37.23%. Comparative analysis using mVISTA demonstrated that the *Fragaria* chloroplast genomes are highly syntenic and structurally conserved, with no evidence of major inversions, translocations, or gene losses among the analyzed accessions, while hypervariable intergenic regions (*trnH-psbA*, *trnK-rps16*, and the *petD* cluster) were identified as promising candidates for DNA barcoding and cultivar identification. Maximum Likelihood-based phylogenomic reconstruction strongly supported the placement of 'Mencir' within the *Fragaria* clade, showing its closest affinity to 'Royal Royce' (ON478194.1), consistent with its hybrid origin.

Structural analysis of the cytochrome b_6f complex genes (*petA*, *petB*, *petD*, *petG*, *petL*, and *petN*) revealed distinct exon intron organizations, where *petA*, *petG*, *petL*, and *petN* are intronless, while *petB* and *petD*

contain group II introns. Conserved motif analysis confirmed that the six *pet* genes exhibit strong sequence conservation across *Fragaria* species, with no unique motifs detected in 'Mencir' compared to other cultivars, indicating strong purifying selection maintaining functional integrity. Protein modeling using AlphaFold3 demonstrated that core subunits PetA, PetB, and PetD exhibit exceptional conformational stability (pTM > 0.83; RMSD < 0.23 Å), while the small subunits PetG, PetL, and PetN exhibited lower pTM scores (0.49–0.55) likely reflects computational limitations in modeling short peptides with high transmembrane domains, rather than inherent structural flexibility.

Additionally, RSCU analysis revealed a strong codon usage bias towards A/T-ending codons, consistent with the overall AT-rich nature of *Fragaria* chloroplast genomes, where mutational pressure serves as the primary driver of codon usage patterns. Genetic diversity and Ka/Ks analyses further confirmed that all *pet* genes are under intensive purifying selection, with Ka/Ks values substantially below 0.3 and nucleotide diversity (π) below 0.003. The extreme conservation of *petG* and *petN* (π = 0.000; Ks = 0) underscores their essential structural roles in maintaining cytochrome b_6f complex stability.

Overall, these findings provide essential genomic resources serve as potential candidate markers for identification, while offering broader evolutionary insights into chloroplast gene regulation. The high conservation of *pet* genes renders them unsuitable as phylogenetic markers but valuable as models for studying purifying selection and functional constraints in chloroplast genomes. Future studies should expand sampling across Indonesian *Fragaria* accessions and integrate transcriptomic data to further elucidate the regulatory mechanisms underlying environmental adaptation in tropical strawberries.

Declarations

Ethics approval and consent to participate

Not applicable. This study does not involve human participants, human data, or animal experiments. All experimental research on plant materials was conducted in strict adherence to institutional, national, and international guidelines. The plant species evaluated in this study, *Fragaria × ananassa*, is not listed as a threatened or endangered species under the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES). The plant material used in this study was obtained from commercial farm, not from wild populations; therefore, no specific collection permits or licences were required. All procedures complied with the applicable regulations of the Republic of Indonesia regarding the use of plant genetic resources for research purposes.

Consent for publication

Not applicable. This manuscript does not contain any individual person's data in any form.

Competing interests

The authors declare that they have no competing interests.

Funding

This research was supported by the Ministry of Education, Culture, Research, and Technology of the Republic of Indonesia under grant number (067/C3/DT.05.00/PL/2025).

Author Contribution

G.R.A conceived and designed the study, performed the molecular genetic studies, participated in sequence alignment, and drafted the manuscript. T.N.A, A.W, T.R, and D.A.W carried out the bioinformatics analyses, including genome assembly, annotation, and comparative genomic analyses. A.W and R.S.K contributed to sample collection and DNA extraction. M.F.A participated in the design of the study, performed the statistical analysis, and helped to draft the manuscript. All authors read and approved the final manuscript.

Acknowledgement

The authors thank the Indonesia Ministry of Education, Culture, Research, and Technology, for funding this research under the BIMA PFR research grant scheme (Grant Numbers: 067/C3/DT.05.00/PL/2025). We also appreciate the help and technical support from the Laboratory of Genetics and Breeding and the Faculty of Biology.

Data Availability

The complete chloroplast genome sequence of Fragaria × ananassa 'Mencir' reported in this study has been deposited in the NCBI GenBank database under accession number PZ351331.1 and is available at https://www.ncbi.nlm.nih.gov/nuccore/PZ351331.1. All other genome sequences used for comparative analyses were obtained from publicly available repositories, with their accession numbers listed in Table S1 (Supplementary Information)

References

1. Cheng H, Li J, Zhang H, Cai B, Gao Z, Qiao Y, et al. The complete chloroplast genome sequence of strawberry (*Fragaria* × *ananassa* Duch.) and comparison with related species of Rosaceae. PeerJ. 2017;5:e3919. 10.7717/peerj.3919.
2. Zahra N, Saeed MK, Fatima N, Shehzad E, Saeed A. Exploring the allure of strawberries. Innovare J Agricultural Sci. 2023;1–6. 10.22159/ijags.2023.v11i5.48533.
3. Wiryanata IWG, Hut S, Suharjo A. Analisis pendapatan usahatani stroberi (*Fragaria chiloensis* L). AGRIMETA: Jurnal Pertanian Berbasis Keseimbangan Ekosistem. 2021;11(1):46–56.
4. Setiawati MR, Rachelita N, Fitriatin BN, et al. Pengaruh Pemberian Asam Humat, Asam Fulvat, dan Pupuk Hayati pada Media Tanam terhadap Beberapa Sifat Kimia Tanah, Hasil, dan Kualitas Buah Stroberi (*Fragaria ananassa*). Agrikultura. 2023;34(2):255. 10.24198/agrikultura.v34i2.46734.

5. Masi R, Dungga NE, Yanti CWB. Peningkatan kualitas produksi stroberi melalui pemanfaatan bio-slurry cair. J Agrotan. 2015;1(1):45–56.
6. Ullah I, Demirsoy L, Basit A, et al. Year-round strawberry production: advancements in forcing techniques to extend the growing season. Planta. 2026;263(2):40. 10.1007/s00425-025-04902-y.
7. Aristya GR, Sasongko AB, Hidayati L, et al. Implementasi Inovasi Budidaya Stroberi di Agrowisata Banyuroto Kabupaten Magelang Melalui Education for Sustainable Development. Jurnal Pengabdian kepada Masyarakat (Indonesian J Community Engagement). 2017;2(2):125. 10.22146/jpkm.26500.
8. Prasad R, Lisiecka J, Raj K. Strawberry– More than a Popular Summer Fruit: A Mini-Review. ANAFS. 2022;2022(02). 10.37722/ANAFS.2022201.
9. Rizal A, Sachoemar SI, Aliah RS, et al. Economic valuation and the determinants of demand factors of Bandung strawberry agrotourism, West Java, Indonesia. Geoj Tourism Geosites. 2022;43(3):1081–90. 10.30892/gtg.43329-923.
10. FAOSTAT statistical database [Internet]. Food and Agriculture Organization of the United Nations. 2023. Available from: https://www.fao.org/faostat/en/#data/QCL [Accessed 31 December 2025].
11. Ismayaningrum D, Widhiyanti N, editors. Angka Tetap Hortikultura Tahun 2024. Jakarta: Directorate General of Horticulture; 2025.
12. Srimiati M, Habibah P, Dinya Zahra A, et al. Effect of Strawberry Powder Substitution on the Organoleptic of Instant Pudding: Pengaruh Substitusi Bubuk Stroberi terhadap Karakteristik Puding Instan. MGI. 2023;18(1):74–81. 10.20473/mgi.v18i1.74-81.
13. Panggabean TSU, Sofian A, Nuraida. Respon Pertumbuhan dan Produksi Stroberi (*Fragaria ananassa*) Varietas Mencir Terhadap Pemberian Pupuk NPK 32-10-10 dan Pupuk Organik Cair Kulit Pisang Pada Media Vertikultur. Jurnal Agrofolium. 2024;4(2):379–85.
14. Haryati S, Yuliastuti ER, Sudiaz R, et al. Pedoman Budi Daya Stroberi. Jakarta: Pertanian; 2023.
15. Romiyadi R, Putri NI. Growth and yield test of various strawberry (*Fragaria* sp.) varieties in lowland areas. Jurnal Greenation Pertanian dan Perkebunan. 2024;2:51–9.
16. Koryati T, Ningsih H, Erdiandini I, Paulina M, Firgiyanto R, Junairiah, et al. Pemuliaan Tanaman. Medan: Yayasan Kita Menulis; 2022.
17. Wicke S, Schneeweiss GM, dePamphilis CW, et al. The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. Plant Mol Biol. 2011;76(3–5):273–97. 10.1007/s11103-011-9762-4.
18. Song Y, Li C, Liu L, et al. The population genomic analyses of chloroplast genomes shed new insights on the complicated ploidy and evolutionary history in *Fragaria*. Front Plant Sci. 2023;13:1065218. 10.3389/fpls.2022.1065218.
19. Aristya GR, Budiman MS, Kasiamdari RS, et al. Genetic Variation and Phylogenetic Analysis of Strawberry (*Fragaria* spp.) on Yogyakarta and Central Java, Indonesia, Based on rbcL DNA Barcoding. Pakistan J Biol Sci. 2025;28(3):121–30. 10.3923/pjbs.2025.121.130.
20. Henry RJ. Progress in Plant Genome Sequencing. Appl Biosci. 2022;1(2):113–28. 10.3390/applbiosci1020008.

