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.

Assembly and analysis of the complete mitochondrial genome of Prunus davidiana (Rosaceae)

Yuyan Zhang

zhangyuyan@jaas.ac.cn

Jiangsu Academy of Agricultural Sciences

Xin Guo

Nanjing Agricultural University

Ruijuan Ma

Jiangsu Academy of Agricultural Sciences

Mingliang Yu

Jiangsu Academy of Agricultural Sciences

Jianlan Xu

Jiangsu Academy of Agricultural Sciences

Zhijun Shen

Jiangsu Academy of Agricultural Sciences

Aihua Ren

Horticulture Branch of Heilongjiang Academy of Agricultural Sciences

Research Article

Keywords: Prunus davidiana, mitochondrial genome, repeats, phylogenetic analysis

Posted Date: April 9th, 2024

DOI: https://doi.org/10.21203/rs.3.rs-4163592/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: *Prunus davidiana* is an excellent woody tree species that has high resistance to various abiotic and biotic stress factors. The species is also a potential gene donor because of its important developmental prospects. While the characteristic of mitochondrial genome (mitogenome) remains unexplored.

Results: In this study, we sequenced and assembled the *P. davidiana* mitogenome by using the reads from Illumina sequencing and Oxford Nanopore platform. According to the results, the *P. davidiana*mitogenome has a length of 375 671 bp, with a 45.5% GC content. It contains 64 genes, including 39 protein-coding genes, 22 tRNA genes, and 3 rRNA genes. The codon usage, repeated sequences, nonsynonymous to synonymous substitution ratios, RNA editing, synteny, and phylogenetic relationships among Rosaceae species and gene migration event were all examined, respectively. The results of our analyses provide valuable information and a theoretical basis for research on *P. davidiana* in future.

Conclusions: In this study, we assembled and annotated the mitogenome of *P. davidiana*, an excellent woody tree species that has high resistance to various abiotic and biotic stress factors. The results of our analyses provide comprehensive information on the *P. davidiana* mitogenome, which should facilitate evolutionary research and molecular barcoding in Rosaceae.

Background

Prunus davidiana is a small shrub that belongs to the Prunoideae subfamily in the family Rosaceae [1]. The tree species is native to China and is reportedly resistant to cold, saline alkalinity, and drought stress, and some pests and diseases [2]. Therefore, it is commonly used as a grafting rootstock for peaches and plums in northern China considering its adaptability under various environmental conditions [3,4]. Accordingly, *P. davidiana* is considered a potential gene donor in future peach breeding programs.

As the“energy factories”of cells, the mitochondria are important for energy synthesis and conversion in cell physiological activities [5]. They play crucial roles in plant growth and development [6]. Plant organelle genomes have genetic systems independent of the nuclear genome that have share the stable regulatory mechanisms with the nuclear genome [7]. The chloroplast (cp) genome is usually structurally conserved with double-stranded stable circular genomes [8]. However, other structures of mitogenomes have been observed, including circular, linear, branching, and numerous small circular molecules [9]. In addition, plant mitochondrial genomes (mitogenomes) vary in size even among related species. The size of most genomes is 200–800 kb, and the largest land plant mitogenome is approximately 11.3 Mb (*Silene conica*) [10]. While the smallest is only approximately 66 kb (*Viscum scurruloideum*) [11]. Mitogenomes contain abundant repetitive sequences and mediate intergenomic gene transfer, which could lead to the variation in genome size and rapid genome rearrangement [12]. Recovering the conformation of plant mitogenomes is challenging because of their redundant sequences and genomic recombination [13]. In addition, gene sequences transfer between nuclear and organelle genomes is a common phenomenon, as reported in recent research [7, 14,15]. Consequently, gene contents also vary considerably [16].

P. davidiana is an excellent germplasm material for the study of peach breeding because of its high resistance to many numerous biotic and abiotic stress. Although the cp genome of *P. davidiana* has been reported [17], the mitogenome has not been reported. In the present study, we report the sequencing, assembly, and annotation of the *P. davidiana* mitogenome using a combination of the Illumina and PacBio sequencing platforms. In addition, we also analyzed the repetitive sequences, selective pressures, RNA editing sites and phylogenetic relationships among Rosaceae species in the present study. Gene transfer among nuclear, cp, and mitochondrial genomes were also surveyed. The results of the present study could enhance our understanding of organelle genome evolution in *P. davidiana*, in addition to facilitating genomic breeding of *P. davidiana*.

Materials and Methods

Materials and sequencing

P. davidiana plants were grown under natural conditions in the Germplasm base of the Jiangsu Academy of Agricultural Sciences, Nanjing, China. Total DNA was isolated from fresh leaves using the CTAB method [18] and then quality was evaluated using 1% agarose gel electrophoresis. The qualified samples were sent to Nanjing Jisi Huiyuan Biotechnology Co. (http://www.genepioneer.com/) for Illumina sequencing and Oxford sequencing. The experimental procedures were carried out according to a standard protocol [19]. To obtain the full-length mitochondrial genome with high accuracy. The paired end sequencing (PE) reads (150 bp) were obtained by Illumina Novaseq 6000 (Illumina, San Diego, CA, USA), and we filter the original data by using the fastp (v 0.20.0, https:// github. com/ OpenG ene/ fastp) software to get high-quality clear reads. The long-read sequences were obtained by Nanopore PromethION (Nanopore, Oxford, UK), which were filtered by filtlong (v0.2.1,https://link.zhihu.com/?target=https%3A//github.com/rrwick/Filtlong) software.

Mitogenome assembly and annotation

Plant mitochondrial genes (CDS and rRNA) are highly conserved. Taking advantage of this feature, the reads were mapped to the reference gene sequences (plant 2mitochondrial core gene (https://github.com/xul962464/plant_mt_ref_gene) using Minimap2 (v2.1) [20]. The candidate sequences contains multiple core genes and higher alignment quality (covering more complete core genes) were selected as seed sequences. The seed sequences were then compared to the long-read sequencing data with minimum overlap of 1kb and at least 70% similarity. All long-read sequencing data of the mitochondrial genome were obtained. The canu [21] and bowtie2 (v2.3.5.1) [22]were used to correct the long-read sequencing data and the short-read sequencing data, respectively. Finally, the mitochondrial genome of *P. davidiana* was obtained.

The structure annotation of the Mitochondrial genome was divided into the following steps: (1) The encoding protein and rRNA were aligned with the published plant mitochondrial sequences using BLAST and further manual adjustments with the related species. (2) The tRNA was annotated using tRNAs-canSE (http:// lowel ab. ucsc. edu/ tRNAs can- SE/) with default settings [23]. To obtain more accurate

annotation results, the above results were checked and manually corrected using Apollo [24]. The mitochondrial genome was then mapped using OGDRAW (https:// chlor obox. mpimp- golm. mpg. de/ OGDraw. html).91

Analysis of RSCU and repeat structures

We conducted the Codon preference analysis for the PCGs of the mitogenome and calculation of RSCU values by using MEGA 7.0 [25]. SSRs were discovered using Perl script MISA (http://pgrc.ipkgatersleben.de/misa/) [26], with sizes of one to six nucleotides and thresholds of 10, 5, 4, 3, 3, and 3. In addition, the REPuter [27] were used to identify the forward, palindromic, and tandem repeats.

Analysis of Selective pressure

We calculated the nonsynonymous (Ka) and synonymous (Ks) substitution rates with PCGs among *P. davidiana*, *M. domestica, S. aucuparia, R. bibas, P. betulifolia*, and *F. orientalis*. DnaSP was used to calculate Ka, Ks, and Ka/Ks values [28].

RNA editing sitespredictionandGenome alignments

The *P. davidiana* cp genome (MH460864) was downloaded from the NCBI Organelle Genome Resources Database. We searched the *P. davidiana* mitogenome against the cp genome using BLASTN 2.9.0+ [29]. In addition, a BLASTN search (E-value = $1 \acute{} 10^{-50}$) was also performed between the mitogenome and the *P. davidiana* nuclear genome sequence (PRJNA655343). PREP suit (http://prep.unl.edu/) (the cutoff value was set as 0.2) was used to predict the RNA editing sites in the PCGs of the *P. davidiana* mitogenome.

Synteny and phylogenetic tree construction

A dot plot pairwise comparison was generated, and the conserved co-linear blocks were plotted. Subsequently, a Multiple Synteny Plot of the *P. davidiana* mitogenome with closely related species was visualized using MCscanX [30]. A total of 29 complete mitogenomes were downloaded (https://www.ncbi.nlm.nih.gov/) to ascertain the phylogenetic position of *P. davidiana*. We aligned the 32 shared genes of analyzed species in Muscle [31] and constructed the Maximum likelihood (ML) phylogenetic tree with MEGA X (1000 bootstrap replications) [25].

Results

P. davidiana mitogenome features

The *P. davidiana* mitogenome was sequenced, assembled, and annotated (Table S1). The graph-based mitochondrial genome of *P. davidiana* was assembled in 15 linear segments (molecules 1-15) with a total length of 375,671 bp and the GC content was 45.5%, as shown in Figure 1. The segments ranged in size from 680 bp to 65,093 bp, with similar depth of coverage (Table S2 and Figure 2). The *P. davidiana* genome sequence was submitted to the GenBank database (SRR25468515).

There were 64 genes annotated in the *P. davidiana* mitogenome, including 39 protein-coding genes (PCGs), 22 tRNA genes, and 3 rRNA genes (Figure 2 and Table 1). Among the 39 PCGs, six contained introns *(nad1*, *nad2*, *nad4*, *nad7*, *ccmFc*, and *rps3*). Interestingly, two-copy genes (*nad1*, *nad2* and *cox1*) and a three-copy gene (*nad5*) were found. Additionally, we also observed five tRNA genes and one rRNA gene located in repeat sequences (*trnN-GTT*, *trnM-CAT*, *trnP-TGG*, *trnH-GTG*, *trnW-CCA*, and *rrn5*) (**Fig. 1**).

Codon usage analysis of PCGs

Codon usage analysis of 39 PCGs was performed. The codon usage of each amino acid is shown in Figure 3. PCGs in *P. davidiana* had a total length of 30,234 nucleotides. The typical ATG start codon was found in the most PCGs; while *nad4L* and *nad1* were ACG, which may be altered by C-to-U RNA editing (Table 2). As with the status in other mitogenomes, TAA, TGA, and TAG served as stop codons.

