



Article

# Characterization and Analysis of the Mitochondrial Genome of Common Bean (*Phaseolus vulgaris*) by Comparative Genomic Approaches

Changwei Bi <sup>1</sup>, Na Lu <sup>1</sup>, Yiqing Xu <sup>2</sup>, Chunpeng He <sup>1,\*</sup> and Zuhong Lu <sup>1,\*</sup>

<sup>1</sup> State Key Laboratory of Bioelectronics, School of Biological Science and Medical Engineering, Southeast University, Nanjing 210096, Jiangsu, China; bichwei@seu.edu.cn (C.B.); nlu@seu.edu.cn (N.L.)

<sup>2</sup> School of Information Science and Technology, Nanjing Forestry University, Nanjing 210037, Jiangsu, China; yiqingxu@njfu.edu.cn

\* Correspondence: cphe@seu.edu.cn (C.H.); zhlu@seu.edu.cn (Z.L.)

Received: 6 May 2020; Accepted: 25 May 2020; Published: 27 May 2020



**Abstract:** The common bean (*Phaseolus vulgaris*) is a major source of protein and essential nutrients for humans. To explore the genetic diversity and phylogenetic relationships of *P. vulgaris*, its complete mitochondrial genome (mitogenome) was sequenced and assembled. The mitogenome is 395,516 bp in length, including 31 unique protein-coding genes (PCGs), 15 transfer RNA (tRNA) genes, and 3 ribosomal RNA (rRNA) genes. Among the 31 PCGs, four genes (*mttB*, *nad1*, *nad4L*, and *rps10*) use ACG as initiation codons, which are altered to standard initiation codons by RNA editing. In addition, the termination codon CGA in the *ccmF<sub>C</sub>* gene is converted to UGA. Selective pressure analysis indicates that the *ccmB*, *ccmF<sub>C</sub>*, *rps1*, *rps10*, and *rps14* genes were under evolutionary positive selection. The proportions of five amino acids (Phe, Leu, Pro, Arg, and Ser) in the whole amino acid profile of the proteins in each mitogenome can be used to distinguish angiosperms from gymnosperms. Phylogenetic analyses show that *P. vulgaris* is evolutionarily closer to the Glycininae than other leguminous plants. The results of the present study not only provide an important opportunity to conduct further genomic breeding studies in the common bean, they also provide valuable information for future evolutionary and molecular studies of leguminous plants.

**Keywords:** common bean; *Phaseolus vulgaris*; mitochondrial genome; comparative genomics; phylogeny

## 1. Introduction

Mitochondria (mt) are semi-autonomous organelles that are part of almost all eukaryotic cells (cells with clearly defined nuclei). Their primary function is to produce a steady supply of adenosine triphosphate (ATP). Mitochondria are thus termed the ‘powerhouses’ or ‘energy factories’ of cells. Chloroplasts (cp) and mitochondria most likely originated from formerly free-living bacteria through endosymbiotic acquisition, which can explain the presence of their own genomes [1,2]. With rapid developments in sequencing and genome assembly methods, an increasing number of complete organelle genomes have been assembled in the last decade. Thus far, over 4900 complete chloroplast and plastid genomes have been assembled but only 321 plant mitogenomes have been assembled and deposited in GenBank Organelle Genome Resources (as of 14 May 2020; <https://www.ncbi.nlm.nih.gov/genome/browse/>), suggesting that their assembly is complex and difficult.

Mitochondria are specific to each plant and have complex genome structures [3–5], variable genome sizes [6,7], numerous repetitive sequences [8,9], multiple RNA editing modifications [10,11], and frequent gene gains or losses during evolution [9,12,13]. In seed plant mitogenomes, the genome sizes are highly variable, ranging from an exceptionally small genome of 66 kb in the parasitic plant

*Viscum scurruloideum* [14] to the largest multi-chromosomal genome of 11.3 Mb in *Silene conica* [15]. Even if two species are evolutionarily close, their genome sizes may vary considerably. The mitogenome sizes of plants in the subfamily Papilionoideae range from 271 kb in *Medicago truncatula* [16] to 588 kb in *Vicia faba* [17], while the mitogenomes of most papilionoid legumes are approximately 400 kb in length [18]. This wide variation in mitogenome size can be attributed to the proliferation of repetitive sequences and the acquisition of foreign DNA from other organisms during evolution [19,20].

Previous studies have documented that the mitogenomes of seed plants are enriched with repetitive sequences, including simple sequence repeats (SSRs), tandem repeats, and dispersed repeats. The SSRs in plant mitogenomes are commonly used as molecular markers for studying genetic diversity and identifying species [21]. The tandem repeats occur in a broad range of plant mitogenomes, which can also serve as molecular markers for unravelling population processes in plants [22]. Large dispersed repeats are the main causes of genome rearrangements, which may generate multipartite structures [13,23–25].

Although the mitogenome sizes of seed plants are variable, the functional genes of NADH dehydrogenase, ubiquinol cytochrome c reductase, ATP synthase, and cytochrome c biogenesis are quite conservative, except for succinate dehydrogenase genes and ribosomal proteins. Many primordial mt genes have been lost during evolution, which has been found to be closely related to their specific functions. For example, *sdh3* and *sdh4* were lost in all gramineous mitogenomes, the *rps11* gene was lost in the differentiation of gymnosperms and angiosperms [26], and the *cox2* gene was lost in the differentiation of the Phaseoleae and Glycininae [18]. Strikingly, nearly all of the universally present NADH dehydrogenase genes were lost from the mitogenome of *Viscum scurruloideum*, with the loss closely associated with its parasitic lifestyle [14].

The Fabaceae, commonly known as legumes, is an economically and ecologically important family of flowering plants ranging from small annual herbs to giant trees, most of which are herbaceous perennials. This family is the third-largest angiosperm family after the Asteraceae and Orchidaceae [27,28], consisting of about 770 genera and more than 20,000 species. A recent study by the Legume Phylogeny Working Group (LPWG) reclassified the three widely-accepted Fabaceae subfamilies (Caesalpinioideae, Mimosoideae, and Papilionoideae) into six new subfamilies (Cercidoideae, Detarioideae, Duparquetioideae, Dialioideae, Caesalpinioideae, and Papilionoideae) based on a taxonomically-comprehensive phylogeny [28]. However, due to the complexity of plant mitogenomes, only 27 mitogenomes of Fabaceae species have been assembled and deposited in the NCBI Nucleotide database (14 May 2020), including 19 species in the Papilionoideae, six species in the Caesalpinioideae, one species of *Cercis canadensis* in the Cercidoideae, and one species of *Tamarindus indica* in the Detarioideae.

In this study, we assembled the complete mitogenome of the common bean *Phaseolus vulgaris*, an herbaceous annual plant grown worldwide for its edible dry seeds or unripe fruit. The common bean is one of the most important grain legumes for human consumption and plays an important role in sustainable agriculture due to its ability to fix atmospheric nitrogen [29]. We analyzed its gene content, repetitive sequences, RNA editing sites, selective pressure, and phylogenetic position, then made comparisons with other plant mitogenomes. The complete mitogenome of *P. vulgaris* will provide important information for the investigation of mitogenomic evolution among the Fabaceae family and aid the functional study of fabaceous mitogenomes. Mitochondrial biogenesis is very important in plant breeding and knowledge of the complete mitogenome provides an opportunity to conduct further important genomic breeding studies in the common bean.

## 2. Materials and Methods

### 2.1. Plant Materials, DNA Extraction, and Sequencing

An inbred landrace of *P. vulgaris* accession G19833 derived from the Andean pool (Race Peru) was selected for sequencing. Seeds were obtained from the germplasm bank of the Embrapa Arroz e Feijão,

Brazil [30]. Root tips obtained from germinated seeds were pre-treated with 2 mM 8-hydroxyquinoline for 18 h at 10 °C, fixed in ethanol-acetic acid (3:1 *v/v*), and stored in fixative at –20 °C for up to several weeks. Total genomic DNA was extracted from root tips using DNAeasy Plant Mini Kits (Qiagen). To construct the shotgun library, DNA was fragmented by nebulization. The raw reads were sequenced with a combination of Roche/454 GS FLX sequencing reads, Illumina HiSeq-2500 sequencing short reads (primarily to correct 454 sequencing errors) and PacBio RS II sequencing long reads (primarily to validate the assembly of the master conformation). The raw reads of *P. vulgaris* used in this study were available in the NCBI Sequence Read Archive (SRA) under accessions SRR069592, SRR5628227, and SRR2912756.

## 2.2. Mitogenome Assembly and Annotation

An efficient procedure for plant mitochondrial genome assembly using whole-genome data from the 454 GS FLX sequencing platform has been applied in many plants, such as *Boea hygrometrica* [31], *Daucus carota* [32], *Gossypium raimondii* [26], and *Salix suchowensis* [33]. Briefly, as shown in Figure S1, we first assembled all the Roche/454 GS FLX sequencing reads using Newbler (version 3.0) [34] with the following parameters: -cpu 20, -het, -sio, -m, -urt, -large, and -s 100. Then, we used custom Perl scripts to construct a draft assembly graph from the file “454AllContigGraph.txt” generated from Newbler. As shown in Figure 1, we obtained six contigs to construct the completed draft mitochondrial graph for assembling the *P. vulgaris* mitogenome. Among the six selected contigs, two (Contig15 and Contig40) were assembled into the mitogenome twice, while the others were assembled only once. To assemble the master conformation (MC), we mapped the PacBio sequencing reads to the mt contigs that spanned repetitive contigs using BLASTN to obtain a major contig relationship map for the repeat regions [35,36].

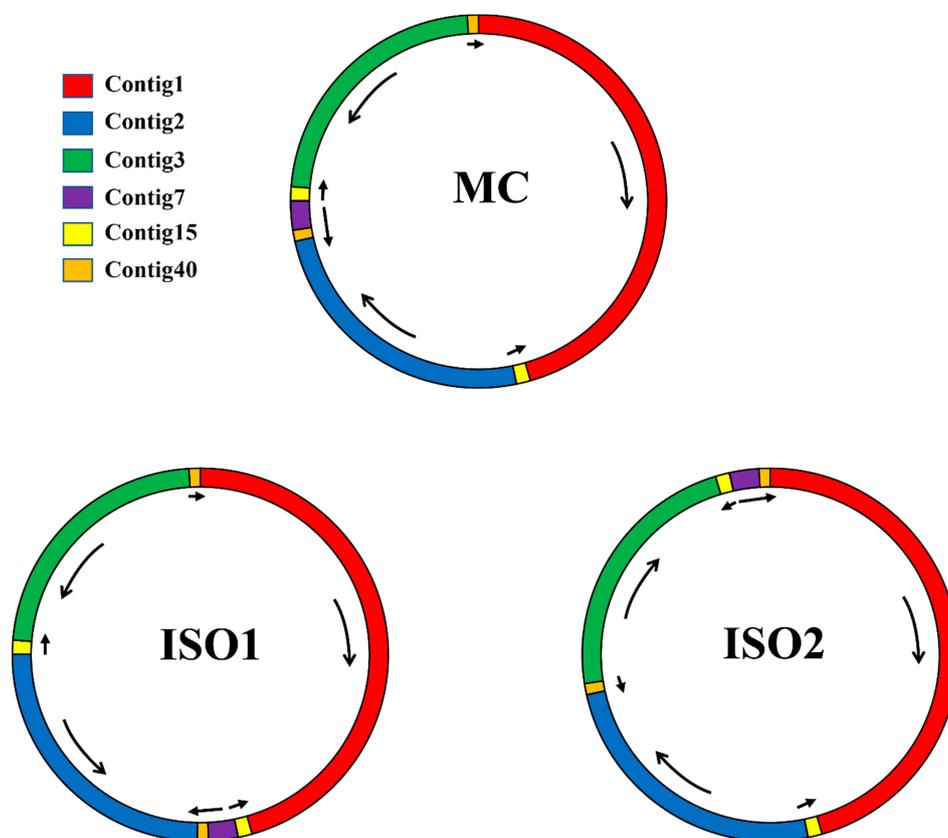
Specifically, for each repeat pair (Contig15 and Contig40), we built four reference sequences according to Dong et al. [37], each with 200 bp up- and down-stream of the two template sequences (original sequences). Then, we searched the PacBio long reads against the database built up from the reference sequences and extracted the matching reads with a blast identity above 80%, an e-value cut-off of  $1e^{-100}$ , and a hit length of over 3000 bp. Next, we mapped the best-matched reads to the four reference sequences in MacVector v17.0.7. As shown in Figure 1, we obtained one master genome and two isomeric genomes (ISO) based on the number of PacBio reads that were mapped to both end contigs of the repetitive contigs (Table S1). We then mapped Illumina sequencing reads to the draft MC mitogenome with BWA [38] and SAMtools [39] softwares to correct the homopolymer length errors (especially in A/T enriched regions) from 454 GS FLX Titanium [26]. Finally, the complete mitogenome sequence of *P. vulgaris* was obtained.