21. Shillito RD, Whitt S, Ross M, Ghavami F, De Vleesschauwer D, D'Halluin K, et al. Detection of genome edits in plants—from editing to seed. Vitro CellDevBiol-Plant. 2021;57(4):595–608. 10.1007/s11627-021-10214-z.
22. Isobe S, Shirasawa K, Hirakawa H. Advances of Whole Genome Sequencing in Strawberry with NGS Technologies. Hort J. 2020;89(2):108–14. 10.2503/hortj.UTD-R012.
23. Lu Y, Li M, Gao Z, et al. Advances in Whole Genome Sequencing: Methods, Tools, and Applications in Population Genomics. IJMS. 2025;26(1):372. 10.3390/ijms26010372.
24. Li C, Cai C, Tao Y, et al. Variation and Evolution of the Whole Chloroplast Genomes of Fragaria spp. (Rosaceae). Front Plant Sci. 2021;12:754209. 10.3389/fpls.2021.754209.
25. Malone LA, Proctor MS, Hitchcock A, et al. Cytochrome b6f – Orchestrator of photosynthetic electron transfer. Biochimica et Biophysica Acta (BBA). - Bioenergetics. 2021;1862(5):148380. 10.1016/j.bbabio.2021.148380.
26. Saudale FZ. Analisis struktur gen dan protein lipid droplet-associated protein (LDAP) yang berperan dalam biogenesis minyak pada tanaman jarak pagar (*Jatropha curcas* L). 2020;9(2), 111–20.
27. Wang J, Ye H, Li X, et al. Genome-Wide Analysis of the MADS-Box Gene Family in Hibiscus syriacus and Their Role in Floral Organ Development. IJMS. 2023;25(1):406. 10.3390/ijms25010406.
28. Wen H, Wang L, Gong Y, et al. Genome-Wide Identification and Characterization of GRAS Transcription Factor Family in Cultivated Hybrid Sugarcane ZZ1 (Saccharum officinarum) and Their Role in Development and Stress. IJMS. 2024;25(24):13470. 10.3390/ijms252413470.
29. Song Y, Shen M, Cao F, Yang X. Compare Analysis of Codon Usage Bias of Nuclear Genome in Eight Sapindaceae Species. IJMS. 2024;26(1):39. 10.3390/ijms26010039.
30. Vaser R, Šikić M. Raven: a de novo genome assembler for long reads [Internet]. Bioinformatics; 2020 [cited 2025 December 21]. Available from: http://biorxiv.org/lookup/doi/10.1101/2020.08.07.242461 doi:10.1101/2020.08.07.242461.
31. Rozas J, Ferrer-Mata A, Sánchez-DelBarrio JC, et al. Mol Biol Evol. 2017;34(12):3299–302. 10.1093/molbev/msx248. DnaSP 6: DNA Sequence Polymorphism Analysis of Large Data Sets.
32. Katoh K, Standley DM. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. Mol Biol Evol. 2013;30(4):772–80. 10.1093/molbev/mst010.
33. Larsson A. AliView: a fast and lightweight alignment viewer and editor for large datasets. Bioinformatics. 2014;30(22):3276–8. 10.1093/bioinformatics/btu531.
34. Darriba D, Posada D, Kozlov AM et al. ModelTest-NG: A New and Scalable Tool for the Selection of DNA and Protein Evolutionary Models. Crandall K, editor. Molecular Biology and Evolution. 2020;37(1):291–4. 10.1093/molbev/msz189
35. Kozlov AM, Darriba D, Flouri T et al. RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. Wren J, editor. Bioinformatics. 2019;35(21):4453–5. 10.1093/bioinformatics/btz305
36. Arshad KT, Li C, Li L, et al. Genome-wide identification and expression profiling of bHLH transcription factors associated with ferulic acid biosynthesis in Angelica sinensis. Front Plant Sci. 2025;16:1718585. 10.3389/fpls.2025.1718585.

37. Okonechnikov K, Golosova O, Fursov M. the UGENE team. Unipro UGENE: a unified bioinformatics toolkit. Bioinformatics. 2012;28(8):1166–7. 10.1093/bioinformatics/bts091.
38. Abramson J, Adler J, Dunger J, et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. Nature. 2024;630(8016):493–500. 10.1038/s41586-024-07487-w.
39. Cock PJA, Antao T, Chang JT, et al. Biopython: freely available Python tools for computational molecular biology and bioinformatics. Bioinformatics. 2009;25(11):1422–3. 10.1093/bioinformatics/btp163.
40. Sawitri WD, Wahyuni DK, Saputro TB, et al. Complete Chloroplast Genome of Indonesian Sugarcane (*Saccharum* 'Bululawang'): Comparative, Structural, and Evolutionary Analysis of Photosystem I-Encoding Genes. J Sust Agri Env. 2026;5(3):e70187. 10.1002/sae2.70187.
41. Hunter JD, Matplotlib. A 2D Graphics Environment. Comput Sci Eng. 2007;9(3):90–5. 10.1109/MCSE.2007.55.
42. Munyao JN, Dong X, Yang JX, Mbandi EM, Wanga VO, Oulo MA, et al. Complete Chloroplast Genomes of Chlorophytum comosum and Chlorophytum gallabatense: Genome Structures, Comparative and Phylogenetic Analysis. Plants. 2020;9(3):296. 10.3390/plants9030296.
43. Morton BR. Selection on the codon bias of chloroplast and cyanelle genes in different plant and algal lineages. J Mol Evol. 1998;46(4):449–59. 10.1007/PL00006325.
44. Ratnere I, Dubchak I. CP Bioinf. 2009;26(1). 10.1002/0471250953.bi1006s26. Obtaining Comparative Genomic Data with the VISTA Family of Computational Tools.
45. Dong W, Xu C, Li C, et al. ycf1, the most promising plastid DNA barcode of land plants. Sci Rep. 2015;5(1):8348. 10.1038/srep08348.
46. Delpierre N, Berveiller D, Granda E, et al. Wood phenology, not carbon input, controls the interannual variability of wood growth in a temperate oak forest. New Phytol. 2016;210(2):459–70. 10.1111/nph.13771.
47. Xiang Y, Huang CH, Hu Y, et al. Evolution of Rosaceae Fruit Types Based on Nuclear Phylogeny in the Context of Geological Times and Genome Duplication. Mol Biol Evol. 2016;msw242. 10.1093/molbev/msw242.
48. De Coster W, D'Hert S, Schultz DT et al. NanoPack: visualizing and processing long-read sequencing data. Berger B, editor. Bioinformatics. 2018;34(15):2666–9. 10.1093/bioinformatics/bty149
49. Wick RR, Judd LM, Gorrie CL, et al. Completing bacterial genome assemblies with multiplex MinION sequencing. Microb Genomics. 2017;3(10). 10.1099/mgen.0.000132.
50. Bowden R, Davies RW, Heger A, et al. Sequencing of human genomes with nanopore technology. Nat Commun. 2019;10(1):1869. 10.1038/s41467-019-09637-5.
51. Jain M, Koren S, Miga KH, et al. Nanopore sequencing and assembly of a human genome with ultra-long reads. Nat Biotechnol. 2018;36(4):338–45. 10.1038/nbt.4060.
52. Daniell H, Lin CS, Yu M, et al. Chloroplast genomes: diversity, evolution, and applications in genetic engineering. Genome Biol. 2016;17(1):134. 10.1186/s13059-016-1004-2.
53. Palmer JD. Comparative Organization of Chloroplast Genomes. Annu Rev Genet. 1985;19(1):325–54. 10.1146/annurev.ge.19.120185.001545.