The relative synonymous codon usage (RSCU) of 39 PCGs was also analyzed. The 39 PCGs comprised 30,234 bp encoding 10,078 codons, excluding termination codons. The RSCU >1 of Codons were considered to be used preferentially by amino acids. All RSCU values of NNT and NNA codons, excluding Ile (ATA), Leu (CTA), and Thr (ACA), were higher than 1.0. The result indicates the existence of a strong As or Ts bias at the third codon position in *P. davidiana* mitogenomes. The phenomenon also exists in other mitogenomes of plant species [15].

Analysis of synonymous and nonsynonymous substitution ratios

The nonsynonymous to synonymous substitution ratios (Ka/Ks) between any two species among six species (*Malus domestica*, *Sorbus aucuparia*, *Rhaphiolepis bibas*, *Pyrus betulifolia*, *Fragaria orientalis*, and *P. davidiana*) were calculated based on the 32 shared genes in Rosaceae (Table S5). According to Figure 4, the mean values of pairwise Ka/Ks of *ccmFn*, *rps4*, and *adh4* were higher than those of other genes. The result suggests that these genes have been under positive selection in the six plant species during evolution.

RNA editing sites Prediction in PCGs

RNA editing events, which are posttranscriptional processes, are enriched in metagenomes [32]. A total of 502 potential RNA editing sites were identified on 32 mitochondrial PCGs according to predictions from the online website PREP suit (http://prep.unl.edu/)(cutoff value = 0.2). All of them were observed to be C-to-U base editing. As shown in Figure 5, the *nad4* gene encoded the most RNA editing sites (40 sites), whereas *nad4L* and *rps7* only encoded one. Among the 502 predicted sites, the results were hydrophilic to hydrophobic (13.15%; 66 sites), hydrophobic to hydrophilic (47.01%; 236 sites), hydrophilic to hydrophilic (7.97%; 40 sites), hydrophobic to hydrophobic (31.27%; 157 sites), and hydrophilic to stop (0.60%, 3 sites).

Analysis of repeats in the P. davidiana mitogenome

Repeated sequences are widespread in the mitochondrial genome, mainly including SSRs, tandem repeats, and dispersed repeats [33]. SSRs are tandem-repeated motifs of one to six bases that are usually used as molecular markers in studying and identifying species, and in analyzing genetic diversity [34]. In

the present study, a total of 132 SSRs were identified in the *P. davidiana* mitogenome, including 49 (37.12%) mono-, 20 (15.15%) di-, 11 (8.33%) tri-, 42 (31.82%) tetra-, 7 (5.30%) penta-, and 3 (2.27%) hexanucleotide repeats (Table 2). More than 68.94% belonged to monomers and tetramers among the 132 SSRs. In addition, the results showed that 69.39% of monomers were A/T contents. In previous studies, the proportion of A/T in the SSRs was contributed to that A/T in the whole mitogenomes [33, 35] .

In addition to the SSRs, 187 forward, 219 palindromic, and 20 tandem repeats were detected in the *P. davidiana* mitogenome. The whole repeats length was 11,574 bp, which accounted for 3.08% of the whole *P. davidiana* mitogenome. Most of the forward and palindromic repeats ranged from 31 to 50 bp, the longest one was 458 bp, whereas tandem repeats were shorter than 39 bp (Figure 6 and Table S3 and Table S4).

Phylogenetic and Synteny analysis

The largest co-linear blocks of 35.0 and 34.9 kb were both identified in the dot plot with *P. avium* and *P. salicina,* respectively (Figure 7). In addition, the co-linear blocks were not arranged in the same order, a larger number of homologous co-linear blocks were detected between *P. davidiana* and the closely related species (Figure 8). A total of 74 homologous co-linear blocks (>500 bp) were identified between *P. davidiana* and *P. avium*, the largest block was 29,172 bp. Similarly, a total of 85 homologous co-linear blocks (>500 bp) were identified between *P. davidiana* and *P. salicina*, with the largest block being 8728 bp in length. The results also indicated that the mitogenomes were extremely unconserved in structure.

The mitogenomes provides an opportunity to confirm plant phylogenetic positions. In the present study, to further explore the evolutionary relationships of *P. davidiana* mitochondria, 31 plant mitogenomes were downloaded from the GenBank database (https://www.ncbi.nlm.nih.gov/genome/browse/#!/overview/). The 31 conserved single-copy orthologous genes present in all 32 mitogenomes were selected to construct phylogenetic tree. Five monocotyledonous species were designated as the outgroup. As shown in Figure 9, 25 out of 27 nodes in the generated tree had bootstrap support values > 70%, and 17 nodes were supported 100%. The phylogenetic tree strongly supports (bootstrap support = 100%) the close phylogenetic relationship between *P. davidiana* and *P. avium* and *P. salicina*. Overall, the results provide a valuable foundation for future analyses of the phylogenetic affinities of Rosaceae species.

Chloroplast and nuclear genes to mitogenome transfer events

DNA fragment transfers are common events during plant evolution among nuclear and organellar genomes. The nuclear and cp genomes of *P. davidiana* were searched using its mitogenome sequences as queries, to further understand the sequence transfer event characteristics. The 303.0 kb sequences were obtained from nuclear genome transferred into the mitogenome (Fig. 10A). In addition, the greatest length was 19,266 bp, and they were mainly between 200 bp and 400 bp (Fig. 10B). The 19 complete genes (*rps13*, *trnP-TGG*, *trnF-GAA*, *trnS-GCT*, *cob*, *rps14*, *rpl5*, *trnH-GTG*, *trnK-TTT*, *ccmB*, *trnN-ATT*, *trnK-TTT*, *ccmC*, *trnD-GTC*, *rps7*, *nad6*, *trnN-GTT*, *rps1*, and *rps4*) were contained in the shared sequences.

The *P. davidiana* mitogenome sequence (375,671 bp) was approximately 2.37 times longer than the cp genome sequence (158,055 bp). The results showed that 16 fragments with a total length of 4,238 bp had migrated from the cp genome to the mitogenome in *P. davidiana* (Table 3). Three of these fragments were more than 500bp in length, and the longest one was 863bp (Figure 11). Eight intact cp genes (*psaJ*, *petN*, *psbJ*, *trnD-GUC*, *trnN-GUU* , *trnW-CCA*, *trnP-UGG*, and *trnH-GUG*) were identified in the fragments. Others were partial sequences of transferred genes (Table S6). The transferred genes would aided the movement of genetic material throughout *Prunus*.

Discussion

Characterization of the P. davidiana mt genome

Mitochondria are vital organelles in organisms that provide energy for the physiological activities of various cells. Studies have reported that plant mitogenomes are more complex than animal mitogenomes in size variation, repetitive content, protein coding sequences, and structure [36]. In previous studies, the circular, linear, branched, and numerous smaller circular molecule structures of mitogenomes have been reported [9]. In the present study, we assembled and characterized the *P. davidiana* mitogenome in detail. The size of the *P. davidiana* mitogenome was 375,671 bp, which is similar to that of *M. domestica* (396,947 bp) [37]. The GC content is an important for species assessment, the GC content of *P. davidiana* was 45.5%, which is similar to the typically stable 43%-46% in higher plants [15, 35]. This supports that GC content is highly conserved in higher plants.

Identification of repeat sequences and RNA editing sites

Repeated sequences are widespread in the mitogenome, including two main categories: tandem and interspersed repeats [22]. They are important for developing markers for population and evolutionary studies [38, 39, 40]. In the present study, forward, palindromic, and tandem repeats were identified in the *P. davidiana* mitogenome. The longest repeat was 458 bp and the A/T proportion was higher in many SSR repeats. Such repeats would accelerate intermolecular recombination in the mitogenomes during Rosaceae evolution [36].

In plants, RNA editing plays an important role in protein folding, which is enriched in mitochondrial and cp genomes, especially, the cytidine (C)-to-uridine (U) RNA editing sites [33, 34]. Previous research has uncovered approximately 491 RNA editing sites within 34 genes in *Oryza sativa* L. [41], 486 RNA editing sites within 31 genes in *Phaseolus vulgaris* L. [33], and 421 RNA editing sites within 26 genes in *Acer truncatum* [32]. In the present study, 502 RNA editing sites were identified in 32 PCGs in the *P. davidiana* mitogenome. The results showed that the number of RNA editing sites varies greatly between the mitogenomes of plant species. Interestingly, all were C-U RNA editing sites. However, the largest numbers of RNA editing sites were associated with cytochrome c biogenesis and NADH dehydrogenase genes, a trend similar to that in *A. truncatum* [34] and *Clematis acerifolia* [40].

DNA fragment transfer events

DNA fragment transfer events are common in the mitochondria, nuclei, and cp of angiosperms. Furthermore, previous studies have suggested that the most prominent transfer direction is from organellar genomes to the nuclear genomethe, and then from chloroplast genomes and nuclei to the mitogenome [33, 42, 43]. According to the results of the present study, we found that the 303.0 kb sequences of nuclear DNA transferred into the 15 molecules in the *P. davidiana* mitogenome. Furthermore, 16 fragments were identified, which had been transferred from the cp genome to the mitogenome.

To explore homologous covariance blocks among the related species, the arrangement of homologous genes or sequences was detected using a covariance study. According to the results, the largest co-linear blocks of 35.0 and 34.9 kb were both identified in the dot plot with *P. avium* and *P. salicina,* respectively. Meanwhile, an inconsistent order of arrangement of co-linear blocks was observed. The findings suggest that the *P. davidiana* mitogenome has undergone extensive genomic rearrangements.

Conclusions

In this study, we sequenced, assembled and annotated the mitogenome of *P. davidiana*. Detailed analyses of the mitogenome were performed based on DNA and amino acid sequences. The total length of the *P. davidiana* mitogenome is 375,671 bp, which including 15 linear molecules, with a 45.5% GC content. We annotated 64 genes, including 39 PCGs, 22 tRNA genes, and 3 rRNA genes. In addition, codon usage, repetitive sequences, Ka/Ks, RNA editing, synteny, and DNA fragment transfer events were analysed. To confirm the evolutionary status, we also performed a phylogenetic analysis of the *P. davidiana* mitogenome and 31 other plants. The results of this study provide valuable information and a theoretical basis for future *P. davidiana* research and breeding activities.