The mitogenome was annotated using the public MITOFY analysis web server (<http://dogma.cccb.utexas.edu/mitofy/>) [8]. The putative genes were manually checked and adjusted by comparing them with other legume mitogenomes in MacVector v.17.07. All transfer RNA genes were confirmed by using tRNAscan-SE with default settings [40]. The start and stop codons of PCGs were manually adjusted to fit open reading frames. The relative synonymous codon usage (RSCU) values and amino acid composition of PCGs were calculated by MEGA X [41]. The OrganellarGenomeDRAW (OGDRAW) program was used to visualize the circular map of the *P. vulgaris* mitogenome [42].

## 2.3. Selective Pressure Analysis

To reflect the selective pressure of PCGs, we calculated the nonsynonymous ( $K_a$ ) and synonymous ( $K_s$ ) substitution rates of each PCG between *P. vulgaris* and other higher plants. *Arabidopsis thaliana* (*A. thaliana*; Brassicaceae) is a popular model organism in plant biology and genetics. *Citrullus lanatus* (*C. lanatus*; Cucurbitaceae) and *Vitis vinifera* (*V. vinifera*; Vitaceae) are highly cultivated fruits worldwide and belong to the Rosids clade, like leguminous plants. Therefore, we selected the mitogenomes of *A. thaliana*, *V. vinifera*, and *C. lanatus* as references to infer the direction and magnitude of natural selection acting on PCGs during the evolution of *P. vulgaris*. The orthologous gene pairs from

*P. vulgaris*, *A. thaliana*, *V. vinifera*, and *C. lanatus* were aligned and formatted by ParaAT2.0 with default parameters [43]. The  $K_a$ ,  $K_s$ , and  $K_a/K_s$  values were calculated using KaKs\_Calculator v.2.0 based on the YN method, and Fisher's exact test was performed to justify the validity of the  $K_a$  and  $K_s$  values [44,45].



**Figure 1.** The master genome and two isomeric genomes observed from *P. vulgaris* mitogenome mediated by two pairs of large repeats (Contig15 and Contig40). The mt contigs were generated and selected from Newbler assembly software. MC and ISO mean the master and isomeric conformations, respectively. Arrows denote the sequence orientation of assembled contigs.

#### 2.4. Prediction of RNA Editing Sites

The online PREP-Mt (predictive RNA editors for plants) suite of servers (<http://prep.unl.edu/>) was used to predict the possible RNA editing sites in the PCGs of *P. vulgaris* and the other four leguminous mitogenomes (*G. max*, *L. japonicus*, *V. radiata*, and *M. pinnata*). In order to predict more true RNA editing sites, the cut-off for prediction score was set as  $C = 0.2$ , which has been proven to be a slight optimum [46]. A low cut-off value will predict more true edit sites but will also increase the probability of misidentifying an unedited site as an edited one.

#### 2.5. Identification of Repeat Sequences in *P. vulgaris* Mitogenome

Three kinds of repeats—SSRs, tandem repeats and dispersed repeats—were detected in the *P. vulgaris* mitogenome. The SSRs were detected using the web-based microsatellite identification tool MISA-web (<https://webblast.ipk-gatersleben.de/misa/>) [47] according to the methods of previous studies [9,26] with a motif size of one to six nucleotides and thresholds of eight, four, four, three, three, and three, respectively. Tandem repeats were identified using the online tool Tandem Repeats Finder 4.09 with default settings (<http://tandem.bu.edu/trf/trf.html>) [48]. AB-BLAST 3.0 was utilized to identify and locate dispersed repeats with the following parameters:  $M = 1$ ,  $N = -3$ ,  $Q = 3$ ,  $R = 3$ ,  $kap$ ,  $span$ ,  $B = 1 \times 10^9$  and  $W = 7$ , which have been proven as effective in many studies [8,9,49]. The BLAST hits

with e-values < 1 and identities > 80% were considered as disperse repeats. The number of dispersed repeats was calculated for seven size intervals (30–49, 50–69, 70–99, 100–149, 150–199, 200–999 and ≥1000 bp). Partly or wholly overlapping repeats were considered as a single repeat unit.

## 2.6. Phylogenetic Analyses

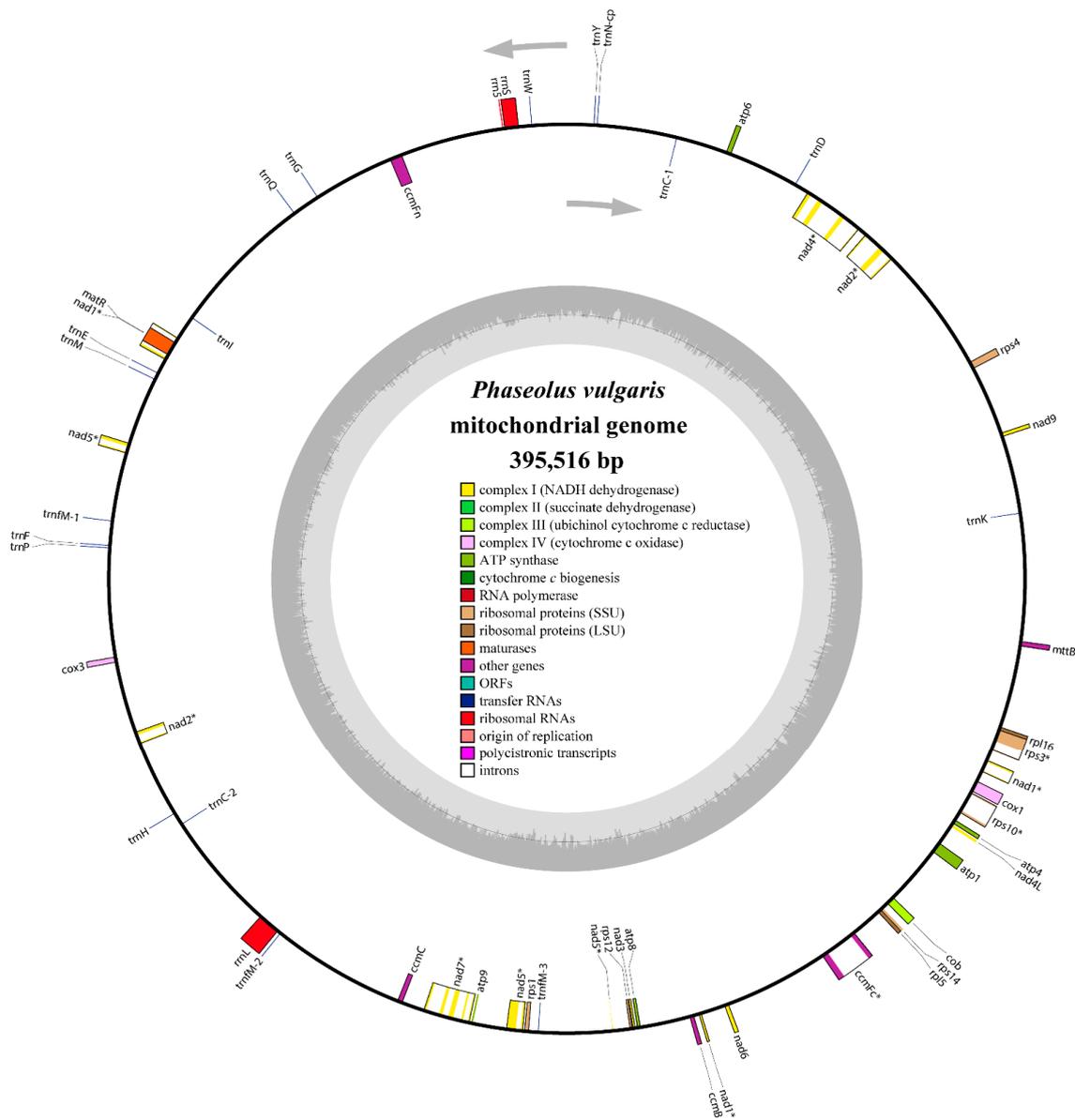
In order to accurately infer the phylogenetic relationships of *P. vulgaris* within the Fabaceae family, maximum likelihood (ML) analysis was performed based on the conserved mitochondrial PCGs (amino acid sequences) of 23 higher plants. The NCBI accession numbers and abbreviations of all these observed mitogenomes are listed in Table S2. Apart from the 19 representative Fabaceae species, taxon sampling also included two species of Solanales (*C. annuum* and *N. tabacum*) and two species of Malpighiales (*P. tremula* and *S. suchowensis*) as outgroups. The single-copy orthologous PCGs common among the 23 analysed species were selected with local Perl scripts. All conserved mitochondrial PCGs were extracted from each mitogenome. The conserved gene sequences from the mitogenome were concatenated into a single dataset and aligned using Muscle software with default settings [50]. Poorly-aligned sequences were deleted or manually adjusted for each alignment. Prior to constructing the phylogenetic tree, we applied MEGA X to determine the most appropriate amino acid substitution model [41]. Based on the model selection results, the ML tree based on a JTT + F model with a gamma distribution was constructed using MEGA X. The bootstrap index value (%) in which the associated taxa clustered together was shown next to the branches and was calculated from 1000 replications.

## 3. Results and Discussion

### 3.1. Genomic Features of the *P. vulgaris* Mitogenome

The complete genomic sequence of the *P. vulgaris* mitogenome was submitted to the NCBI Genome Database (<https://www.ncbi.nlm.nih.gov/genome/browse/>) under accession number NC\_045135.1. The mitogenome was assembled into a typical circular molecule 395,516 bp in length (Figure 2), similar to the mitochondria of some papilionoid legumes such as *G. max* (402,558 bp), *G. soja* (402,545 bp), *L. japonicus* (380,861 bp), *V. angularis* (404,466 bp), and *V. radiata* var. *radiata* (401,262 bp; Table S2). In fact, the mitogenome sizes vary considerably among the papilionoid legumes, ranging from 271,618 bp in *Medicago truncatula* to 588,000 bp in *Vicia faba*. Mitogenome sizes can vary greatly in different cultivars of the same species. For example, the mitogenome size of *G. max* Aiganhuang (N21249) is 402,558 bp, whereas that of *G. max* cultivar Zhonghuang 13 is 513,779 bp [51].