54. Wang Y, Wang S, Liu Y, et al. Chloroplast genome variation and phylogenetic relationships of Atractylodes species. BMC Genomics. 2021;22(1):103. 10.1186/s12864-021-07394-8.
55. Li JH, Xu WB, Guo CH. Comparative Analysis of Chloroplast Genome Between Widely Distributed and Locally Distributed Lysionotus (Gesneriaceae) Related Members. IJMS. 2025;26(15):7031. 10.3390/ijms26157031.
56. Quax TEF, Claassens NJ, Söll D, van der Oost J. Codon Bias as a Means to Fine-Tune Gene Expression. Mol Cell. 2015;59(2):149–61. 10.1016/j.molcel.2015.05.035.
57. Sharp PM, Li WH. The codon adaptation index-a measure of directional synonymous codon usage bias, and its potential applications. Nucl Acids Res. 1987;15(3):1281–95. 10.1093/nar/15.3.1281.
58. Wicke S, Schäferhoff B, dePamphilis CW, et al. Disproportional Plastome-Wide Increase of Substitution Rates and Relaxed Purifying Selection in Genes of Carnivorous Lentibulariaceae. Mol Biol Evol. 2014;31(3):529–45. 10.1093/molbev/mst261.
59. Goulding SE, Wolfe KH, Olmstead RG, et al. Ebb and flow of the chloroplast inverted repeat. Molec Gen Genet. 1996;252(1–2):195–206. 10.1007/BF02173220.
60. Li C, Cai C, Tao Y, et al. Variation and Evolution of the Whole Chloroplast Genomes of Fragaria spp. (Rosaceae). Front Plant Sci. 2021;12:754209. 10.3389/fpls.2021.754209.
61. Shaw J, Lickey EB, Schilling EE, Small RL. Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: the tortoise and the hare III. Am J Bot. 2007;94(3):275–88. 10.3732/ajb.94.3.275.
62. Amiryousefi A, Hyvönen J, Poczai P. IRscope: an online program to visualize the junction sites of chloroplast genomes. Hancock J, editor. Bioinformatics. 2018;34(17):3030–1. 10.1093/bioinformatics/bty220
63. CBOL Plant Working Group1, Hollingsworth PM, Forrest LL, Spouge JL, Hajibabaei M, Ratnasingham S, et al. A DNA barcode for land plants. Proc Natl Acad Sci USA. 2009;106(31):12794–7. 10.1073/pnas.0905845106.
64. Baniulis D, Yamashita E, Zhang H, Hasan SS, Cramer WA. Structure–Function of the Cytochrome b_6 f Complex†. Photochem Photobiology. 2008;84(6):1349–58. 10.1111/j.1751-1097.2008.00444.x.
65. Robbins EH, Kelly S. The Evolutionary Constraints on Angiosperm Chloroplast Adaptation. Genome Biol Evol. 2023;15(6):evad101.
66. Stoppel R, Meurer J. Complex RNA metabolism in the chloroplast: an update on the psbB operon. Planta. 2013;237(2):441–9. 10.1007/s00425-012-1782-z.
67. Tikhonov AN. The cytochrome b6f complex at the crossroad of photosynthetic electron transport pathways. Plant Physiol Biochem. 2014;81:163–83. 10.1016/j.plaphy.2013.12.011.
68. Malone LA, Proctor MS, Hitchcock A, et al. Cytochrome b6f – Orchestrator of photosynthetic electron transfer. Biochimica et Biophysica Acta (BBA). - Bioenergetics. 2021;1862(5):148380. 10.1016/j.bbabio.2021.148380.
69. Hamel P, Olive J, Pierre Y, et al. A New Subunit of Cytochromeb 6 f Complex Undergoes Reversible Phosphorylation upon State Transition. J Biol Chem. 2000;275(22):17072–9. 10.1074/jbc.M001468200.

70. Monzon AM, Zea DJ, Fornasari MS et al. Conformational diversity analysis reveals three functional mechanisms in proteins. Orengo CA, editor. PLoS Comput Biol. 2017;13(2):e1005398. 10.1371/journal.pcbi.1005398
71. Dobson L, Szekeres LI, Gerdán C, et al. TmAlphaFold database: membrane localization and evaluation of AlphaFold2 predicted alpha-helical transmembrane protein structures. Nucleic Acids Res. 2023;51(D1):D517–22. 10.1093/nar/gkac928.
72. Suzuki H, Morton BR. Codon Adaptation of Plastid Genes. Li XQ, editor. PLoS ONE. 2016;11(5):e0154306. 10.1371/journal.pone.0154306
73. Hao J, Liang Y, Ping J, et al. Chloroplast gene expression level is negatively correlated with evolutionary rates and selective pressure while positively with codon usage bias in Ophioglossum vulgatum L. BMC Plant Biol. 2022;22(1):580. 10.1186/s12870-022-03960-8.
74. Wang Z, Xu B, Li B, et al. Comparative analysis of codon usage patterns in chloroplast genomes of six *Euphorbiaceae* species. PeerJ. 2020;6:8:e8251. 10.7717/peerj.8251.
75. Rono PC, Dong X, Yang JX, et al. Initial Complete Chloroplast Genomes of Alchemilla (Rosaceae): Comparative Analysis and Phylogenetic Relationships. Front Genet. 2020;9:11:560368. 10.3389/fgene.2020.560368.
76. Mahai R, Liu R, Du X, et al. Comparative and Phylogenetic Analysis of the Complete Chloroplast Genomes of 19 Species in Rosaceae Family. Phyton. 2024;93(6):1203–19. 10.32604/phyton.2024.051559.
77. Geng X, Xue Y, Wang H, et al. Comprehensive analysis of chloroplast genome evolution in Poaceae: codon usage patterns, selection pressures, and phylogenomic relationships. BMC Genomics. 2026;27(1):217. 10.1186/s12864-025-12498-6.
78. Shaw J, Shafer HL, Leonard OR, et al. Chloroplast DNA sequence utility for the lowest phylogenetic and phylogeographic inferences in angiosperms: The tortoise and the hare IV. Am J Bot. 2014;101(11):1987–2004. 10.3732/ajb.1400398.