Abbreviations

PCR: Polymerase chain reaction; rRNA:Ribosomal RNA; Ka/Ks: Snonsynonymous-to-synonymous substitution ratio; Leu: Leucine; DNA: Deoxyribonucleic acid; tRNA:Tranfer RNA; PCGs:protein-coding genes; Ile: Isoleucine; Ser: Serine Thr:Threonine; SSRs:Simple sequence repeat; PCG: Protein-coding genes; RSCU:Relative synonymous codon usage; mitogenomes: mitochondrial genomes; cp gemome: chloroplast genome; Arg: Arginine; Cys: Cysteine.

Declarations

Acknowledgments:Thanks to all the members of the Institute of Pomology team for their contributions.

Authors' Contributions

YYZ, XG, RJM designed the project and the strategy, MLY, AHR and JLX contributed to plant sample collection; YYZ, AHR and ZJS work on genome assembly, annotation and comparative analyses; YYZ and XG wrote and revised the manuscript. All authors read and approved the final manuscript.

Funding

This work was supported by Jiangsu Key Research and Development Program (Modern Agriculture) (BE2022381), the National Natural Science Foundation of China (31901979) and the National Key Research and Development Program subproject of China (2019YFD1000801-02).

Data availability

The supporting datasets and supplemented materials are included within the article as additional files. The *P. davidiana* Mitochondrial genome sequences were submitted in the GenBank database (No. SRR25468515).

Declarations

Ethics approval and consent to participate our experimental research on the collection and handling of 'DXS1' in accordance with the relevant national/institutional guidelines.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

1. Liu J, Liu Y Q, Jing L I, Sun S X, Wang Y Q. Effects of alkali stress on *Prunus davidiana (Carr.)* leaf morphological structure and photosynthetic characteristics. Southwest China J. Agri. Sci. 2017; 30:327–332. https://doi.org/10.16213/j.cnki.scjas.2017.2.015.
2. Hesse CO. Advances in fruit breeding Peaches. Purdue University Press. West Lafayette, Indiana, pp.1975; 325–326. https://doi.org/10.1016/S0304-4238(76)80013-1.
3. Zhang J H, Wang J Y, Hu Y L, Wang Y S, Zhu S H, Mao Z Q. Root lifespan and root respiration of fruit of fruit tree species. Acta. Hortic. 2008; 772: 269. https://doi.org/10.17660/ActaHortic.2008.772.42.
4. Xu D, Li H, Lin L, Liao M, Xia H. Effects of carboxymethyl chitosan on the growth and nutrient uptake in *Prunus davidiana* seedlings. Physiol. Mol. Biol. Plants. 2020; 26: 661–668. https://doi.org/0.1007/s12298-020-00791-5.
5. Ye N, Wang X, Li J, Bi C, Xu Y, Wu D, et al. Assembly and comparative analysis of complete mitochondrial genome sequence of an economic plant *Salix suchowensis*. Peer. J. 2017; 5: e3148.
6. Ogihara Y, Yamazaki Y, Murai K, Kanno A, Terachi T, Shiina T, et al. Structural dynamics of cereal mitochondrial genomes as revealed by complete nucleotide sequencing of the wheat mitochondrial genome. Nucleic Acids Res. 2005; 33: 6235–6250. https://doi.org/10.1093/nar/gki925.
7. Li J, Li J, Ma Y, Kou L,Wei J, Wang W. The complete mitochondrial genome of okra (*Abelmoschus esculentus*): using nanopore long reads to investigate gene transfer from chloroplast genomes and rearrangements of mitochondrial DNA molecules. BMC Genomics. 2022; 23(1):1-14. https://doi.org/10.1186/s12864-022-08706-2.

8. Ma Q, Wang Y, Zhu L, Bi C, Li S, Li S. Characterization of the Complete Chloroplast Genome of *Acer truncatum* Bunge (*Sapindales: Aceraceae*): A New Woody Oil Tree Species Producing Nervonic Acid. BioMed. Res. Int. 2019; (11/12):1-13. https://doi.org/10.1155/2019/7417239.
9. Sloan DB. One ring to rule them all? Genome sequencing provides new insights into the 'master circle' model of plant mitochondrial DNA structure. New Phytol. 2013; 200(4):978–85. https://doi.org/10.1111/nph.12395.
10. Sloan D, Alverson A, Chuckalovcak J, Wu M, Mccauley D. Rapid evolution of enormous, multichromosomal genomes in flowering plant mitochondria with exceptionally high mutation rates. PLoS Biol. 2012; 10(1): e1001241. https://doi.org/10.1371/journal.pbio.1001241.
11. Skippington E, Barkman T, Rice D, Palmer J. Miniaturized mitogenome of the parasitic plant *Viscum scurruloideum* is extremely divergent and dynamic and has lost all nad genes. P. Natl. Acad. Sci. USA. 2015; 112(27): E3515. https://doi.org/10.1073/pnas.1504491112.
12. Zou Y, Zhu WD, Sloan D.B, Wu ZQ. Long-read sequencing characterizes mitochondrial and plastid genome variants in *Arabidopsis* msh1 mutants. Plant J. 2022; 112:738–755. https://doi.org/10.1111/tpj.15976.
13. Christensen AC. Plant mitochondrial genome evolution can be explained by DNA repair mechanisms. Genome Biol. Evol. 2013; 5(6):1079–86. https://doi.org/10.1093/gbe/evt069.
14. Bergthorsson U, Adams K, Thomason B, Palmer J. Widespread horizontal transfer of mitochondrial genes in flowering plants. Nature. 2003; 424(6945): 197-201. https://doi.org/10.1038/nature01743.
15. Ma Q, Wang Y, Li S, Wen J, Zhu L, Yan K, et al. Assembly and comparative analysis of the first complete mitochondrial genome of *Acer truncatum* Bunge: a woody oil-tree species producing nervonic acid. BMC Plant Biol. 2022; 22(1):29. https://doi.org/10.1186/s12870-021-03416-5.
16. Wynn E, Christensen A. Repeats of Unusual Size in Plant Mitochondrial Genomes: Identification, Incidence and Evolution. G3: Genes, Genomes, Genetics. 2019; 9:549-559. https://doi.org/10.1534/g3.118.200948.
17. Zhang X, Yan J, Ling Q, et al. Complete Chloroplast Genome Sequence Of *Prunus Davidiana (Rosaceae)*. 2023. https://doi.org/10.1080/23802359.2018.1501325.
18. Varré JS, D'Agostino N, Touzet P, Gallina S, Tamburino R, Cantarella C, et al. Complete sequence, multichromosomal architecture and transcriptome analysis of the *Solanum tuberosum* mitochondrial genome. Int. J. Mol. Sci. 2019; 20(19):4788. https://doi.org/10.3390/ijms20194788.
19. Bi C, Lu N, Xu Y, et al. Characterization and Analysis of the Mitochondrial Genome of Common *Bean (Phaseolus vulgaris)* by Comparative Genomic Approaches[J]. Int. J. Mol. Sci. 2020; 21(11):3778. https://doi.org/10.3390/ijms21113778.
20. Ma Q, Li S, Bi C, Hao Z, Sun C, Ye N. Complete chloroplast genome sequence of a major economic species, *Ziziphus jujuba (Rhamnaceae)*. Curr. Genet. 2017; 63:117-129. https://doi.org/10.1007/s00294-016-0612-4.
21. Ye N , Wang X , Li J ,et al.Assembly and comparative analysis of complete mitochondrial genome sequence of an economic plant *Salix suchowensis*. Peer J. 2017; 5(1):e3148.

https://doi.org/10.7717/peerj.3148.

22. Cheng Y, He X, Priyadarshani S, Wang Y, Y L, Shi C, Ye K, et al. Assembly and comparative analysis of the complete mitochondrial genome of *suaeda glauca*. BMC Genomics. 2021; 22(1):167. https://doi.org/10.1186/s12864-021-07490-9.
23. Goremykin V V, Lockhart P J , Viola R, et al. The mitochondrial genome of *Malus domestica* and the import-driven hypothesis of mitochondrial genome expansion in seed plants. Plant J. 2012; 71(4):615-626. https://doi.org/10.1111/j.1365-313X.2012.05014.x.
24. Gualberto JM, Mileshina D, Wallet C, Niazi AK, Weber-Lotfi F, Dietrich A. The plant mitochondrial genome: dynamics and maintenance. Biochimie. 2014; 100:107–20. https://doi.org/10.1016/j.biochi.2013.09.016.
25. Xiong Y, Lei X, Bai S. Xiong Y, Liu W, Wu W, et al. Genomic survey sequencing, development and characterization of single- and multi-locus genomic SSR markers of *Elymus sibiricus* L. BMC Plant Biol. 2021, 21: 3. https://doi.org/10.1186/s12870-020-02770-0.
26. Liu D, Qu K, Yuan YC, Zhao ZH, Chen Y, Han B, et al. Complete sequence and comparative analysis of the mitochondrial genome of the rare and endangered *Clematis acerifolia*, the first clematis mitogenome to provide new insights into the phylogenetic evolutionary status of the genus. Front. Genet. 2023 , 13, 1050040. https://doi.org/10.3389/fgene.2022.1050040
27. Clifton S.W, Minx P, Fauron C.M.-R, Gibson M, Allen J.O, Sun H, et al. Sequence and Comparative Analysis of the *Maize* NB Mitochondrial Genome. Plant Physiol. 2004; 136:3486–3503. https://doi.org/10.2307/4356703.
28. Rice D, Alverson A, Richardson A, Young G, Sanchez-Puerta M, Munzinger J, et al. Horizontal transfer of entire genomes via mitochondrial fusion in the angiosperm *Amborella*. Science. 2013; 342(6165):1468-73. https://doi.org/10.1126/science.1246275.
29. Zhao N, Grover C, Chen Z, Wendel J, Hua J. Intergenomic gene transfer in diploid and allopolyploid *Gossypium*. BMC Plant Biol. 2019; 19(1):492. https://doi.org/10.1186/s12870-019-2041-2.
30. Arseneau JR, Steeves R, Laflamme M. Modified low-salt CTAB extraction of high-quality DNA from contaminant-rich tissues. Mol. Ecol. Resour. 2017; 17(4):686-93. https://doi.org/10.1111/1755-0998.12616.
31. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics (Oxford, England). 2014; 30(15):2114–20. https://doi.org/10.1007/s11515-015-1364-0.
32. Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics, 34:3094-3100. doi:10.1093/bioinformatics/bty191.
33. Koren S, Walenz BP, Berlin K, Miller JR, Phillippy AM. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Research. (2017). https://doi.org/10.1101/gr.215087.11.
34. Langmead B, Salzberg S L. Fast gapped-read alignment with Bowtie 2. Nature methods, 2012(4). https://doi.org/10.1038/nmeth.1923.