The nucleotide composition of the whole mitogenome is A: 27.37%, C: 22.40%, G: 22.71%, and T: 27.52% (Table 1). The overall GC content is 45.11%, which is consistent with other leguminous plants (*G. max*: 45.03%, *V. faba*: 45.04%, and *V. radiata* var. *radiata*: 45.11%). Strikingly, the GC content of the PCGs is very small compared to those of other regions. As shown in Table 2, a total of 49 unique genes were detected in the *P. vulgaris* mitogenome, comprising 31 PCGs, 15 tRNA genes and 3 rRNA genes. However, none of the genes encodes subunits of Complex II (succinate dehydrogenase), which has also been lost in some other leguminous plants. Additionally, two tRNA genes located in repeat sequences were found to contain two or three copies (*trnC-GCA* and *trnfM-CAU*). The total lengths of the PCGs and *cis*-spliced introns comprise 7.26% and 8.24% of the whole mitogenome, while tRNA and rRNA genes only comprise 0.34% and 1.33% of the mitogenome, respectively. Most PCGs have no introns; however, eight genes (Table 2; *nad1*, *nad2*, *nad4*, *nad5*, *nad7*, *ccmFC*, *rps3*, and *rps10*) were found to contain one or more introns. Three genes (*nad1*, *nad2*, and *nad5*) required trans-splicing to assembly fully-translatable mRNA (Figure 2).



**Figure 2.** Circular map of the *P. vulgaris* mitogenome. Genes shown on the outside of the circle are transcribed clockwise, whereas genes on the inside are transcribed counterclockwise. GC content is represented on the inner circle by the dark gray plot. The asterisks besides genes denote intron-containing genes.

**Table 1.** Genomic features of *P. vulgaris* mitogenome.

Feature	A %	C %	G %	T %	GC %	Size (bp)	Proportion in Genome (%)
Whole genome	27.37	22.40	22.71	27.52	45.11	395,516	100
Protein-coding genes <sup>a</sup>	26.34	21.29	21.62	30.75	42.91	28,725	7.26
<i>cis</i> -spliced introns <sup>a</sup>	24.49	25.56	24.79	25.16	50.34	32,584	8.24
tRNA genes <sup>a</sup>	23.07	24.34	27.49	25.09	51.84	1335	0.34
rRNA genes <sup>a</sup>	26.14	22.62	29.04	22.20	51.66	5252	1.33
Non-coding regions	27.78	22.18	22.48	27.56	44.65	327,620	82.83

<sup>a</sup> Protein-coding genes, *cis*-spliced introns, tRNAs, and rRNAs belong to coding regions.

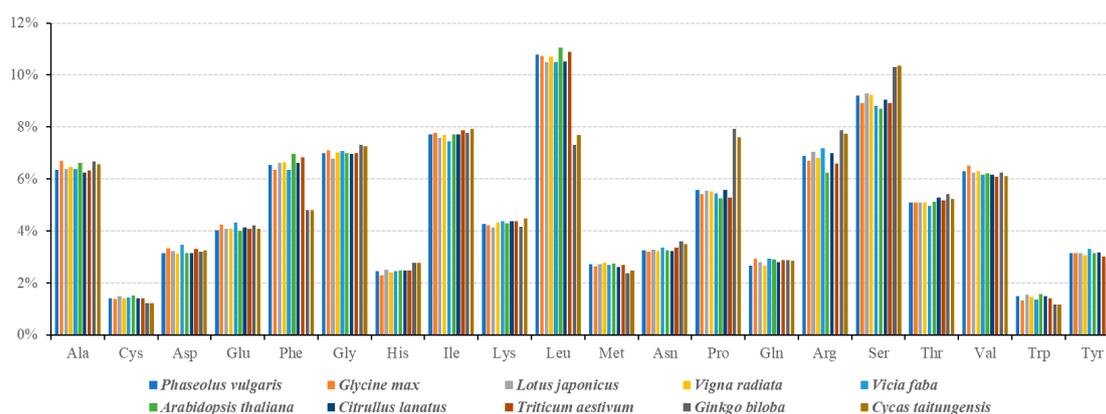
**Table 2.** Gene content of *P. vulgaris* mitogenome.

Group of Genes	Gene Name
Complex I (NADH dehydrogenase)	<i>nad1</i> *, <i>nad2</i> *, <i>nad3</i> , <i>nad4</i> *, <i>nad4L</i> , <i>nad5</i> *, <i>nad6</i> , <i>nad7</i> *, <i>nad9</i>
Complex II (succinate dehydrogenase)	-
Complex III (ubiquinol cytochrome c reductase)	<i>cob</i>
Complex IV (cytochrome c oxidase)	<i>cox1</i> , <i>cox3</i>
Complex V (ATP synthase)	<i>atp1</i> , <i>atp4</i> , <i>atp6</i> , <i>atp8</i> , <i>atp9</i>
Cytochrome c biogenesis	<i>ccmB</i> , <i>ccmC</i> , <i>ccmF<sub>C</sub></i> *, <i>ccmF<sub>N</sub></i>
Ribosomal proteins (SSU)	<i>rps1</i> , <i>rps3</i> *, <i>rps4</i> , <i>rps10</i> *, <i>rps12</i> , <i>rps14</i>
Ribosomal proteins (LSU)	<i>rpl5</i> , <i>rpl16</i>
Maturases	<i>matR</i>
Transport membrane protein	<i>mttB</i>
Ribosomal RNAs	<i>rrn5</i> , <i>rrnS</i> , <i>rrnL</i>
Transfer RNAs	<i>trnC-GCA</i> (2 copies), <i>trnD-GUC</i> , <i>trnE-UUC</i> , <i>trnF-GAA</i> , <i>trnG-GCC</i> , <i>trnM-CAU</i> (3 copies), <i>trnH-GUG</i> , <i>trnI-CAU</i> , <i>trnK-UUU</i> , <i>trnM-CAU</i> , <i>trnN-GUU</i> , <i>trnP-UGG</i> , <i>trnQ-UUG</i> , <i>trnW-CCA</i> , <i>trnY-GUA</i>

\* The asterisks besides genes denotes intron-containing genes.

### 3.2. Codon Usage Analysis of PCGs

In the *P. vulgaris* mitogenome, most of the PCGs use ATG as the start codon, while *mttB* and *nad1* start with ACG (C to U RNA editing on the second site is presumed) as the start codon (Table 3). Four types of stop codons were found in the PCGs: (1) TAA (15 genes; *atp4*, *atp8*, *atp9*, *cox1*, *nad1*, *nad2*, *nad3*, *nad4L*, *nad5*, *nad6*, *nad9*, *rpl5*, *rpl16*, *rps1*, and *rps4*), (2) TGA (10 genes; *atp1*, *ccmB*, *ccmC*, *ccmF<sub>N</sub>*, *cox3*, *matR*, *mttB*, *nad4*, *rps10*, and *rps12*), (3) TAG (5 genes; *atp6*, *cob*, *nad7*, *rps3*, and *rps14*), and (4) CGA (*ccmF<sub>C</sub>*; C to U RNA editing on the first site is presumed). As shown in Figure 3, the codon usage analysis revealed that leucine (Leu) and serine (Ser) are the most frequently-used amino acid residues, while cysteine (Cys) and tryptophan (Trp) are the least-used amino acid residues in the plant mitochondrial proteins. By comparison of the composition of *P. vulgaris* with other angiosperms plants, we found that the distribution of amino acid residues across the mitochondrial proteins are very similar in angiosperms (Figure 3). In addition, most of the amino acid residues were found to be very conserved between angiosperms (*P. vulgaris*, *G. max*, *L. japonicus*, *V. radiata*, *V. faba*, *A. thaliana*, *C. lanatus*, and *T. aestivum*) and gymnosperms (*Ginkgo biloba* and *Cycas taitungensis*), except for five of them (Phe, Leu, Pro, Arg, and Ser).



**Figure 3.** Codon usage pattern of *P. vulgaris* mitogenome compared with nine other higher plants. The proportion of each amino acid residues to the whole mitochondrial proteins is shown on the Y-axis. *Ginkgo biloba* and *Cycas taitungensis* are gymnosperms, while others are angiosperms.

The relative synonymous codon usage (RSCU) analysis for the *P. vulgaris* mitogenome is shown in Figure 4, which indicates that all codons are present in the PCGs. Excluding the termination codons, the 31 PCGs in the *P. vulgaris* mitogenome consist of 9545 codons in total. Additionally, the codon usage

showed that the RSCU values of the NNT and NNA codons are higher than 1.0 except for Ile (AUA) and Leu (CUA; Figure 4), suggesting a strong As or Ts bias in the third codon position of *P. vulgaris* mitochondrial PCGs, which is a very common phenomenon observed in all studied mitogenomes (Table S3). The codon usage pattern of *P. vulgaris* mitogenome is highly consistent with two other papilionoid legumes. The distributions of some codons encoding Pro (CCU, CCA, and CCG) differ between dicotyledons (*P. vulgaris*, *G. max*, *V. angularis*, *C. lanatus*, and *A. thaliana*) and monocotyledons (*T. aestivum*), and some codons (UCG, AGU, AGC, CCU, CCG, ACG, CGG, and AGA) are distributed differently between angiosperms and gymnosperms.

**Table 3.** Gene profile and organization of PCGs in *P. vulgaris* mitogenome.

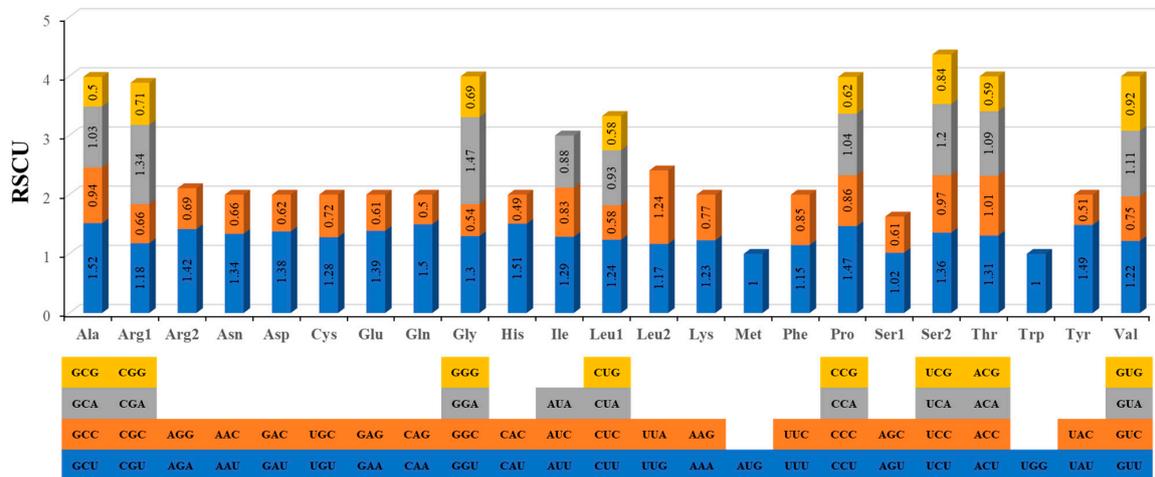
Gene Name	Length	Start Codon	Stop Codon	Direction
<i>atp1</i>	1527	ATG	TGA	F
<i>atp4</i>	588	ATG	TAA	F
<i>atp6</i>	726	ATG	TAG	F
<i>atp8</i>	483	ATG	TAA	R
<i>atp9</i>	225	ATG	TAA	R
<i>ccmB</i>	621	ATG	TGA	F
<i>ccmC</i>	741	ATG	TGA	R
<i>ccmF<sub>C</sub></i>	1329	ATG	CGA	F
<i>ccmF<sub>N</sub></i>	1740	ATG	TGA	R
<i>cob</i>	1176	ATG	TAG	F
<i>cox1</i>	1584	ATG	TAA	F
<i>cox3</i>	798	ATG	TGA	F
<i>matR</i>	2010	ATG	TGA	F
<i>mttB</i>	723	ACG	TGA	F
<i>nad1</i>	978	ACG	TAA	F
<i>nad2</i>	1467	ATG	TAA	R
<i>nad3</i>	357	ATG	TAA	R
<i>nad4</i>	1488	ATG	TGA	R
<i>nad4L</i>	303	ACG	TAA	F
<i>nad5</i>	2019	ATG	TAA	F/R
<i>nad6</i>	618	ATG	TAA	F
<i>nad7</i>	1185	ATG	TAG	R
<i>nad9</i>	573	ATG	TAA	F
<i>rpl5</i>	558	ATG	TAA	F
<i>rpl16</i>	516	ATG	TAA	F
<i>rps1</i>	618	ATG	TAA	R
<i>rps3</i>	1689	ATG	TAG	F
<i>rps4</i>	1041	ATG	TAA	F
<i>rps10</i>	363	ACG	TGA	F
<i>rps12</i>	378	ATG	TGA	R
<i>rps14</i>	303	ATG	TAG	F