Figures

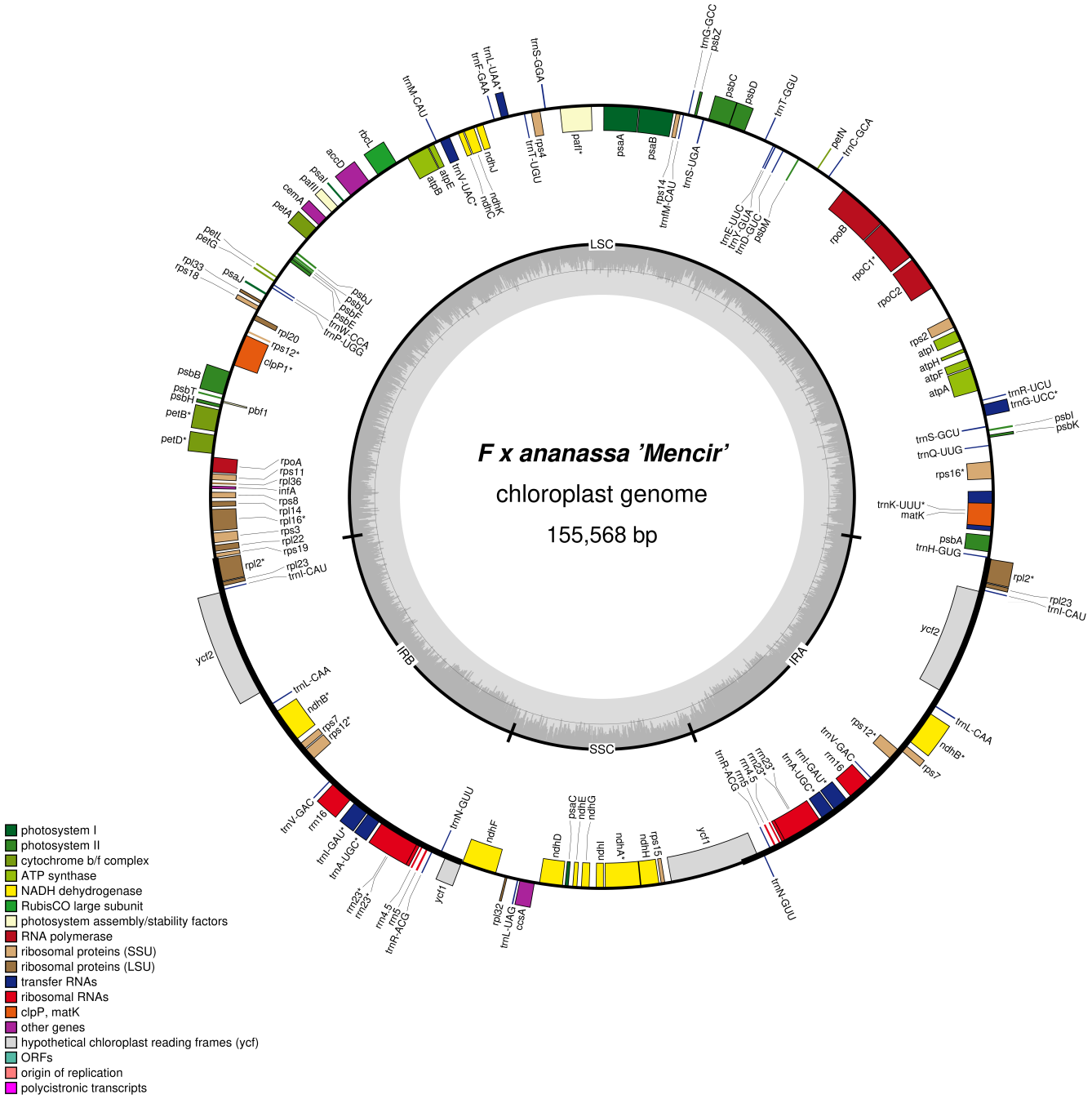


Figure 1

Circular maps of *F. × ananassa*'Mencir' chloroplast genome. Genes shown on the outside of the circle are transcribed clockwise, while those on the inside are transcribed counterclockwise. The inner track shows the extent of the inverted repeats (IRA and IRB), the large single-copy (LSC) region, and the small single-copy (SSC) region. Genome annotation was performed using CpGAVAS2, and the circular map was visualized using OGDRAW.

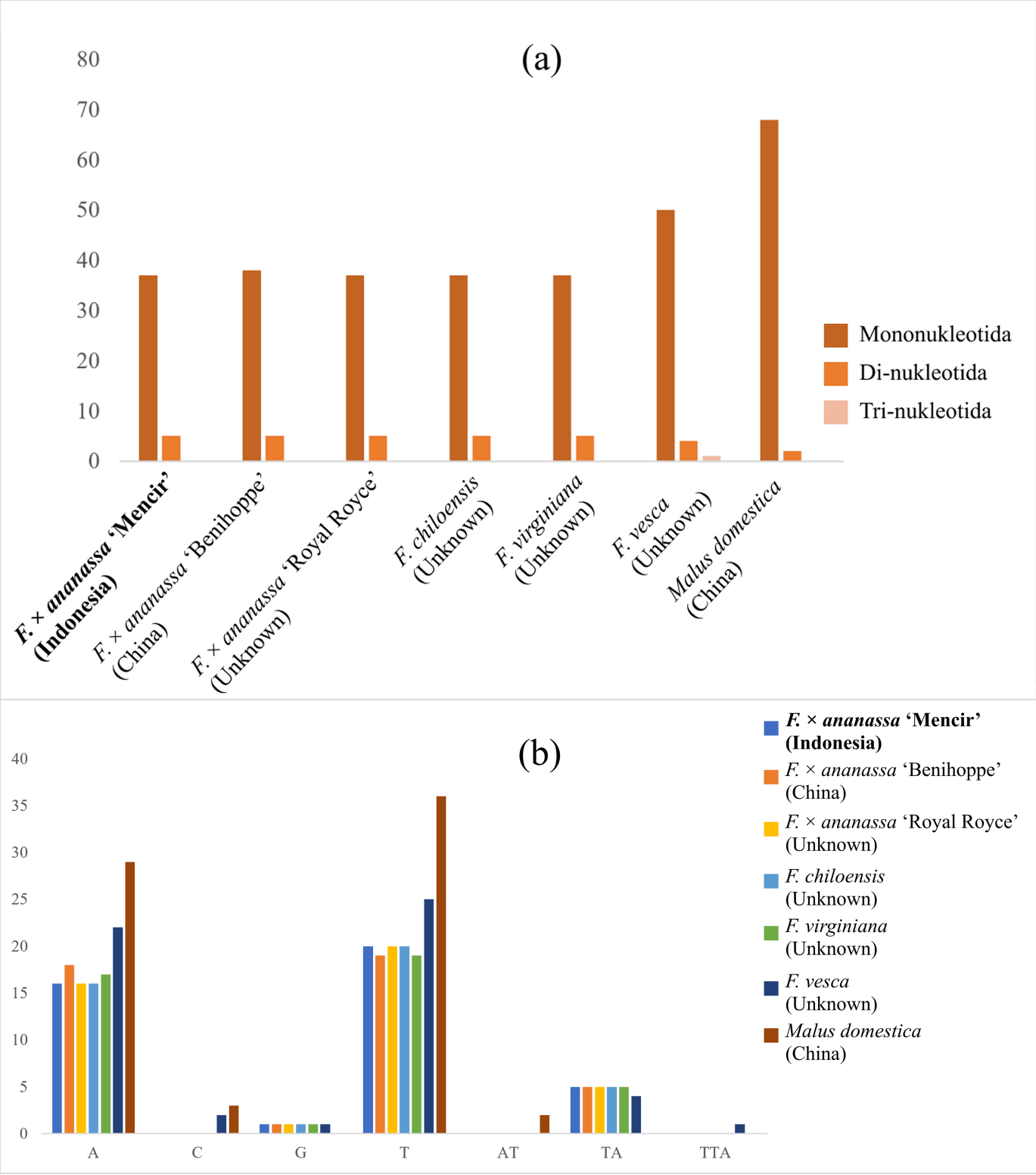


Figure 2

Comparison of the distribution of chloroplast SSR motifs. (a) Nucleotide types of SSR motifs; (b) distribution of SSR motif sequence types.