35. Schattner P, Brooks AN, Lowe TM. The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. Nucleic Acids Res. 2005; 33:W686–W689. https://doi.org/10.1093/nar/gki366.
36. Lewis S.E, Searle S, Harris N, Gibson, M, Iyer V, Richter J, et al. Apollo: A sequence annotation editor. Genome Biol. 2002; 3: research0082. https://doi.org/10.1093/nar/gki366.
37. Kumar S, Stecher G, Tamura K. MEGA7: Molecular evolutionary genetics analysis version 7. 0 for bigger datasets. Mol. Biol. Evol. 2016; 33: 1870–1874. https://doi.org/10.1093/molbev/msw054.
38. Thiel T, Michalek W, Varshney R, Graner A. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). Theor. Appl. Genet. 2003; 106:411-22. https://doi.org/10.1007/s00122-002-1031-0.
39. Kurtz S, Choudhuri J, Enno O, Chris S, Jens S, Robert G. REPuter: the manifold applications of repeat analysis on a genomic scale, Nucleic Acids Res. 2001; 29(22): 4633-42. https://doi.org/10.1093/nar/29.22.4633.
40. Librado, Rozas J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics. 2009; 25(11): 1451-1452. https://doi.org/10.1093/bioinformatics/btp187.
41. Cheng Y, He X, Priyadarshani S.V.G.N, Wang Y, Ye L, Shi C, et al. Assembly and comparative analysis of the complete mitogenomes of *Suaeda Glauca*. BMC Genom. 2021; 22: 167. https://doi.org/10.1186/s12864-021-07490-9.
42. Wang Y, Tang H, DeBarry J.D, Tan X, Li J, Wang X, et al. MCScanX: A toolkit for detection and evolutionary analysis of gene synteny and collinearity. Nucleic Acids Res. 2012; 40:e49. https://doi.org/10.1093/nar/gkr1293.
43. Edgar R. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004; 32(5): 1792-1797. https://doi.org/10.1093/nar/gkh340.

Tables

Table 1. The characteristics of genes in *P. davidiana* mitogenome.

Group	Name	Length	Start codon	Stop codon	Amino acids
ATP synthase	*atp1*	1524	ATG	TGA	507
	atp4	597	ATG	TAG	198
	atp6	774	ACG	TAA	282
	atp8	480	ATG	TAA	159
	atp9	225	ATG	TGA	74
NADH dehydrogenase	*nad1(2)*[a]	978	ACG	TAA	325
	nad2(2)[a]	1467	ATG	TAA	488
	nad3	357	ATG	TAA	118
	nad4[a]	1488	ATG	TGA	495
	nad4L	303	ACT	TAA	100
	nad5(3)	2013	ATG	TAA	670
	nad6	618	ATG	TAA	205
	nad7[a]	1185	ATG	TAG	394
	nad9	573	ATG	TAA	190
Cytochrome c biogenesis	*ccmB*	621	ATG	TGA	206
	ccmC	753	ATG	TGA	250
	ccmFc[a]	1320	ATG	TAG	439
	ccmFn	1731	ATG	TGA	576
Maturases	*matR*	1968	ATG	TAG	655
Ubichinol cytochrome c reductase	*cob*	1182	ATG	TGA	393
Cytochrome c oxidase	*cox1(2)*	1566	ATG	TAA	521
	cox2	783	ATG	TAA	260
	cox3	798	ATG	TGA	265
Transport membrane protein	*mttB*	792	ATA	TAG	264
Ribosomal proteins (LSU)	*rpl5*	555	ATG	TAA	184
	rpl10	453	ATG	TAA	150

Ribosomal proteins (SSU)	*rps3*[a]	1650	ATG	TAA	549
	rps4	831	ATG	TAA	276
	rps1	642	ATG	TAA	213
	rps7	447	ATG	TAA	148
	rps12	378	ATG	TGA	125
	rps13	351	ATG	TGA	116
Succinate dehydrogenase	*sdh3*	324	ATG	TGA	107
	sdh4	387	ATG	TGA	128
Transfer RNAs	*trnY-GTA*	83	_	_	_
	trnN-GTT[b]	72	_	_	_
	trnC-GCA	71	_	_	_
	trnS-GCT(3)	65/88	_	_	_
	trnF-GAA	74	_	_	_
	trnK-TTT	73	_	_	_
	trnM-CAT(3)	73/74/77	_	_	_
	trnP-TGG[b]*(2)*	74/75	_	_	_
	trnE-TTC	72	_	_	_
	trnW-CCA[b]	74	_	_	_
	trnS-GCT	65/88	_	_	_
	trnS-TGA	87	_	_	_
	trnD-GTC(2)[b]	70	_	_	_
	trnQ-TTG	72	_	_	_
	trnG-GCC	72	_	_	_
	trnH-GTG[b]	74	_	_	_
Ribosomal RNAs	*rrn5*	118	_	_	_
	rrn18	1863	_	_	_
	rrn26	3156	_	_	_

Note: The number of copies are showed after gene names. The genes containing introns and chloroplast-derived genes showed with a and b indicate, respectively.

Table 2. The SSR motifs frequency in the *P. davidiana* mitogenome.

Motif Type	Number													Total	Proportion (%)
	3	4	5	6	7	8	9	10	11	12	13	15	16		
Monomer	-	-	-	-	-	-	-	26	6	10	4	1	2	49	37.12
Dimer	-	-	16	-	1	1	2	-	-	-	-	-	-	20	15.15
Trimer	-	8	3	-	-	-	-	-	-	-	-	-	-	11	8.33
Tetramer	40	1	1	-	-	-	-	-	-	-	-	-	-	42	31.82
Pentamer	7	-	-	-	-	-	-	-	-	-	-	-	-	7	5.30
Hexamer	3	-	-	-	-	-	-	-	-	-	-	-	-	3	2.27
Total	50	9	20	0	1	1	2	26	6	10	4	1	2	132	100

Table 3. Fragments transferred from chloroplasts to mitochondria in *P. davidiana.*

No.	Identity (%)	Mismatch	Gap opens	CP Start	CP End	Mt Start	Mt End	Gene
Molecule 1	74.103	171	46	139995	140858	29658	28799	rrn16 (partical)
Molecule 1	74.103	171	46	102892	103755	28799	29658	rrn16 (partical)
Molecule 2	83.294	45	16	68184	68589	60695	60278	trnW-CCAtrnP-UGG
Molecule 2	89.865	12	3	36207	36351	48887	49034	psbC (partical)
Molecule 2	94.937	4	0	53938	54016	5848	5926	trnM-CAU
Molecule 2	100	0	0	10543	10573	48146	48116	atpA (partica)
Molecule 3	82.847	62	15	28856	29374	15356	14812	petN
Molecule 3	85.627	39	3	66284	66604	32605	32281	psbE (partical)
Molecule 3	85.156	20	9	68831	69078	14686	14441	psaJ
Molecule 3	96.97	0	1	66248	66279	32657	32625	psbE (partica)
Molecule 4	81.771	25	6	31690	31878	27123	26939	trnD-GUC
Molecule 4	96.471	3	0	82	166	7395	7311	trnH-GUG
Molecule 6	87.342	6	3	88183	88257	4769	4691	trnI-CAU
Molecule 6	87.342	6	3	155493	155567	4691	4769	trnI-CAU
Molecule 9	96.512	2	1	110893	110978	6359	6275	trnN-GUU
Molecule 9	96.512	2	1	132772	132857	6275	6359	trnN-GUU