### 3.3. Selective Pressure Analysis

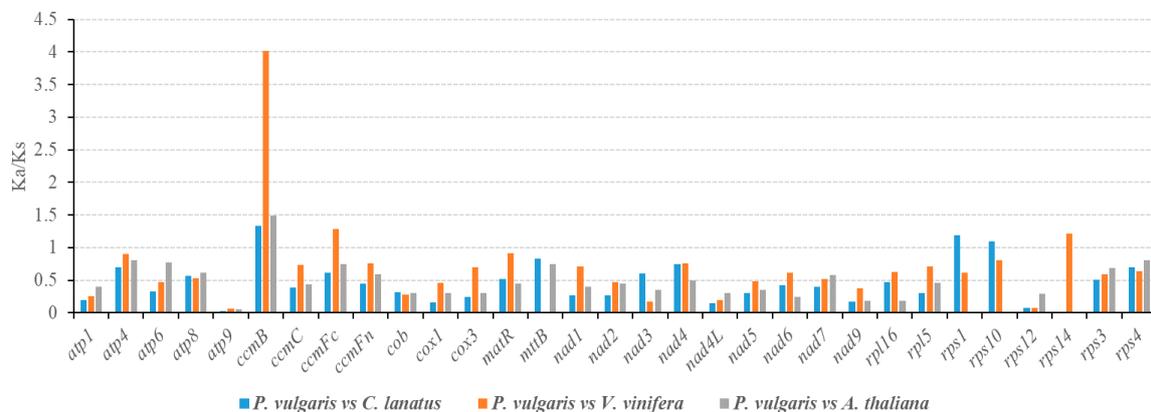
In genetics, the  $K_a/K_s$  ratio is useful for inferring the direction and magnitude of natural selection acting on homologous PCGs across diverged species. The ratio is a more powerful test of the neutral model of evolution than many others available in population genetics as it requires fewer assumptions [52]. A  $K_a/K_s$  ratio  $<1$  implies purifying or stabilizing selection (acting against change), while a ratio of  $>1$  implies positive or Darwinian selection (driving change) and a ratio of exactly 1 indicates neutral selection. Importantly, the  $K_a/K_s$  ratio is unlikely to be significantly above 1 without at least some of the mutations being advantageous.

In this study, the  $K_a/K_s$  ratio was determined for all 31 PCGs following comparison of the *P. vulgaris* mitogenome with those of *C. lanatus*, *V. vinifera* and *A. thaliana* (Figure 5). Nearly all of the  $K_a/K_s$  ratios were  $<1.0$ , suggesting that most of the PCGs were under stabilizing selection during evolution.

Combining the information in Figure 5 and Table 1, the  $K_a/K_s$  ratios of all Complex I–V genes were  $<1$ , indicating that these genes were highly conserved in the evolutionary process of higher plants. The large number of mitochondrial genes under stabilizing selection ( $K_a/K_s < 1$ ) may play important roles in stabilizing the normal functioning of mitochondria [53,54].



**Figure 4.** Relative synonymous codon usage (RSCU) of *P. vulgaris* mitogenome. Codon families are on the X-axis. RSCU values are the number of times of a particular codon, relative to the number of times that the codon would be observed for a uniform synonymous codon usage.



**Figure 5.**  $K_a/K_s$  ratios for 31 protein coding genes of *P. vulgaris*, *C. lanatus*, *V. vinifera*, and *A. thaliana*. The blue, orange, and gray boxes indicate  $K_a/K_s$  ratios of *P. vulgaris* vs. *C. lanatus*, *P. vulgaris* vs. *V. vinifera*, and *P. vulgaris* vs. *A. thaliana*.

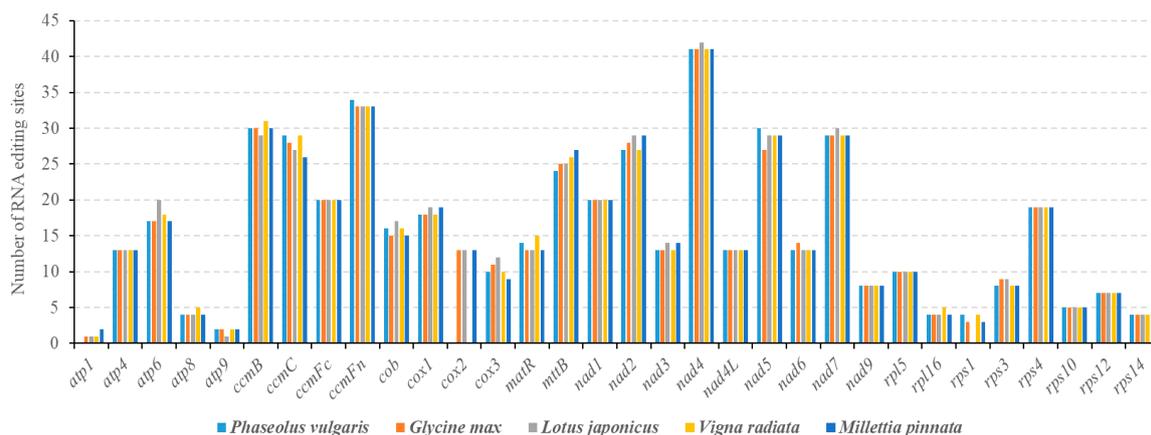
As shown in Figure 5, the  $K_a/K_s$  ratios of *ccmB* were  $>1$  between *P. vulgaris* and all of the three selected species, indicating that *ccmB* may have suffered from positive selection since divergence from their last common ancestor. Particularly, the  $K_a/K_s$  ratio of *ccmB* between *P. vulgaris* and *V. vinifera* was significantly  $>1$  (4.01), suggesting that some advantage occurred during evolution. Additionally, the  $K_a/K_s$  ratios of *ccmF<sub>C</sub>*, *rps1*, *rps10*, and *rps14* were also  $>1$ , indicating that these genes were under positive selection after divergence of the last common ancestor. Since *CcmB* and *ccmF<sub>C</sub>* genes encode for some important components of the *c*-type cytochrome maturation pathway in mitochondria, we speculate that the adaptive evolution of *P. vulgaris* is closely related to the roles of *c*-type cytochromes in respiratory and photosynthetic electron transport [55–57]. Additionally, *rps1*, *rps10*, and *rps14* genes encode small mitoribosomal subunit proteins, which have been reported to play crucial roles in various biological processes in eukaryotic organisms, such as embryogenesis, leaf morphogenesis,

and the formation of reproductive tissues [58–60]. The high  $K_a/K_s$  ratios of *rps* genes observed here may be very important for the evolution of *P. vulgaris*.  $K_a/K_s$  ratios  $>1$  have also been reported for some other mitochondrial genes, including *atp8*, *ccmF<sub>N</sub>*, *matR*, and *mttB* [26,33,61,62], indicating that mitochondrial genes in different plant species may be subjected to diverse selection pressures during evolution. Most importantly, the  $K_a/K_s$  ratio of the orthologous gene-pairs is an average over all sites and, even under positive selection, it can be  $<1$  because some sites might be under positive selection while others are under purifying selection [53,61,63].

### 3.4. Prediction of RNA Editing Sites in PCGs

Many previous studies have documented that RNA editing is one of the necessary steps for gene expression in the mitochondrial and chloroplast genomes of higher plants [64–67]. RNA editing is a post-transcriptional modification that converts specific cytidines (C) to uridines (U) and uridines to uridines in the transcripts of nearly all mitochondrial PCGs. Based on the web-based PREP-mt program, we predicted a total of 486 RNA editing sites in 31 PCGs and 100% C-to-U RNA editing. Among the 486 RNA editing sites, 34.57% (168 sites) were predicted at the first base position of the codon and 65.43% (318 sites) were found in the second position, while none were found in the third position. The lack of predicted RNA editing sites in the silent position is probably due to the limitation of the PREP-Mt predictive methodology rather than there being no RNA editing in this position. Since most of the RNA editing sites in third codon positions did not change the amino acid encoded by the codon, the tie-breaking rules used by PREP-Mt could not select the edited state [68]. Therefore, RNA editing in the silent editing position needs to be further identified by experimental methods.

The occurrence of RNA editing can cause alteration of initiation and termination codons in PCGs, and the frequency of their generation is much higher than that of their removal. As shown in Table 3, *mttB*, *nad1*, *nad4L*, and *rps10* genes use ACG as their initiation codons, which may be altered to the normal AUG by RNA editing modification. Additionally, the *ccmF<sub>C</sub>* gene uses CGA as its termination codon, which may be altered to UGA by RNA editing modification. As shown in Figure 6, the number of RNA editing sites in different genes varies greatly, and the Complex I (NADH dehydrogenase) and Cytochrome *c* biogenesis genes (*ccmB*, *ccmC*, *ccmF<sub>C</sub>*, and *ccmF<sub>N</sub>*) encode the most predicted RNA editing sites. Based on a comparison of the predicted RNA editing sites in five leguminous plants, the *nad4* gene encodes the most RNA editing sites, while *atp1* encodes the fewest (Figure 6).



**Figure 6.** Predicted RNA editing sites of the *P. vulgaris* mitogenome compared with four other leguminous plants. RNA-editing sites are predicted on PREP-Mt sites (<http://prep.unl.edu/>) with the cut-off value of 0.2.