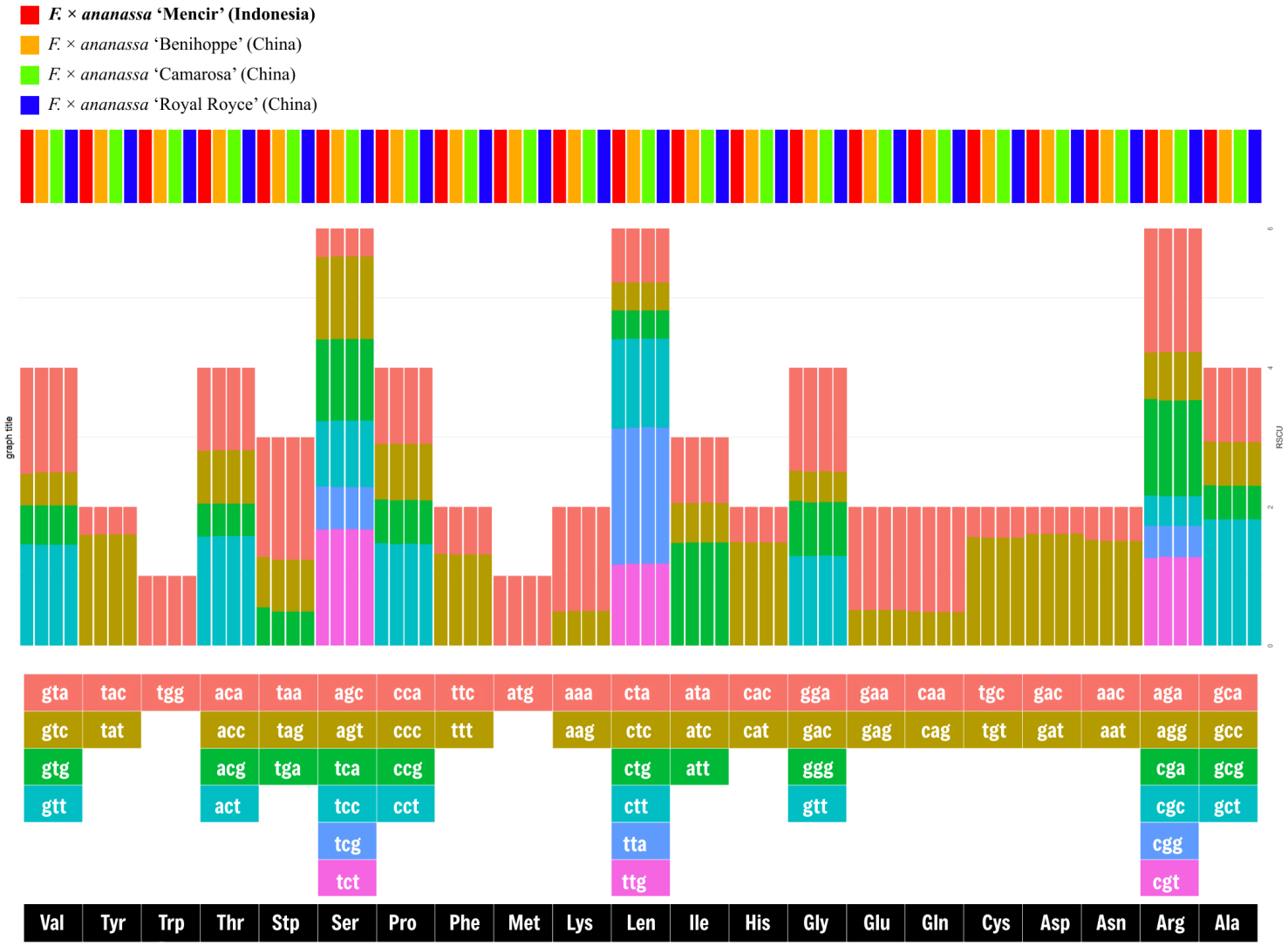


Figure 3

RSCU patterns of chloroplast protein-coding genes in F. × ananassa cultivars.

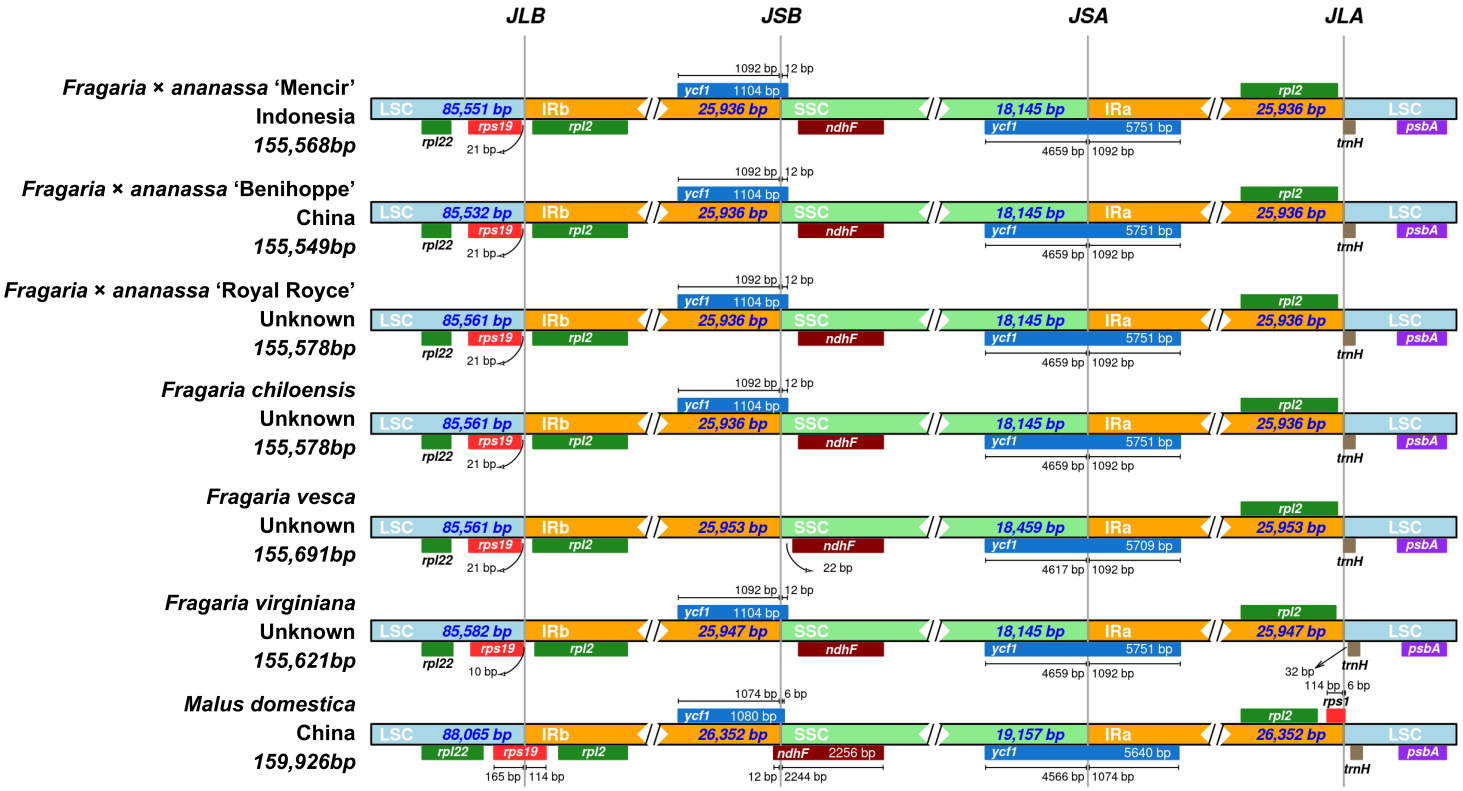


Figure 4

Comparison of the junction between the LSC, SSC, and IR regions across 6 Fragaria species and Malus domestica as outgroup