Figures

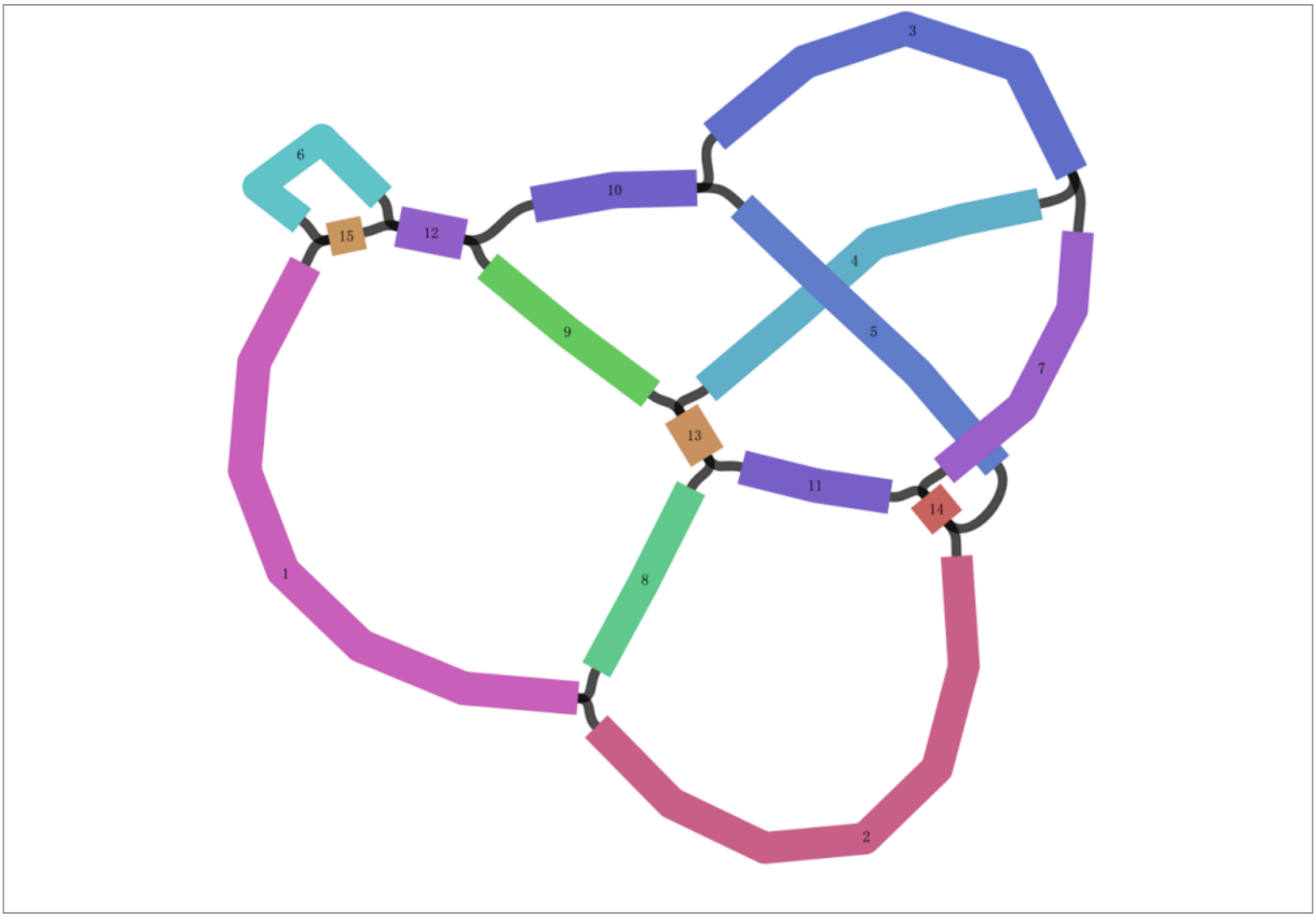

Figure 1

The assembly graph of the mitochondrial genome of *P. davidiana.* Each colored segments named 1-15, all segment adjacencies are supported by the long reads.

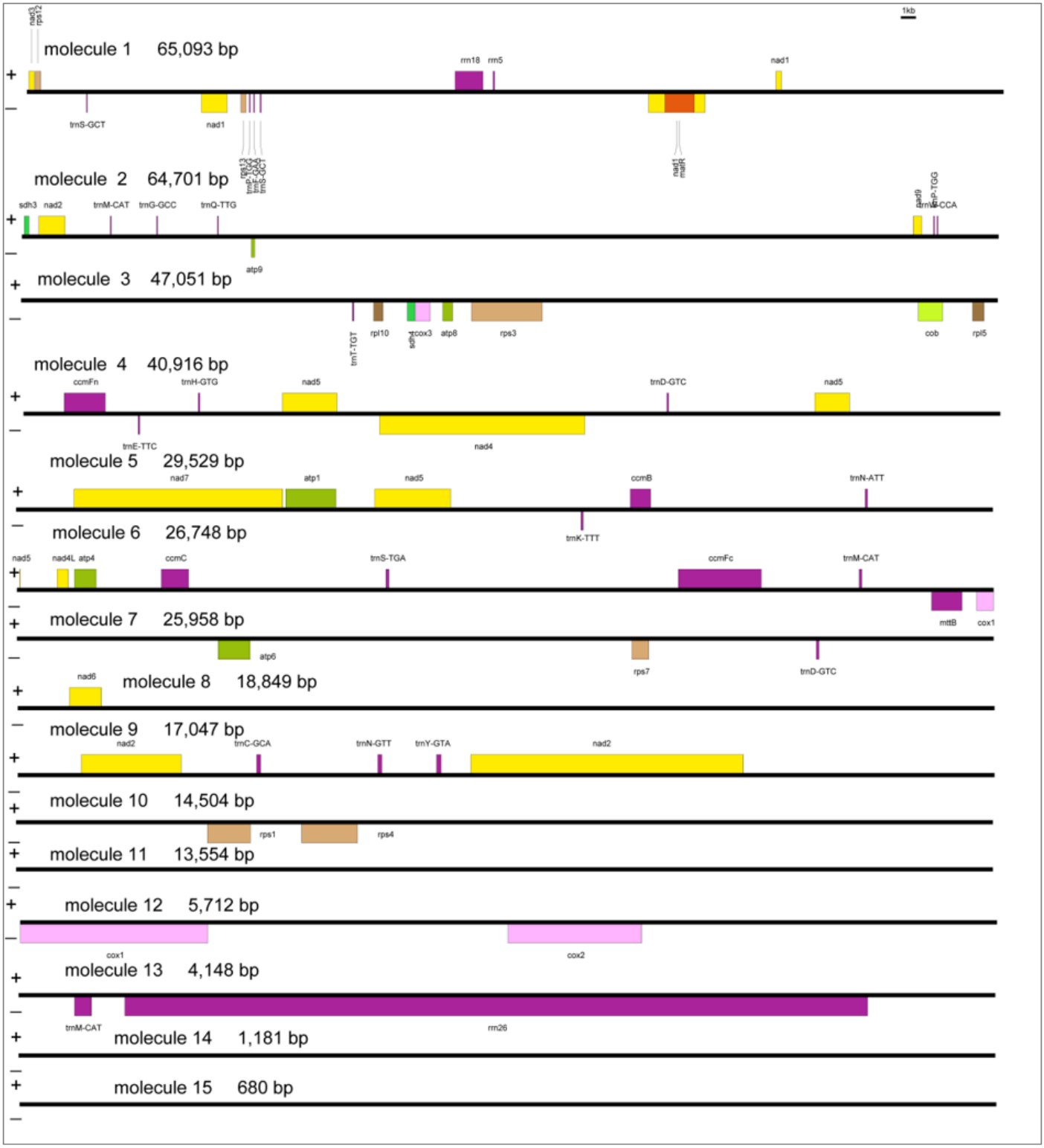

Figure 2

The order, orientation and size of gene within the *P. davidiana* mitogenome. Each box is proportional to size of the gene including introns. The annotated genes were not found in Molecule 11, 14 and 15.

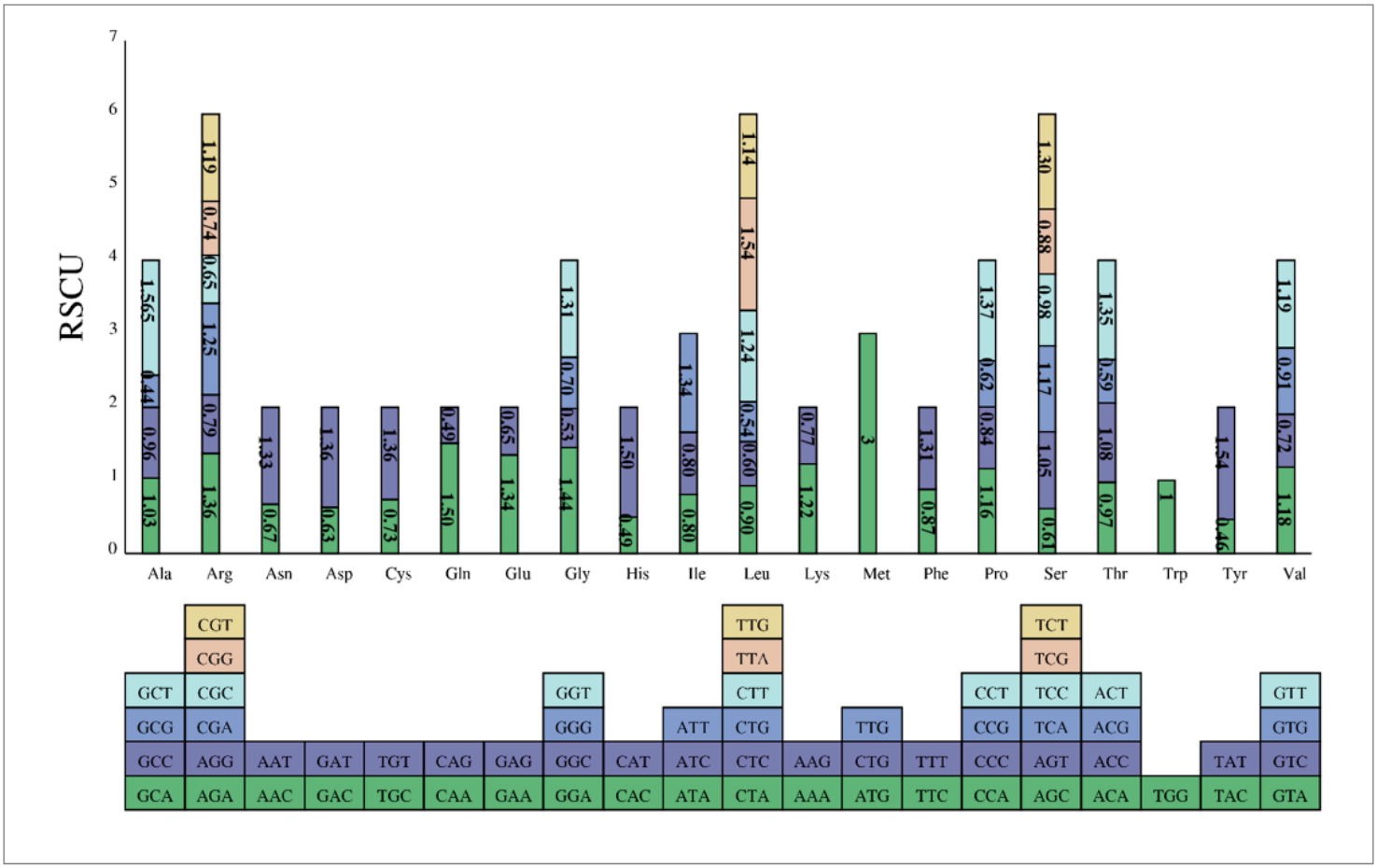

Figure 3

The RSCU analysisin the *P. davidiana* mitogenome. The *x*-axis was Codon families. RSCU values displayed on bar.