Previous studies have shown that the frequency and type of RNA editing in each organelle is highly lineage-specific [26,65,69,70]. As shown in Figure 6, the number of predicted RNA editing sites in different papilionoid legume mitogenomes is very conserved, ranging from 486 sites in *P. vulgaris* to

503 sites in *Lotus japonicus*, suggesting that they share extremely conserved PCGs. In angiospermous mitogenomes, nearly all of the RNA editing sites are C to U, and the number of RNA editing sites is concentrated between 400 to 500. For example, 463 and 444 RNA-editing sites were found in the *C. lanatus* and *C. pepo* mitogenomes, of which 394 are shared [8]; 441 and 427 RNA-editing sites were found in *A. thaliana* and *B. napus* mitogenomes, of which 358 are shared [10]. In the gymnosperm *Cycas taitungensis*, 1084 RNA editing sites were found in its mitogenome [71]. The clearly descending number of RNA editing sites is in accordance with gene losses from gymnosperms to angiosperms. In contrast to angiosperms and gymnosperms, both types of C-to-U and U-to-C conversions are found in the mitochondrial transcripts of ferns and hornworts [69,72].

### 3.5. Analysis of Repeat Sequences in the *P. vulgaris* Mitogenome

The vast majority of the variance in genome size of plant mitogenomes can be explained by differences in the sizes of repeat sequences, which are composed of SSRs, tandem repeats and dispersed repeats. Plant mitogenomes, particularly those of angiosperms, were already well known for its sizeable fractions of repetitive sequences even before any complete mitogenomes were available. SSRs are DNA tracts of tandem-repeated motifs of one to six bases that are useful molecular markers in studying genetic diversity and species identification [21]. In this study, a total of 314 perfect SSRs were identified in the *P. vulgaris* mitogenome, including 139 mono-, 140 di-, 5 tri-, 22 tetra-, 3 penta-, and 5 hexa-nucleotide repeats (Table 4). The mononucleotide repeats of A/T (129 repeats) were found to be more prevalent than other repeat types. The dinucleotide repeats, TA/AT, are the second most numerous (50 repeats), while tri-, tetra-, penta-, and hexa-nucleotide repeats are fewer in number and only observed in intronic or intergenic regions. As shown in Table 5, seven tandem repeats with lengths ranging from 13 bp to 57 bp were also detected in the *P. vulgaris* mitogenome. Among these seven tandem repeats, only one is localized in a coding region (*rnlL*), while the others are all found in intergenic spacers.

Besides SSRs and tandem repeats, 143 dispersed repeats with lengths > 30 bp (total length: 35,000 bp; 8.85% of the genome) were also identified in the *P. vulgaris* mitogenome (Figure 7; Table S4). Most of the repeats (77 repeats, 53.85%) are 30 bp to 59 bp long, and 25 repeats are longer than 100 bp, with only two longer than 1 kb (R1: 4866 bp; R2: 3529 bp). Previous studies have documented the importance of large repeats (>1 kb) in genomic structural changes, and pairwise direct and inverted large repeats may produce two small subgenomic conformations or isomeric conformations, respectively. As shown in Figure 1, the largest repeat was assembled as Contig15 and the second largest was assembled as Contig40, both of which were inverted repeats. By aligning the PacBio long reads to both ends of the two large repeats, we constructed the master circle and two isomeric molecules (Figure 1). Repeats are commonly found in plant mitogenomes but are poorly conserved across species, even within the same family. As shown in Figure 7 and Table S4, the total number of repeats ranges from 59 in *V. angularis* to 215 in *M. pinnata*, and the total length of repeats ranges from 9224 bp (2.28% of the whole genome) in *V. angularis* to 411,265 bp (69.94% of the whole genome) in *V. faba*. Mitogenome enlargement in *V. faba* is mainly caused by the expansion of repeated sequences. Thirteen large (>1 kb) repeats covered 398.8 kb or 68% of the whole mitogenome size [17]. However, when all but single copies of the large repeat sequences were excluded, the *V. faba* mitogenome size is 388.6 kb, which is similar to other Papilionoideae mitogenomes [18]. The extremely complex repeat patterns should be responsible for the various genome sizes of the plant mitogenome. However, genome size is by no means only determined by the size of repeats. The mitochondrial genome of *Vitis vinifera* has only 7% repeats despite a genome size of nearly 773 kb [73], while the moderately-sized (404.5 kb) *V. angularis* genome has fewer and smaller repeats than those found in the much smaller genomes of *Brassica napus* (222 kb) and *Silene latifolia* (253 kb; Table S2) [10,74].

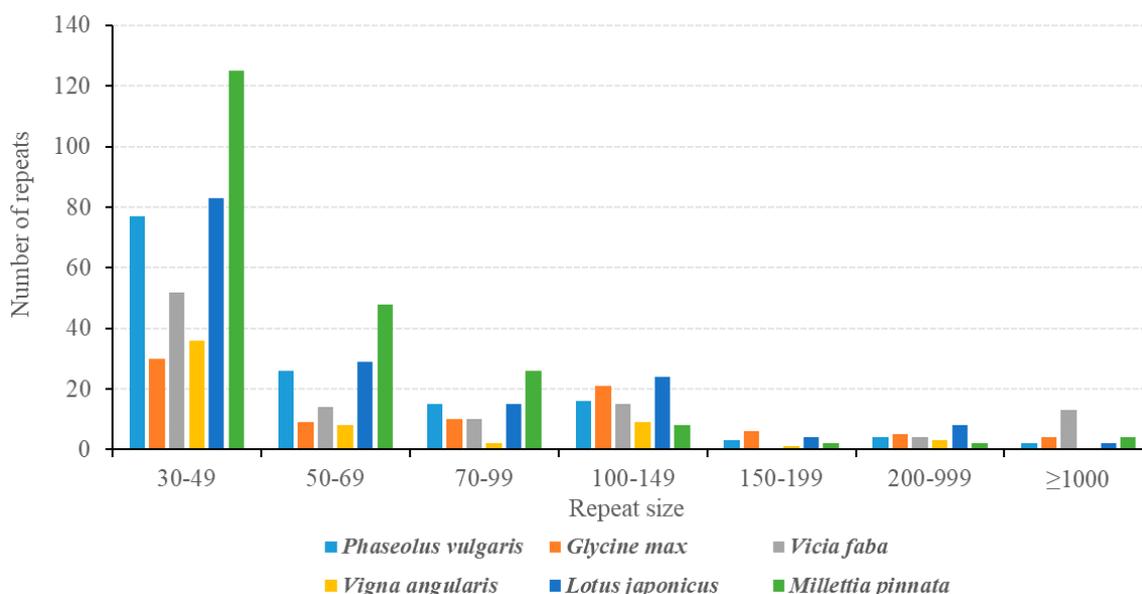
**Table 4.** Frequency of identified SSR motifs in *P. vulgaris* mitogenome.

Motif Type	Number of Repeats											Total	Proportion (%)
	3	4	5	6	7	8	9	10	11	12	13		
Monomer	-	-	-	-	-	91	32	10	4	1	1	139	44.27
Dimer	-	120	16	4	0	0	0	0	0	0	0	140	44.59
Trimer	-	5	0	0	0	0	0	0	0	0	0	5	1.59
Tetramer	21	1	0	0	0	0	0	0	0	0	0	22	7.01
Pentamer	3	0	0	0	0	0	0	0	0	0	0	3	0.96
Hexamer	5	0	0	0	0	0	0	0	0	0	0	5	1.59
Total	29	126	16	4	0	91	32	10	4	1	1	314	100

**Table 5.** Distribution of tandem repeats in *P. vulgaris* mitogenome.

No	Size (bp)	Start	End	Repeat (bp) × Copy Number	Location
1	57	75,208	75,322	(TTGGATCAAAACGATGTTCAACAACCTTTGCCCGCTCTGTTTCTTGGAGGAAAATAG) × 2	IGS ( <i>trnD</i> , <i>atp6</i> )
2	27	92,189	92,243	(AGAGCAGGTCGGTCTAGGTAGTTGAAA) × 2	IGS ( <i>trnC</i> , <i>trnN</i> )
3	38	108,846	108,922	(AAAAATATACATAACATATCCCAAACCTATAGAGATA) × 2	IGS ( <i>rrn5</i> , <i>ccmF<sub>N</sub></i> )
4	13	231,727	231,753	(TCTTAAGTAAAGT) × 2	IGS ( <i>nad2</i> -exon1, <i>trnH</i> )
5	18	252,332	252,368	(CATAGTCGCGAGCTGTT) × 2	<i>rrnL</i>
6	15	315,763	315,793	(GTATAGTATAGTAGG) × 2	IGS ( <i>nad1</i> -exon1, <i>nad6</i> )
7	33	356,560	356,626	(CCTTGCCCCCTGCAGAGCCTCAAGCCCCCTGAGC) × 2	IGS ( <i>atp1</i> , <i>nad4L</i> )

IGS: intergenic pacers.



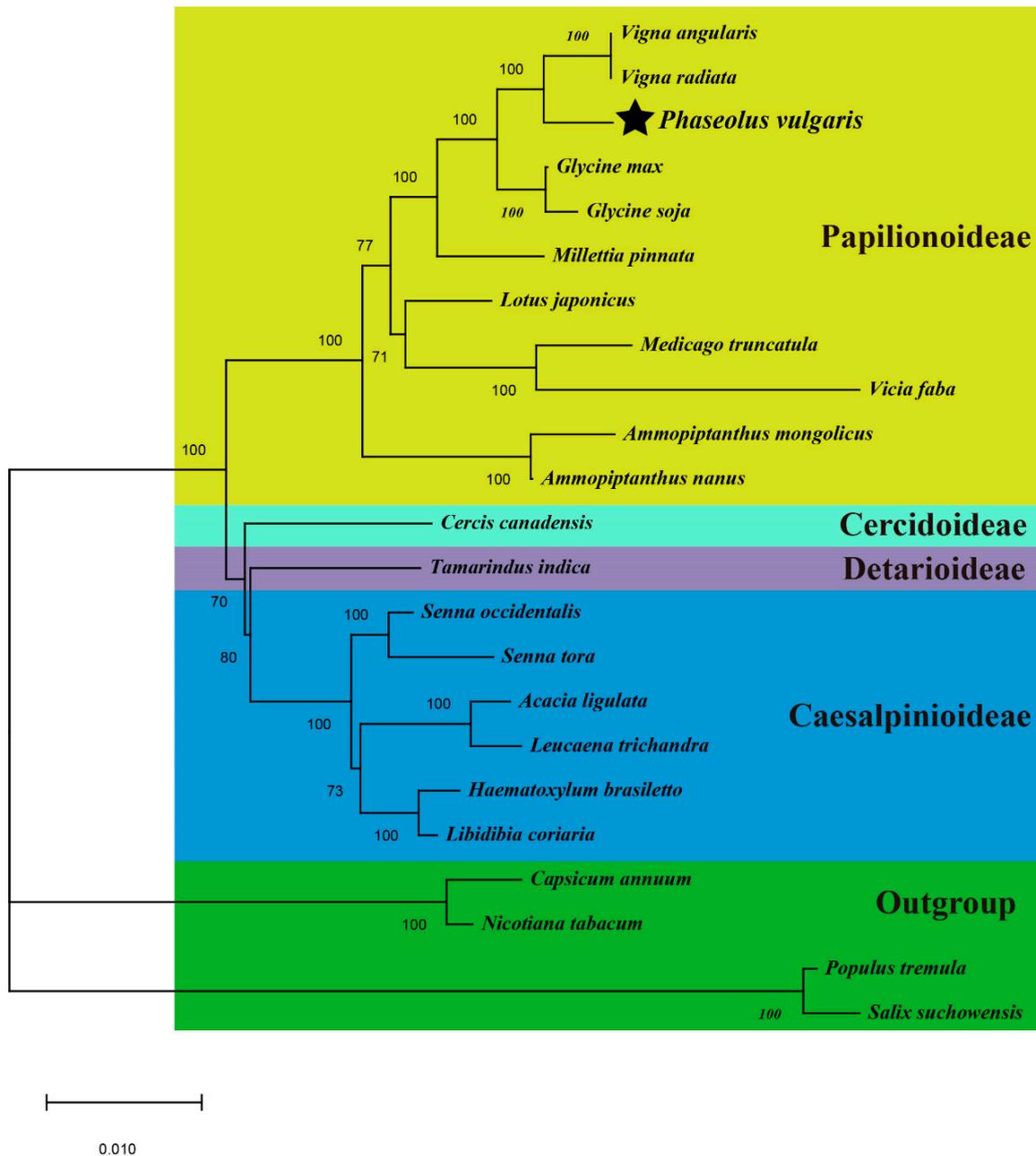
**Figure 7.** Frequency distribution of dispersed repeat in the *P. vulgaris* mitogenome compared with five other leguminous plants. The number of dispersed repeats in *Phaseolus vulgaris*, *Glycine max*, *Vicia faba*, *Vigna faba*, *Vigna angularis*, *Lotus japonicus*, and *Millettia pinnata* mitogenomes are shown by blue, orange, gray, yellow, blue, and green, respectively.