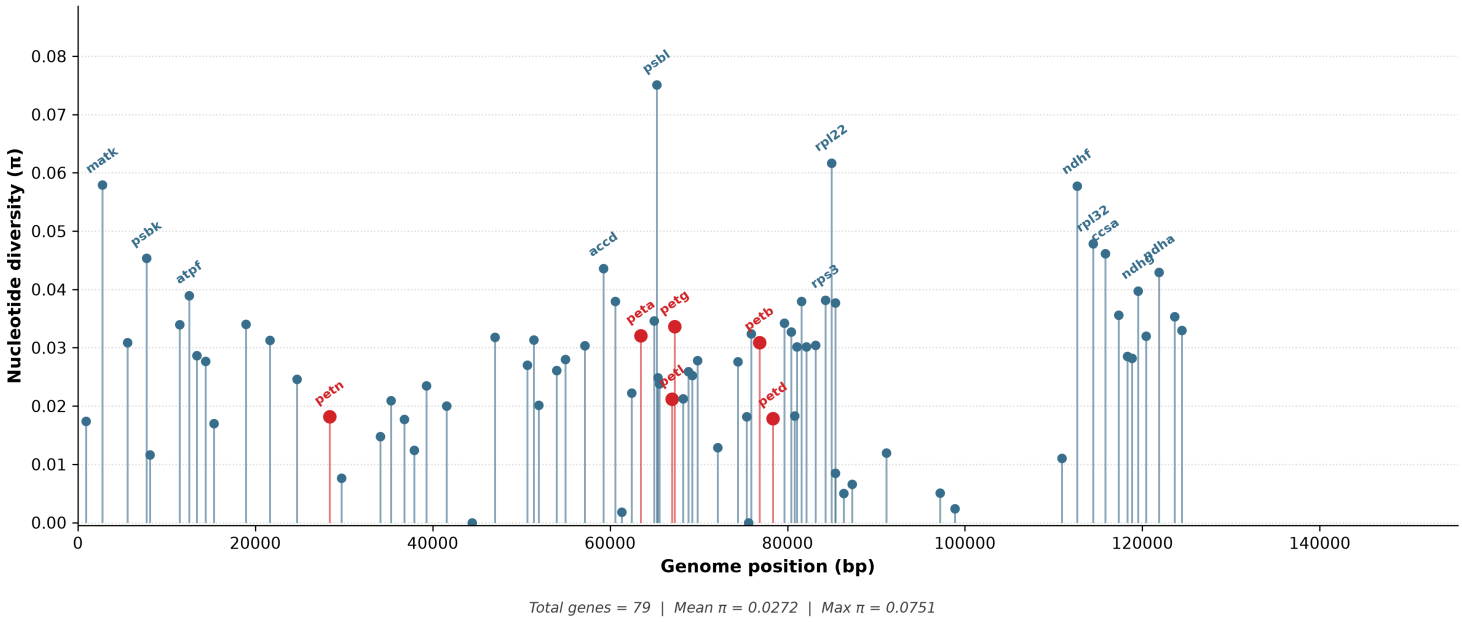


Figure 5

Distribution of nucleotide diversity (π) among 78 plastid protein-coding genes in 11 species. Each lollipop represents one gene, with the stem anchored at its genomic position (midpoint) and the circle height

indicating its π value.

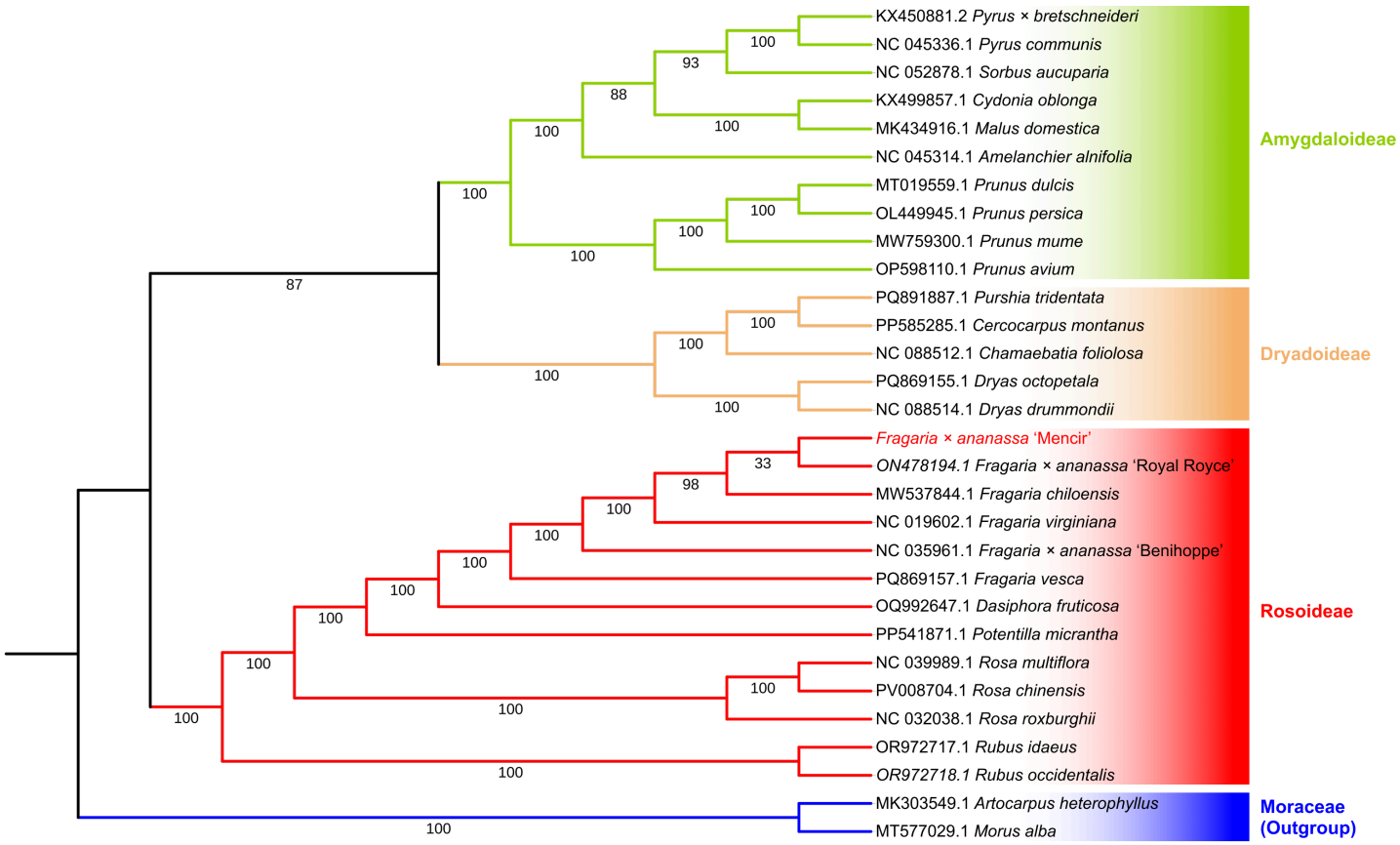


Figure 6

Phylogenomic trees of F. × ananassa with related species.