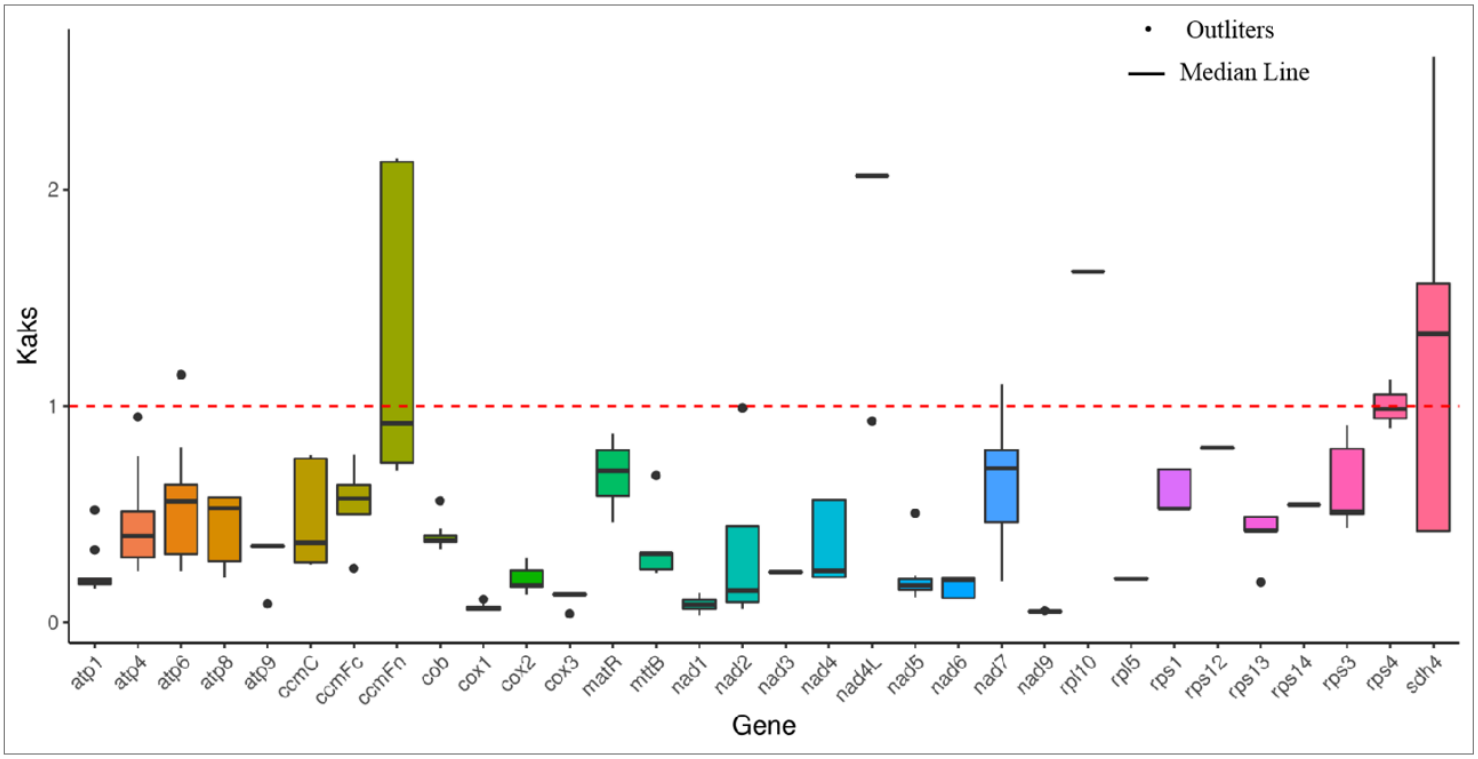

Figure 4

Ka/Ks calculated based on 32 shared genes among the six Rosaceae mitogenomne.

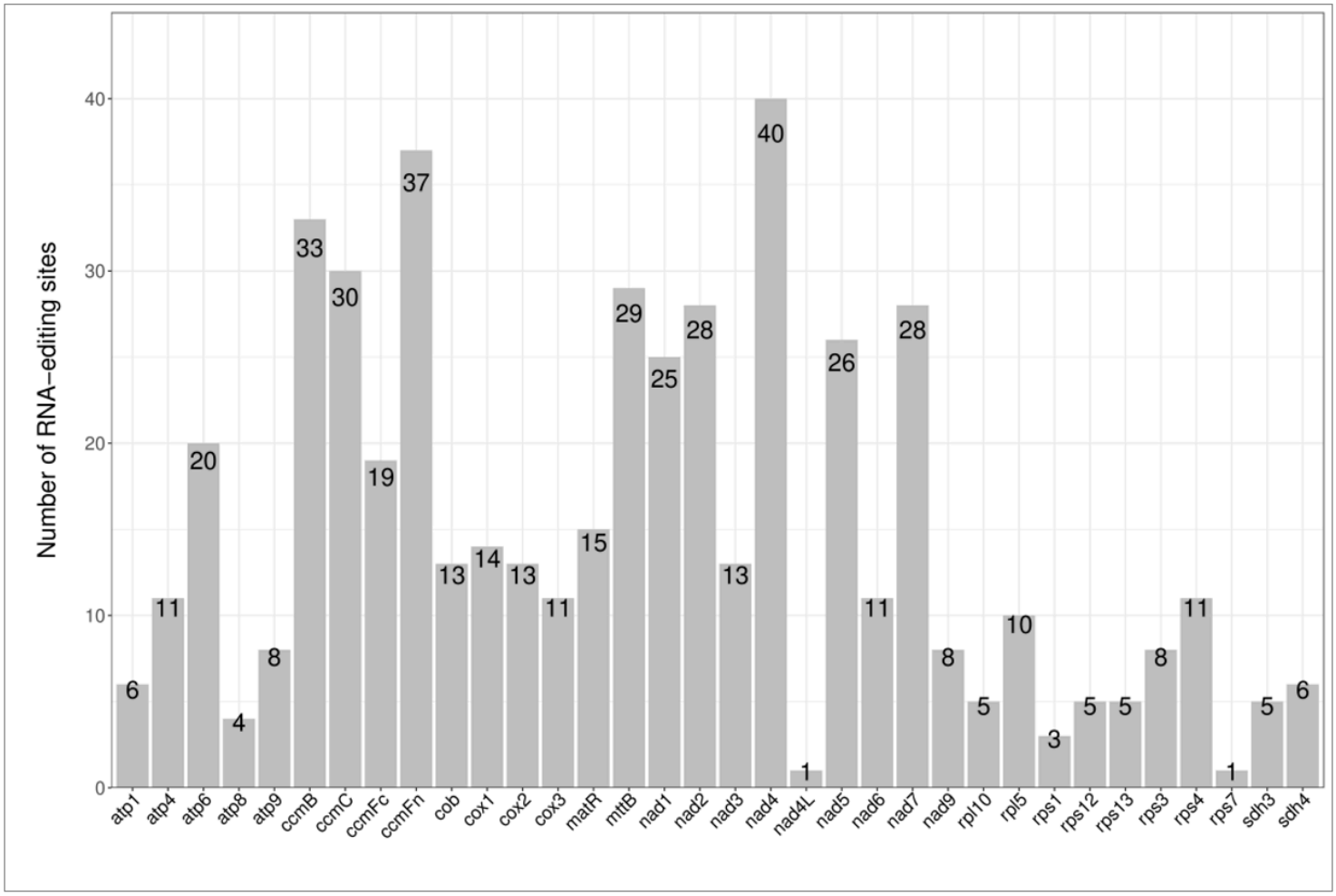

Figure 5

The RNA editing sites distribution in the *P. davidiana* mitogenome. The number shown by gray box represents the RNA editing sites of each gene.

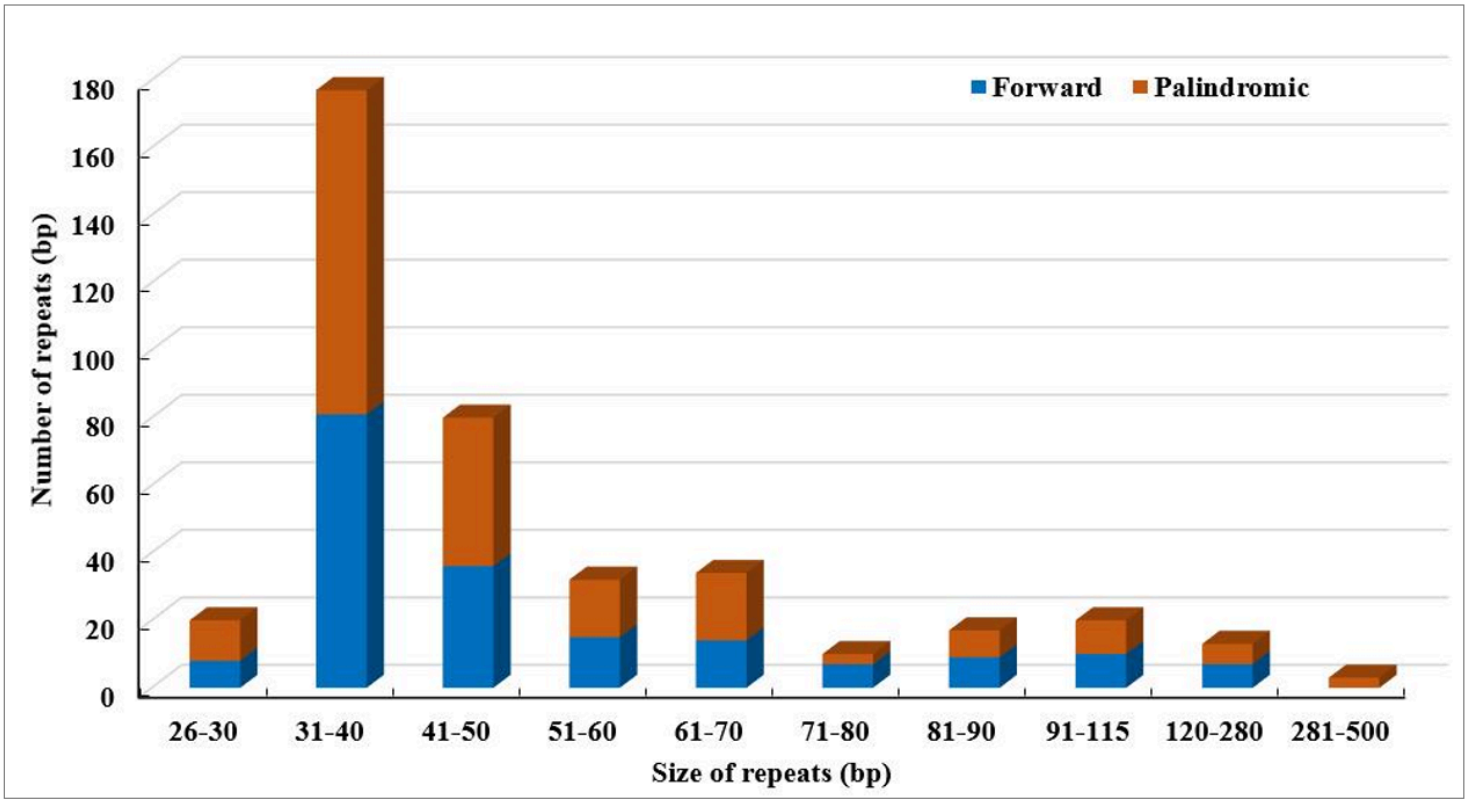

Figure 6

The frequency distribution of repeat lengths in the *P. davidiana* mt genome.