### 3.6. Phylogenetic Analyses and Multiple Losses of PCGs during Evolution

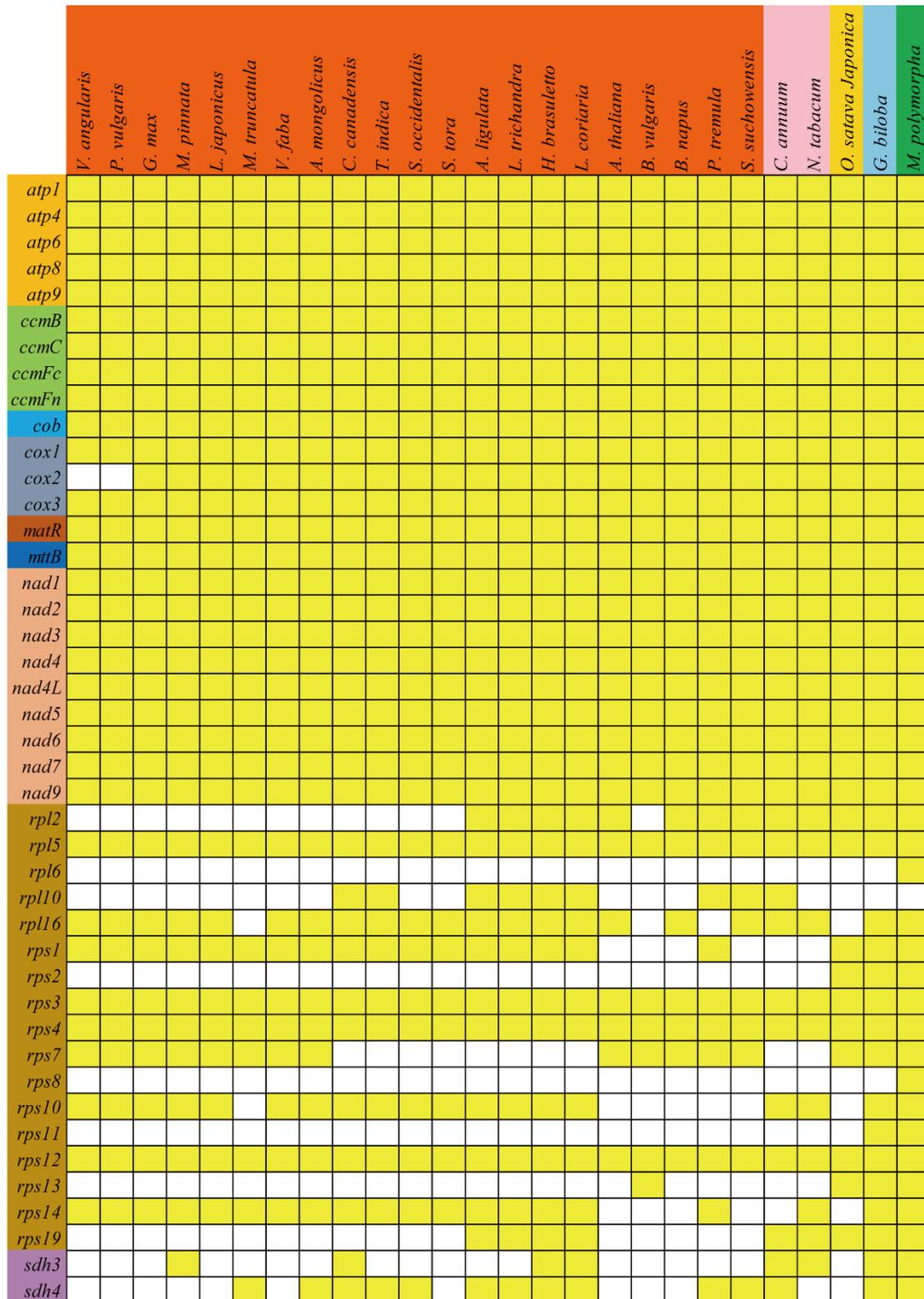
With rapid developments in sequencing technology and assembly methods, an increasing number of complete plant mitogenomes has been assembled, providing an important opportunity for phylogenetic analyses using mitogenomes. In this study, to determine the phylogenetic position of *P. vulgaris*, we downloaded 23 plant mitogenomes from the GenBank database (<https://www.ncbi.nlm.nih.gov/genome/browse/>), including 19 species of Fabales, two species of Solanales, and two species of Malpighiales. A set of 26 conserved single-copy orthologous genes (*atp1*, *atp4*, *atp6*, *atp8*, *atp9*, *ccmB*, *ccmC*, *ccmF<sub>C</sub>*, *ccmF<sub>N</sub>*, *cob*, *cox1*, *cox3*, *matR*, *nad1*, *nad2*, *nad3*, *nad4*, *nad4L*, *nad5*, *nad6*, *nad7*, *nad9*, *rps3*, *rps4*, and *rps12*) present in all of the 23 analyzed mitogenomes was used to construct the phylogenetic tree, and species from the Solanales and Malpighiales were designated as the outgroup. As shown in Figure 8, the bootstrap values of each node are all over 70% supported and 15 nodes are supported 100%. The ML phylogenetic tree strongly supports that *P. vulgaris* is evolutionarily close to the clade formed by two *Vigna* species. The tree also strongly supports the separation of Fabales from the clade composed of Solanales and Malpighiales (100% bootstrap value), as well as the separation of Papilionoideae from the clade composed of Cercidoideae, Detarioideae, and Caesalpinioideae (100%). The bootstrap value for the separation of Detarioideae and Caesalpinioideae is 80%, and the value for the separation of Cercidoideae from the clade composed of Detarioideae and Caesalpinioideae is 70%.

As described by Richardson et al. [75], the mitochondrial genomes of higher plants vary significantly in genome size, gene content and order. Losses of PCGs occurred frequently during the evolution of higher plants. The phylogenetic tree provides a backdrop for the further analysis of gene loss during evolution, and the gene contents of all observed species are summarized in Figure 9. Most of the PCGs were conserved in different plant mitogenomes, especially for the genes in the groups of Complex I, Complex III, Complex V, cytochrome *c* biogenesis, maturases, and transport membrane protein [13]. The conservation of these genes suggests that they play crucial roles in the function of mitochondria. However, the ribosomal proteins and succinate dehydrogenase genes were highly variable. As shown in Figure 9, the *cox2* gene was only lost in the subfamily Phaseolinae (*V. angularis*, *V. radiata*, and *P. vulgaris*) but retained in other leguminous plants, suggesting that this gene was lost after separation from the subfamily Glycininae. The *rpl2* gene was lost in most leguminous

plants but regained in *A. ligulate*, *L. trichandra*, *H. brasiletto*, and *L. coriaria*, suggesting that this gene was lost before the emergence of Fabaceae but could be regained in some leguminous plants. Similar phenomena were found in many ribosomal proteins (*rpl10*, *rpl16*, *rps7*, *rps10*, and *rps19*). Additionally, *rpl6* and *rps8* genes were lost from liverworts (*M. polymorpha*) during evolution [76], the *rps11* gene was lost from gymnosperms (*G. biloba*) and liverworts during the divergence of the angiosperms and gymnosperms [77], and the *rpl10* gene was lost in monocots and gymnosperms but regained in dicots [33,78]. The enhanced loss of ribosomal proteins in plant mitogenomes indicates that these genes were encoded partly by mitochondrial native genes and partly by nuclear genes, due to the gene transfer between mitochondria and nucleus [79–81].



**Figure 8.** Maximum likelihood phylogenies of *P. vulgaris* within Fabaceae. Relationships were inferred employing 26 conserved PCGs of 23 plant mitogenomes. Numbers on each node are bootstrap support values. NCBI accession numbers are listed in Table S2. Scale indicates number of nucleotide substitutions per site.



**Figure 9.** Distribution of PCGs in plant mitogenomes. White boxes indicate that the gene is not present in the mitogenome. The colors of genes indicate their corresponding categories. The colors of species represent the classes of rosids (orange), asterids (pink), monocotyledons (gold), gymnosperms (light blue), and liverworts (green).

#### 4. Conclusions

In this study, we first assembled and characterized the complete mitogenome of *P. vulgaris*. By aligning the PacBio sequencing reads to the draft mitogenome, one master circle and two isomeric molecules were assembled based on two large repeats. Selective-pressure analysis of PCGs indicates that *ccmB*, *ccmFC*, *rps1*, *rps10*, and *rps14* genes with  $K_a/K_s$  ratios  $> 1$  might play important roles during evolution, whereas all Complex I–V genes with  $K_a/K_s$  ratios  $< 1$  were highly conserved in the evolutionary process of higher plants. The C-to-U conversions may generate initiation, termination, or internal codons with completely unpredictable functions. The prediction of RNA editing sites in *P. vulgaris* mt PCGs will provide important clues for the investigation of gene functions with novel codons. The comparison of genomic features in all sequenced leguminous plants should contribute to a comprehensive understanding of the evolutionary process of legumes. The sequencing of the *P. vulgaris* mitogenome not only provides an important opportunity to conduct further genomic breeding studies in the common bean, it also provides valuable information for future evolutionary and molecular studies of leguminous plants.

**Supplementary Materials:** The following are available online at <http://www.mdpi.com/1422-0067/21/11/3778/s1>, Figure S1: The flowchart of *P. vulgaris* mitogenome assembly and annotation, Table S1: PacBio reads of spanning repetitive contigs, Table S2: The abbreviations and genome sizes of studied mitogenomes, Table S3: The relative synonymous codon usage (RSCU) of mitogenome genome across seven higher plants, Table S4: Comparison of dispersed repeats within six mitogenomes of Papilionaceae subfamily.

**Author Contributions:** Conceptualization, C.H.; Formal analysis, C.B. and N.L.; Funding acquisition, Z.L.; Methodology, C.B.; Project administration, Z.L.; C.B., N.L., and Y.X.; Writing—original draft, C.B.; Writing—review and editing, C.H. and Z.L. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was funded by the National Science and Technology Major Project of China (6307030004) and Jiangsu Basic Research Program (Natural Science Foundation) Project (no. BK20191393).