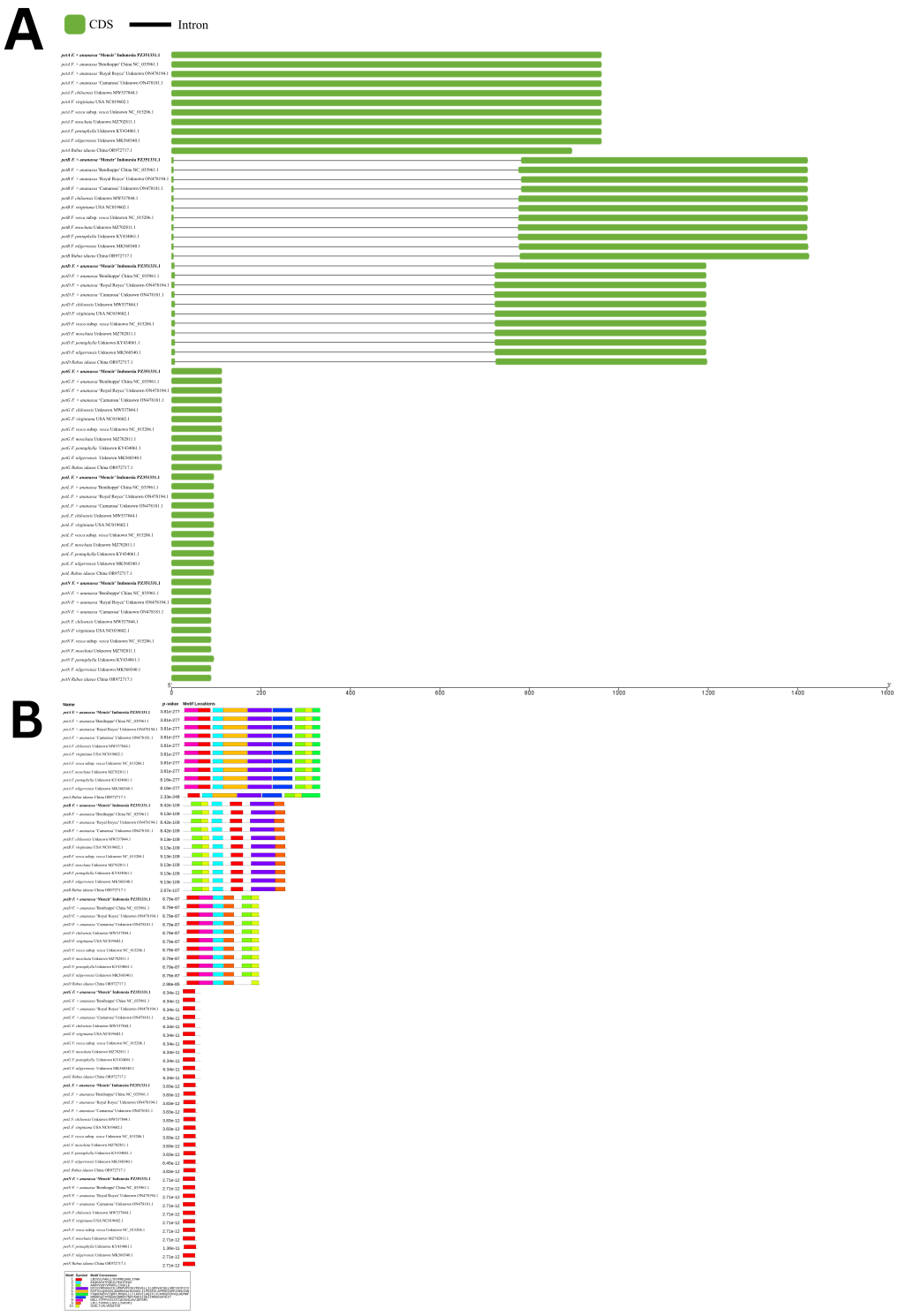


Figure 7

Comparative gene structure and motif analysis of chloroplast pet genes across Fragaria and related species. A.) Exon–intron organization (B) Distribution of conserved motifs

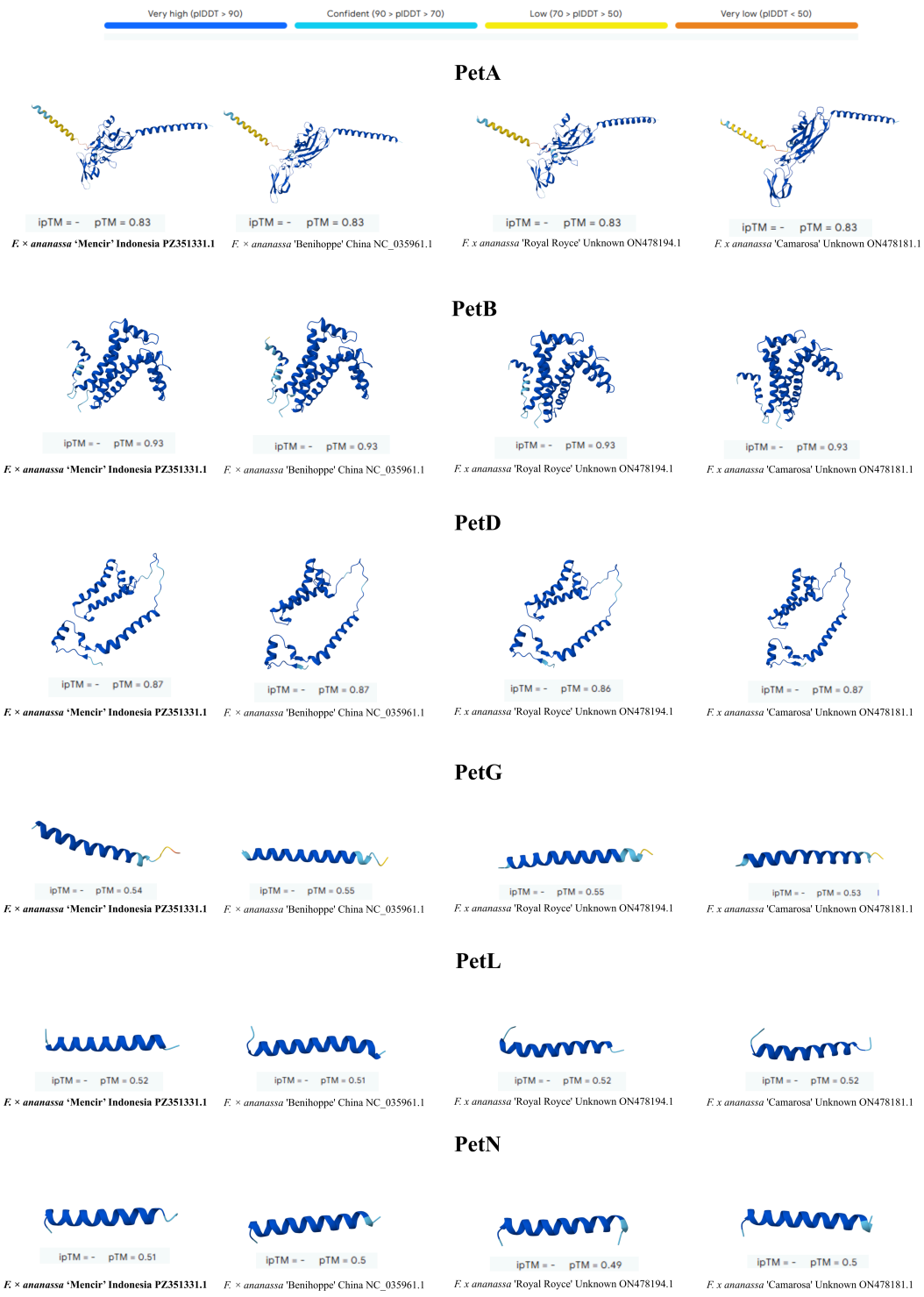


Figure 8

Three-Dimensional Protein Structure Visualization of pet Genes from the Chloroplast Genome of Strawberry (F. ×ananassa) Cultivars: Indonesian 'Mencir'(this study, PZ351331.1), Chinese 'Benihoppe' (NC_035961.1) (Cheng et al., 2017), 'Royal Royce' (ON478194.1), and 'Camarosa' (ON478181.1) (Sun et al., 2022, unpublished).