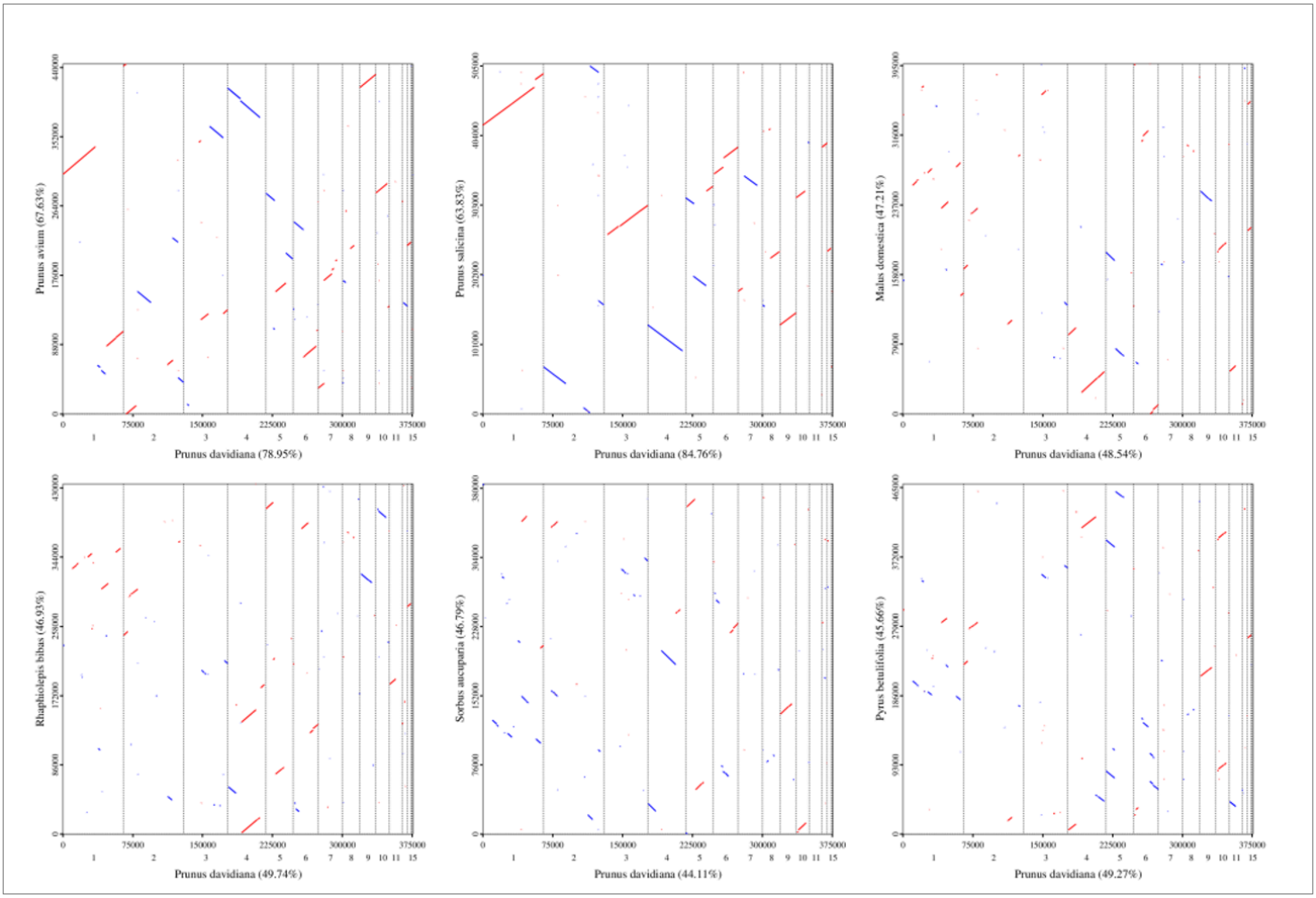

Figure 7

Dot-plots analysis of *P. davidiana* with closely related species.

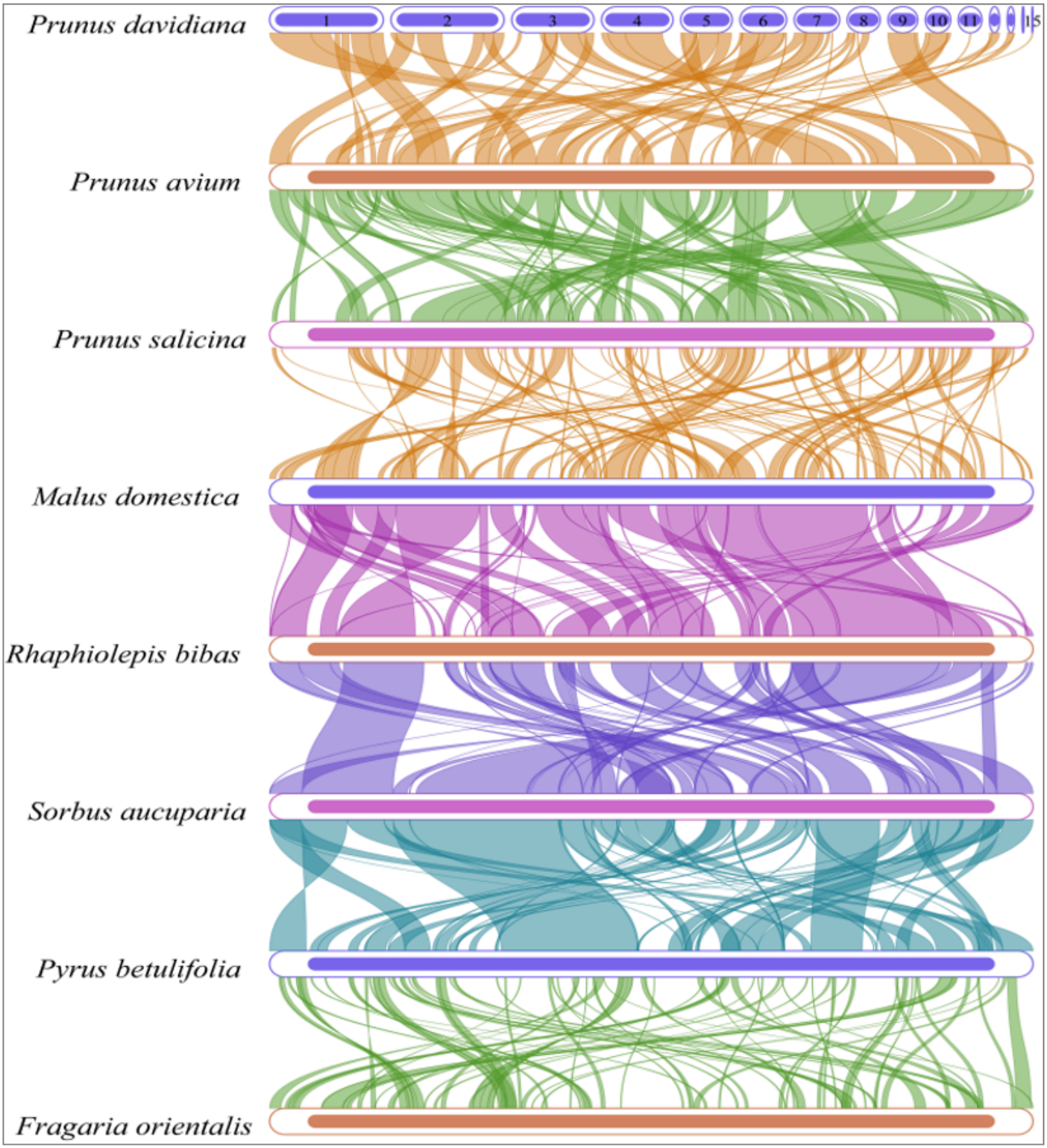

Figure 8

Synteny analysis of six mitogenomes. Bars indicated the mitogenomes, and the ribbons showed the homologous sequences between the adjacent species. The red areas indicate where the inversionoccurred, common blocks less than 0.5 kb in length are not retained.