**Conflicts of Interest:** The authors declare no conflict of interest.

#### References

1. Greiner, S.; Bock, R. Tuning a menage a trois: Co-evolution and co-adaptation of nuclear and organellar genomes in plants. *Bioessays* **2013**, *35*, 354–365. [\[CrossRef\]](#)
2. Timmis, J.N.; Ayliffe, M.A.; Huang, C.Y.; Martin, W. Endosymbiotic gene transfer: Organelle genomes forge eukaryotic chromosomes. *Nat. Rev. Genet.* **2004**, *5*, 123–135. [\[CrossRef\]](#)
3. Hsu, C.L.; Mullin, B.C. Physical characterization of mitochondrial DNA from cotton. *Plant Mol. Biol.* **1989**, *13*, 467–468. [\[CrossRef\]](#)
4. Kubo, T.; Mikami, T. Organization and variation of angiosperm mitochondrial genome. *Physiol. Plant.* **2007**, *129*, 6–13. [\[CrossRef\]](#)
5. Palmer, J.D.; Herbon, L.A. Plant mitochondrial DNA evolves rapidly in structure, but slowly in sequence. *J. Mol. Evol.* **1988**, *28*, 87–97. [\[CrossRef\]](#)
6. Gray, M.W.; Burger, G.; Lang, B.F. Mitochondrial Evolution. *Science* **1999**, *283*, 1476–1481. [\[CrossRef\]](#)
7. Lang, B.F.; Gray, M.W.; Burger, G. Mitochondrial Genome Evolution and the Origin of Eukaryotes. *Annu. Rev. Genet.* **1999**, *33*, 351–397. [\[CrossRef\]](#)
8. Alverson, A.J.; Wei, X.X.; Rice, D.W.; Stern, D.B.; Barry, K.; Palmer, J.D. Insights into the Evolution of Mitochondrial Genome Size from Complete Sequences of *Citrullus lanatus* and *Cucurbita pepo* (Cucurbitaceae). *Mol. Biol. Evol.* **2010**, *27*, 1436–1448. [\[CrossRef\]](#)
9. Alverson, A.J.; Zhuo, S.; Rice, D.W.; Sloan, D.B.; Palmer, J.D. The mitochondrial genome of the legume *Vigna radiata* and the analysis of recombination across short mitochondrial repeats. *PLoS ONE* **2011**, *6*, e16404. [\[CrossRef\]](#)
10. Handa, H. The complete nucleotide sequence and RNA editing content of the mitochondrial genome of rapeseed (*Brassica napus* L.): Comparative analysis of the mitochondrial genomes of rapeseed and *Arabidopsis thaliana*. *Nucleic Acids Res.* **2003**, *31*, 5907–5916. [\[CrossRef\]](#)

11. Mulligan, R.M.; Chang, K.L.; Chou, C.C. Computational analysis of RNA editing sites in plant mitochondrial genomes reveals similar information content and a sporadic distribution of editing sites. *Mol. Biol. Evol.* **2007**, *24*, 1971–1981. [[CrossRef](#)]
12. Andre, C.; Levy, A.; Walbot, V. Small repeated sequences and the structure of plant mitochondrial genomes. *Trends Genet.* **1992**, *8*, 128–132. [[CrossRef](#)]
13. Chang, S.; Wang, Y.; Lu, J.; Gai, J.; Li, J.; Chu, P.; Guan, R.; Zhao, T. Correction: The Mitochondrial Genome of Soybean Reveals Complex Genome Structures and Gene Evolution at Intercellular and Phylogenetic Levels. *PLoS ONE* **2013**, *8*, e56502. [[CrossRef](#)]
14. Skippington, E.; Barkman, T.J.; Rice, D.W.; Palmer, J.D. Miniaturized mitogenome of the parasitic plant *Viscum scurruloideum* is extremely divergent and dynamic and has lost all nad genes. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, E3515–E3524. [[CrossRef](#)] [[PubMed](#)]
15. Sloan, D.B.; Alverson, A.J.; Chuckalovcak, J.P.; Wu, M.; McCauley, D.E.; Palmer, J.D.; Taylor, D.R. Rapid evolution of enormous, multichromosomal genomes in flowering plant mitochondria with exceptionally high mutation rates. *PLoS Biol.* **2012**, *10*, e1001241. [[CrossRef](#)] [[PubMed](#)]
16. Bi, C.; Wang, X.; Xu, Y.; Wei, S.; Shi, Y.; Dai, X.; Yin, T.; Ye, N. The complete mitochondrial genome of *Medicago truncatula*. *Mitochondrial DNA Part B* **2016**, *1*, 122–123. [[CrossRef](#)]
17. Negruk, V. Mitochondrial genome sequence of the legume *Vicia faba*. *Front. Plant Sci.* **2013**, *4*, 128. [[CrossRef](#)]
18. Choi, I.-S.; Schwarz, E.N.; Ruhlman, T.A.; Khiyami, M.A.; Sabir, J.S.; Hajarrah, N.H.; Sabir, M.J.; Rabah, S.O.; Jansen, R.K. Fluctuations in Fabaceae mitochondrial genome size and content are both ancient and recent. *BMC Plant Biol.* **2019**, *19*, 448. [[CrossRef](#)]
19. Wynn, E.L.; Christensen, A.C. Repeats of Unusual Size in Plant Mitochondrial Genomes: Identification, Incidence and Evolution. *G3 Genes Genomes Genet.* **2019**, *9*, 549–559. [[CrossRef](#)]
20. Bergthorsson, U.; Adams, K.L.; Thomason, B.; Palmer, J.D. Widespread horizontal transfer of mitochondrial genes in flowering plants. *Nature* **2003**, *424*, 197–201. [[CrossRef](#)]
21. 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. [[CrossRef](#)]
22. Sperisen, C.; Büchler, U.; Gugerli, F.; Mátyás, G.; Geburek, T.; Vendramin, G. Tandem repeats in plant mitochondrial genomes: Application to the analysis of population differentiation in the conifer Norway spruce. *Mol. Ecol.* **2001**, *10*, 257–263. [[CrossRef](#)]
23. Alverson, A.J.; Rice, D.W.; Dickinson, S.; Barry, K.; Palmer, J.D. Origins and Recombination of the Bacterial-Sized Multichromosomal Mitochondrial Genome of Cucumber. *Plant Cell* **2011**, *23*, 2499–2513. [[CrossRef](#)]
24. Backert, S.; Nielsen, B.L.; Börner, T. The mystery of the rings: Structure and replication of mitochondrial genomes from higher plants. *Trends Plant Sci.* **1997**, *2*, 477–483. [[CrossRef](#)]
25. Ogihara, Y.; Yamazaki, Y.; Murai, K.; Kanno, A.; Terachi, T.; Shiina, T.; Miyashita, N.; Nasuda, S.; Nakamura, C.; Mori, N. Structural dynamics of cereal mitochondrial genomes as revealed by complete nucleotide sequencing of the wheat mitochondrial genome. *Nucleic Acids Res.* **2005**, *33*, 6235–6250. [[CrossRef](#)]
26. Bi, C.; Paterson, A.H.; Wang, X.; Xu, Y.; Wu, D.; Qu, Y.; Jiang, A.; Ye, Q.; Ye, N. Analysis of the complete mitochondrial genome sequence of the diploid cotton *Gossypium raimondii* by comparative genomics approaches. *BioMed. Res. Int.* **2016**, *2016*, 5040598. [[CrossRef](#)]
27. Christenhusz, M.J.; Byng, J.W. The number of known plants species in the world and its annual increase. *Phytotaxa* **2016**, *261*, 201–217. [[CrossRef](#)]
28. Azani, N.; Babineau, M.; Bailey, C.D.; Banks, H.; Barbosa, A.R.; Pinto, R.B.; Boatwright, J.S.; Borges, L.M.; Brown, G.K.; Bruneau, A.; et al. A new subfamily classification of the Leguminosae based on a taxonomically comprehensive phylogeny: The Legume Phylogeny Working Group (LPWG). *Taxon* **2017**, *66*, 44–77. [[CrossRef](#)]
29. Schmutz, J.; McClean, P.E.; Mamidi, S.; Wu, G.A.; Cannon, S.B.; Grimwood, J.; Jenkins, J.; Shu, S.; Song, Q.; Chavarro, C.; et al. A reference genome for common bean and genome-wide analysis of dual domestications. *Nat. Genet.* **2014**, *46*, 707. [[CrossRef](#)]
30. Fonsêca, A.; Ferreira, J.; dos Santos, T.R.B.; Mosiolek, M.; Bellucci, E.; Kami, J.; Gepts, P.; Geffroy, V.; Schweizer, D.; dos Santos, K.G. Cytogenetic map of common bean (*Phaseolus vulgaris* L.). *Chrom. Res.* **2010**, *18*, 487–502. [[CrossRef](#)]