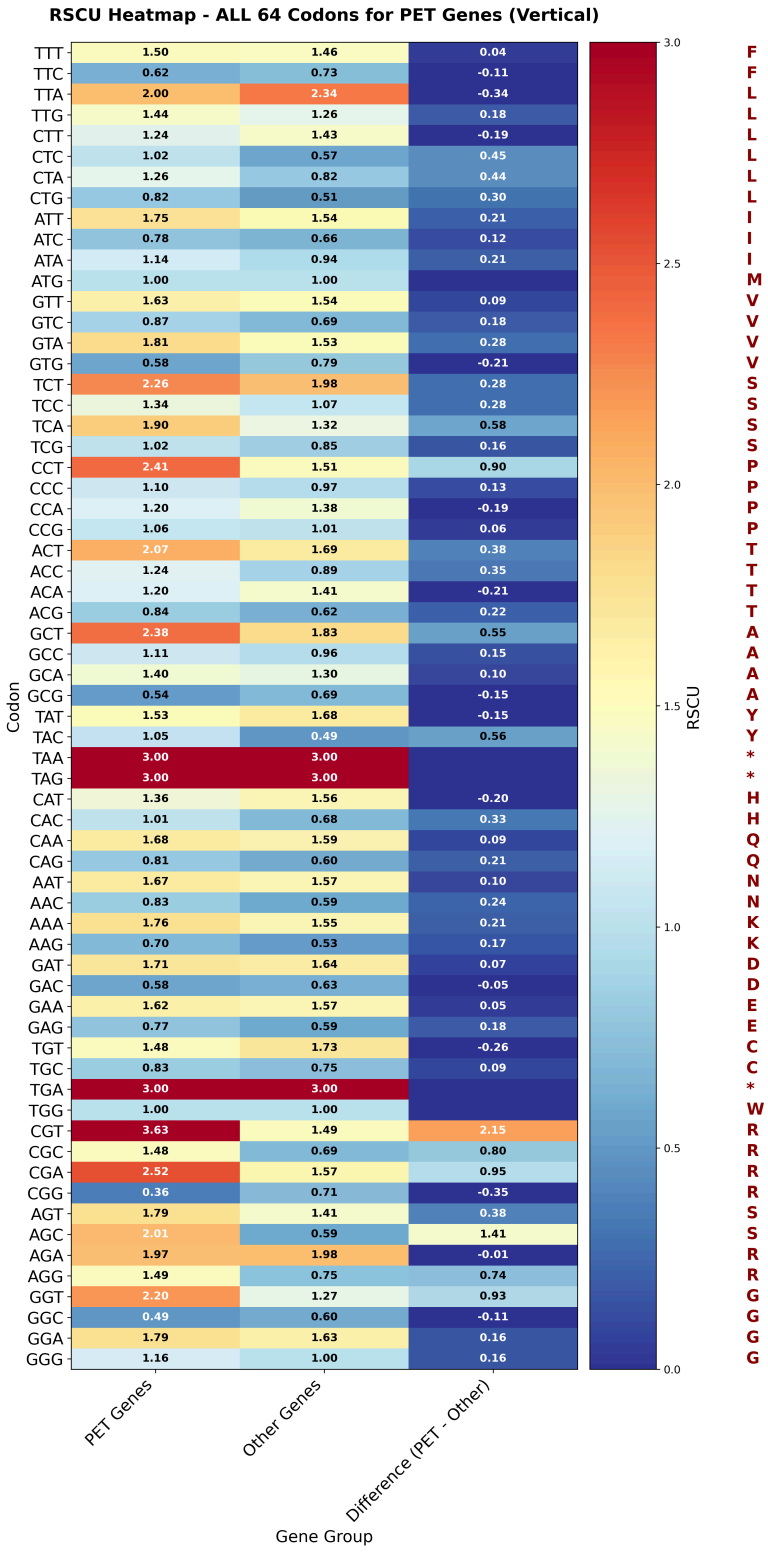


Figure 9

Relative Synonymous Codon Usage (RSCU) Heatmap of 64 Codons in pet Genes Compared to Other Genes in the Chloroplast Genome.

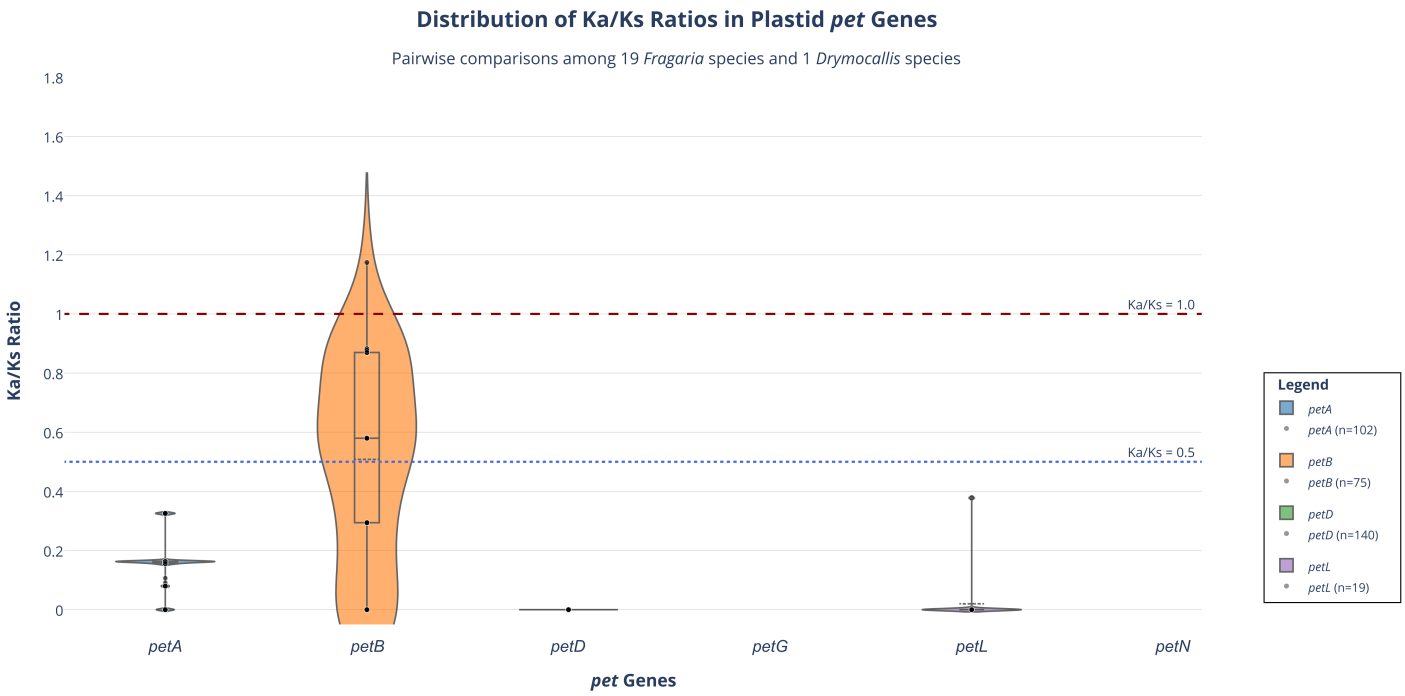


Figure 10

Distribution of pairwise Ka/Ks ratios among concatenated species across 19 *Fragaria* species and one *Drymocallis*species. A violin plot depicts the density and full distribution of Ka/Ks values derived from all valid pairwise interspecific comparisons for each of the five genes. Black dots represent individual pairwise comparisons. Sample sizes (n) indicate the number of valid pairwise comparisons retained after filtering (Ks > 0.0001, Ka>0, Ka/Ks < 0.2).

Supplementary Files

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

- SupplementaryData1.docx