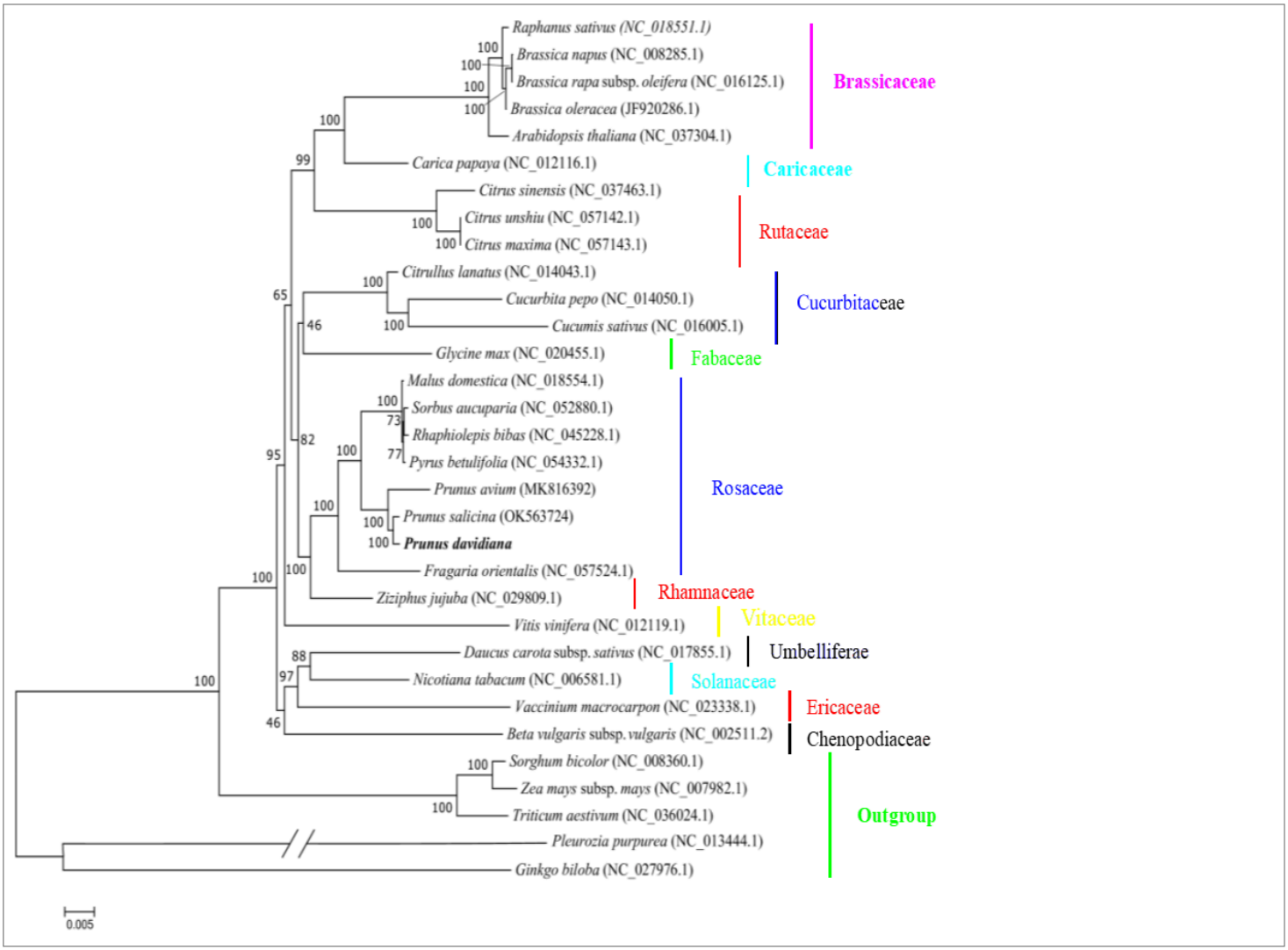

Figure 9

The phylogenetic tree based on 31 single-copy orthologous genes shared among 32 species. The bootstrap support values show at nodes . *Sorghum bicolor*, *Zea mays*, *Triticum aestivum*, Pleurozia *purpurea*, and *Ginkgo biloba* served as outgroups.

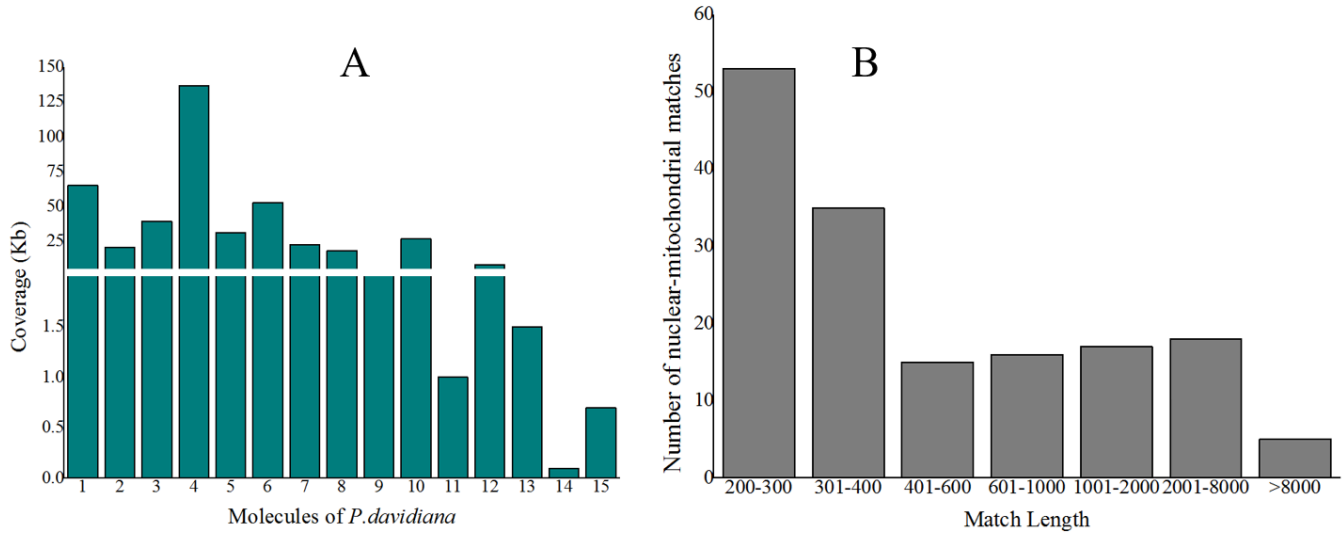

Figure 10

The nuclear-mitochondrial sequences analysisin *P. davidiana.* **A.** Distributions of sequencesshared nuclear-mitochondrial matches. **B**. Distributions of lengths between shared nuclear-mitochondrial matches.

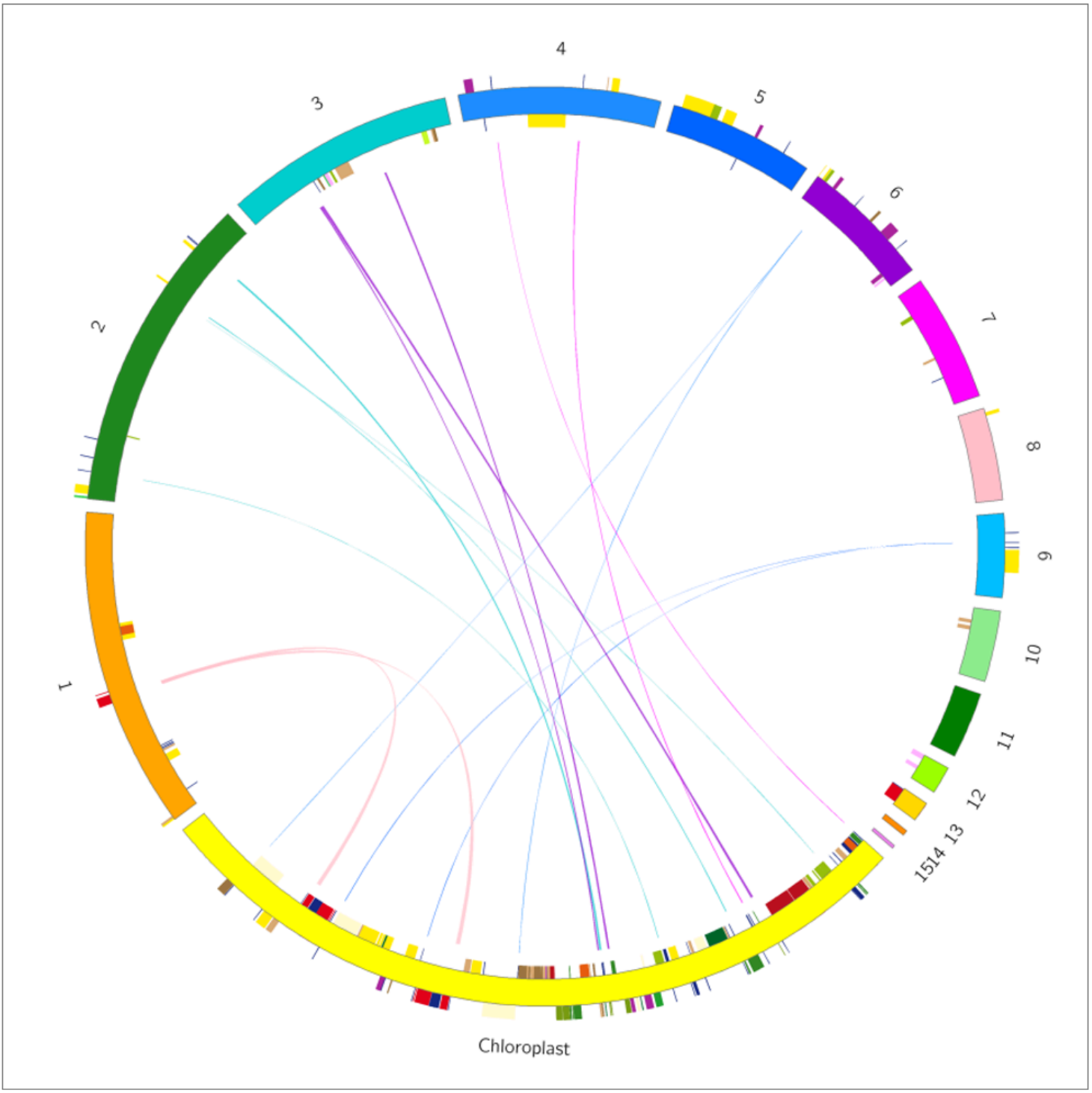

Figure 11

Schematic representation of gene transfers between chloroplast and mitogenome in.*P. davidiana.* The yellow arcs represent the chloroplast genome.

Supplementary Files

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

- TableS1.xlsx
- TableS2.xls
- TableS3.xls
- TableS4.xls
- TableS5.xls
- TableS6.xls