31. Zhang, T.; Fang, Y.; Wang, X.; Deng, X.; Zhang, X.; Hu, S.; Yu, J. The Complete Chloroplast and Mitochondrial Genome Sequences of *Boea hygrometrica*: Insights into the Evolution of Plant Organellar Genomes. *PLoS ONE* **2012**, *7*, e30531. [[CrossRef](#)]
32. Iorizzo, M.; Senalik, D.; Szklarczyk, M.; Grzebelus, D.; Spooner, D.; Simon, P. *De novo* assembly of the carrot mitochondrial genome using next generation sequencing of whole genomic DNA provides first evidence of DNA transfer into an angiosperm plastid genome. *BMC Plant Biol.* **2012**, *12*, 61. [[CrossRef](#)]
33. Ye, N.; Wang, X.; Li, J.; Bi, C.; Xu, Y.; Wu, D.; Ye, Q. Assembly and comparative analysis of complete mitochondrial genome sequence of an economic plant *Salix suchowensis*. *PeerJ* **2017**, *5*, e3148. [[CrossRef](#)]
34. Nederbragt, A.J. On the middle ground between open source and commercial software—the case of the Newbler program. *Genome Biol.* **2014**, *15*, 113. [[CrossRef](#)]
35. Camacho, C.; Coulouris, G.; Avagyan, V.; Ma, N.; Papadopoulos, J.; Bealer, K.; Madden, T.L. BLAST+: Architecture and applications. *BMC Bioinform.* **2009**, *10*, 421. [[CrossRef](#)]
36. Zhang, T.; Zhang, X.; Hu, S.; Yu, J. An efficient procedure for plant organellar genome assembly, based on whole genome data from the 454 GS FLX sequencing platform. *Plant Methods* **2011**, *7*, 38. [[CrossRef](#)]
37. Dong, S.; Zhao, C.; Chen, F.; Liu, Y.; Zhang, S.; Wu, H.; Zhang, L.; Liu, Y. The complete mitochondrial genome of the early flowering plant *Nymphaea colorata* is highly repetitive with low recombination. *BMC Genom.* **2018**, *19*, 614. [[CrossRef](#)]
38. Li, H.; Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **2009**, *25*, 1754–1760. [[CrossRef](#)]
39. Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **2009**, *25*, 2078–2079. [[CrossRef](#)]
40. Schattner, P.; Brooks, A.N.; Lowe, T.M. The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. *Nucleic Acids Res.* **2005**, *33*, W686–W689. [[CrossRef](#)]
41. Kumar, S.; Stecher, G.; Li, M.; Niyaz, C.; Tamura, K. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol. Biol. Evol.* **2018**, *35*, 1547–1549. [[CrossRef](#)]
42. Greiner, S.; Lehwark, P.; Bock, R. OrganellarGenomeDRAW (OGDRAW) version 1.3. 1: Expanded toolkit for the graphical visualization of organellar genomes. *Nucleic Acids Res.* **2019**, *47*, W59–W64. [[CrossRef](#)]
43. Zhang, Z.; Xiao, J.; Wu, J.; Zhang, H.; Liu, G.; Wang, X.; Dai, L. ParaAT: A parallel tool for constructing multiple protein-coding DNA alignments. *Biochem. Biophys. Res. Commun.* **2012**, *419*, 779–781. [[CrossRef](#)]
44. Yang, Z.; Nielsen, R. Estimating Synonymous and Nonsynonymous Substitution Rates Under Realistic Evolutionary Models. *Mol. Biol. Evol.* **2000**, *17*, 32–43. [[CrossRef](#)]
45. Wang, D.; Zhang, Y.; Zhang, Z.; Zhu, J.; Yu, J. KaKs\_Calculator 2.0: A Toolkit Incorporating Gamma-Series Methods and Sliding Window Strategies. *Genom. Proteom. Bioinform.* **2010**, *8*, 77–80. [[CrossRef](#)]
46. Mower, J.P. The PREP suite: Predictive RNA editors for plant mitochondrial genes, chloroplast genes and user-defined alignments. *Nucleic Acids Res.* **2009**, *37*, W253–W259. [[CrossRef](#)]
47. Beier, S.; Thiel, T.; Münch, T.; Scholz, U.; Mascher, M. MISA-web: A web server for microsatellite prediction. *Bioinformatics* **2017**, *33*, 2583–2585. [[CrossRef](#)]
48. Benson, G. Tandem repeats finder: A program to analyze DNA sequences. *Nucleic Acids Res.* **1999**, *27*, 573–580. [[CrossRef](#)]
49. Liu, G.; Cao, D.; Li, S.; Su, A.; Geng, J.; Grover, C.E.; Hu, S.; Hua, J. The Complete Mitochondrial Genome of *Gossypium hirsutum* and Evolutionary Analysis of Higher Plant Mitochondrial Genomes. *PLoS ONE* **2013**, *8*, e69476. [[CrossRef](#)]
50. Edgar, R.C. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **2004**, *32*, 1792–1797. [[CrossRef](#)]
51. Shen, Y.; Du, H.; Liu, Y.; Ni, L.; Wang, Z.; Liang, C.; Tian, Z. Update soybean Zhonghuang 13 genome to a golden reference. *Sci. China Life Sci.* **2019**, *62*, 1257–1260. [[CrossRef](#)] [[PubMed](#)]
52. Li, W.-H.; Wu, C.-I.; Luo, C.-C. A new method for estimating synonymous and nonsynonymous rates of nucleotide substitution considering the relative likelihood of nucleotide and codon changes. *Mol. Biol. Evol.* **1985**, *2*, 150–174. [[PubMed](#)]
53. Betrán, E.; Bai, Y.; Motiwale, M. Fast Protein Evolution and Germ Line Expression of a *Drosophila* Parental Gene and Its Young Retroposed Paralog. *Mol. Biol. Evol.* **2006**, *23*, 2191–2202. [[CrossRef](#)] [[PubMed](#)]
54. Arbiza, L.; Dopazo, J.; Dopazo, H. Positive Selection, Relaxation, and Acceleration in the Evolution of the Human and Chimp Genome. *PLoS Comp. Biol.* **2006**, *2*, e38. [[CrossRef](#)] [[PubMed](#)]

55. Meyer, E.H.; Giegé, P.; Gelhaye, E.; Rayapuram, N.; Ahuja, U.; Thöny-Meyer, L.; Grienenberger, J.-M.; Bonnard, G. AtCCMH, an essential component of the c-type cytochrome maturation pathway in *Arabidopsis* mitochondria, interacts with apocytochrome c. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 16113–16118. [[CrossRef](#)] [[PubMed](#)]
56. Faivre-Nitschke, S.E.; Nazoa, P.; Gualberto, J.M.; Grienenberger, J.M.; Bonnard, G. Wheat mitochondria *ccmB* encodes the membrane domain of a putative ABC transporter involved in cytochrome c biogenesis. *Biochim. Biophys. Acta* **2001**, *1519*, 199–208. [[CrossRef](#)]
57. Sanders, C.; Turkarlan, S.; Lee, D.-W.; Daldal, F. Cytochrome c biogenesis: The Ccm system. *Trends Microbiol.* **2010**, *18*, 266–274. [[CrossRef](#)]
58. Robles, P.; Quesada, V. Emerging Roles of Mitochondrial Ribosomal Proteins in Plant Development. *Int. J. Mol. Sci.* **2017**, *18*, 2595. [[CrossRef](#)]
59. Mauro, V.P.; Edelman, G.M. The Ribosome Filter Redux. *Cell Cycle* **2007**, *6*, 2246–2251. [[CrossRef](#)]
60. Schippers, J.H.M.; Mueller-Roeber, B. Ribosomal composition and control of leaf development. *Plant Sci.* **2010**, *179*, 307–315. [[CrossRef](#)]
61. Cui, P.; Liu, H.; Lin, Q.; Ding, F.; Zhuo, G.; Hu, S.; Liu, D.; Yang, W.; Zhan, K.; Zhang, A.; et al. A complete mitochondrial genome of wheat (*Triticum aestivum* cv. Chinese Yumai), and fast evolving mitochondrial genes in higher plants. *J. Genet.* **2009**, *88*, 299–307. [[CrossRef](#)] [[PubMed](#)]
62. Feng, L.; Li, N.; Yang, W.; Li, Y.; Wang, C.-M.; Tong, S.-W.; He, J.-X. Analyses of mitochondrial genomes of the genus *Ammopiptanthus* provide new insights into the evolution of legume plants. *Plant Syst. Evol.* **2019**, *305*, 385–399. [[CrossRef](#)]
63. Wernegreen, J.J.; Riley, M.A. Comparison of the evolutionary dynamics of symbiotic and housekeeping loci: A case for the genetic coherence of rhizobial lineages. *Mol. Biol. Evol.* **1999**, *16*, 98–113. [[CrossRef](#)] [[PubMed](#)]
64. Bock, R.; Khan, M.S. Taming plastids for a green future. *Trends Biotechnol.* **2004**, *22*, 311–318. [[CrossRef](#)]
65. Chen, H.; Deng, L.; Jiang, Y.; Lu, P.; Yu, J. RNA Editing Sites Exist in Protein-coding Genes in the Chloroplast Genome of *Cycas taitungensis*. *J. Integr. Plant Biol.* **2011**, *53*, 961–970. [[CrossRef](#)]
66. Raman, G.; Park, S. Analysis of the Complete Chloroplast Genome of a Medicinal Plant, *Dianthus superbus* var. *longicalycinus*, from a Comparative Genomics Perspective. *PLoS ONE* **2015**, *10*, e0141329. [[CrossRef](#)]
67. Wakasugi, T.; Hirose, T.; Horihata, M.; Tsudzuki, T.; Kössel, H.; Sugiura, M. Creation of a novel protein-coding region at the RNA level in black pine chloroplasts: The pattern of RNA editing in the gymnosperm chloroplast is different from that in angiosperms. *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 8766–8770. [[CrossRef](#)]
68. Mower, J.P. PREP-Mt: Predictive RNA editor for plant mitochondrial genes. *BMC Bioinform.* **2005**, *6*, 96. [[CrossRef](#)] [[PubMed](#)]
69. Malek, O.; Lättig, K.; Hiesel, R.; Brennicke, A.; Knoop, V. RNA editing in bryophytes and a molecular phylogeny of land plants. *EMBO J.* **1996**, *15*, 1403–1411. [[CrossRef](#)]
70. Steinhauser, S.; Beckert, S.; Capesius, I.; Malek, O.; Knoop, V. Plant Mitochondrial RNA Editing. *J. Mol. Evol.* **1999**, *48*, 303–312. [[CrossRef](#)]
71. Shu-Miaw, C.; Arthur, C.C.S.; Wang, D.; Yu-Wei, W.; Shu-Mei, L.; The-Yuan, C. The Mitochondrial Genome of the Gymnosperm *Cycas taitungensis* Contains a Novel Family of Short Interspersed Elements, Bpu Sequences, and Abundant RNA Editing Sites. *Mol. Biol. Evol.* **2008**, *25*, 603–615.
72. Hiesel, R.; Combettes, B.; Brennicke, A. Evidence for RNA editing in mitochondria of all major groups of land plants except the Bryophyta. *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 629–633. [[CrossRef](#)] [[PubMed](#)]
73. Goremykin, V.V.; Salamini, F.; Velasco, R.; Viola, R. Mitochondrial DNA of *Vitis vinifera* and the Issue of Rampant Horizontal Gene Transfer. *Mol. Biol. Evol.* **2009**, *26*, 99–110. [[CrossRef](#)] [[PubMed](#)]
74. Sloan, D.B.; Alverson, A.J.; Štorchová, H.; Palmer, J.D.; Taylor, D.R. Extensive loss of translational genes in the structurally dynamic mitochondrial genome of the angiosperm *Silene latifolia*. *BMC Evol. Biol.* **2010**, *10*, 274. [[CrossRef](#)] [[PubMed](#)]
75. Richardson, A.O.; Rice, D.W.; Young, G.J.; Alverson, A.J.; Palmer, J.D. The “fossilized” mitochondrial genome of *Liriodendron tulipifera*: Ancestral gene content and order, ancestral editing sites, and extraordinarily low mutation rate. *BMC Biol.* **2013**, *11*, 29. [[CrossRef](#)] [[PubMed](#)]
76. Bowman, J.L.; Kohchi, T.; Yamato, K.T.; Jenkins, J.; Shu, S. Insights into Land Plant Evolution Garnered from the *Marchantia polymorpha* Genome. *Cell* **2017**, *171*, 287–304. [[CrossRef](#)]

77. Guo, W.; Felix, G.; Fan, W.; Young, G.J.; Volker, K.; Palmer, J.D.; Mower, J.P. *Ginkgo* and *Welwitschia* Mitogenomes Reveal Extreme Contrasts in Gymnosperm Mitochondrial Evolution. *Mol. Biol. Evol.* **2016**, *33*, 1448–1460. [[CrossRef](#)]
78. Notsu, Y.; Masood, S.; Nishikawa, T.; Kubo, N.; Akiduki, G.; Nakazono, M.; Hirai, A.; Kadowaki, K. The complete sequence of the rice (*Oryza sativa* L.) mitochondrial genome: Frequent DNA sequence acquisition and loss during the evolution of flowering plants. *Mol. Genet. Genom.* **2002**, *268*, 434–445. [[CrossRef](#)]
79. Clifton, S.W.; Minx, P.; Fauron, C.M.-R.; Gibson, M.; Allen, J.O.; Sun, H.; Thompson, M.; Barbazuk, W.B.; Kanuganti, S.; Tayloe, C. Sequence and Comparative Analysis of the Maize NB Mitochondrial Genome. *Plant Physiol.* **2004**, *136*, 3486–3503. [[CrossRef](#)]
80. Unseld, M.; Marienfeld, J.R.; Brandt, P.; Brennicke, A. The mitochondrial genome of *Arabidopsis thaliana* contains 57 genes in 366,924 nucleotides. *Nat. Genet.* **1997**, *15*, 57–61. [[CrossRef](#)]
81. Sugiyama, Y.; Watase, Y.; Nagase, M.; Makita, N.; Yagura, S.; Hirai, A.; Sugiura, M. The complete nucleotide sequence and multipartite organization of the tobacco mitochondrial genome: Comparative analysis of mitochondrial genomes in higher plants. *Mol. Genet. Genom.* **2005**, *272*, 603–615. [[CrossRef](#)]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).