

## Article

# Deciphering the Plastomic Code of Chinese Hog-Peanut (*Amphicarpaea edgeworthii* Benth., Leguminosae): Comparative Genomics and Evolutionary Insights within the Phaseoleae Tribe

Yi-Nan Xiang<sup>1,2</sup>, Xiao-Qun Wang<sup>1</sup>, Lu-Lu Ding<sup>1</sup>, Xin-Yu Bai<sup>1</sup>, Yu-Qing Feng<sup>1</sup>, Zhe-Chen Qi<sup>1,\*</sup> , Yong-Tao Sun<sup>3</sup> and Xiao-Ling Yan<sup>2,\*</sup>

<sup>1</sup> Zhejiang Province Key Laboratory of Plant Secondary Metabolism and Regulation, College of Life Sciences and Medicine, Zhejiang Sci-Tech University, Hangzhou 310018, China; yinanxiang9@outlook.com (Y.-N.X.); llding6@163.com (L.-L.D.); fyq071008@163.com (Y.-Q.F.)

<sup>2</sup> Eastern China Conservation Centre for Wild Endangered Plant Resources, Shanghai Chenshan Botanical Garden, Shanghai 201602, China

<sup>3</sup> East China Survey and Planning Institute, The National Forestry and Grassland Administration, Hangzhou 310019, China; sunyongtao329@sina.com

\* Correspondence: zqi@zstu.edu.cn (Z.-C.Q.); xlyan@cemps.ac.cn (X.-L.Y.)

**Abstract:** The classification and phylogenetic relationships within the Phaseoleae tribe (Leguminosae) have consistently posed challenges to botanists. This study addresses these taxonomic intricacies, with a specific focus on the Glycininae subtribe, by conducting a comprehensive analysis of the highly conserved plastome in *Amphicarpaea edgeworthii* Benth., a critical species within this subtribe. Through meticulous genomic sequencing, we identified a plastome size of 148,650 bp, composed of 128 genes, including 84 protein-coding genes, 36 tRNA genes, and 8 rRNA genes. Comparative genomic analysis across seven Glycininae species illuminated a universally conserved circular and quadripartite structure, with nine genes exhibiting notable nucleotide diversity, signifying a remarkable genomic variability. Phylogenetic reconstruction of 35 Phaseoleae species underscores the affinity of *Amphicarpaea* with *Glycine*, placing *Apios* as a sister lineage to all other Phaseoleae species, excluding Clitorinae and Diocleinae subtribes. Intriguingly, *Apios*, *Butea*, *Erythrina*, and *Spatholobus*, traditionally clumped together in the Erythrinae subtribe, display paraphyletic divergence, thereby contesting their taxonomic coherence. The pronounced structural differences in the quadripartite boundary genes among taxa with unresolved subtribal affiliations demand a reevaluation of Erythrinae's taxonomic classification, potentially refining the phylogenetic contours of the tribe.

**Keywords:** chloroplast genomics; evolutionary biology; glycininae; Leguminosae phylogeny; phylogenetic relationships; plastome characterization; subtribal classification; systematics



**Citation:** Xiang, Y.-N.; Wang, X.-Q.; Ding, L.-L.; Bai, X.-Y.; Feng, Y.-Q.; Qi, Z.-C.; Sun, Y.-T.; Yan, X.-L. Deciphering the Plastomic Code of Chinese Hog-Peanut (*Amphicarpaea edgeworthii* Benth., Leguminosae): Comparative Genomics and Evolutionary Insights within the Phaseoleae Tribe. *Genes* **2024**, *15*, 88. <https://doi.org/10.3390/genes15010088>

Academic Editor: Linley Jesson

Received: 30 November 2023

Revised: 5 January 2024

Accepted: 8 January 2024

Published: 11 January 2024



**Copyright:** © 2024 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 (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Leguminous plants, a large and economically significant family that have a pivotal role in agricultural systems as food, feed, and biofertilizers, present a compelling model for genetic and evolutionary studies [1–3]. The distinct symbiotic nitrogen-fixing capabilities of legumes underscore their ecological significance and offer a window into plant–microbe interactions and biogeochemical processes [4–6]. However, despite the significance of legumes, foundational taxonomic research within the family remains fraught with unresolved issues [7–9].

The Phaseoleae tribe belonging to the subfamily Papilionaceae comprises seven subtribes with eighty-four genera, which are classified under the unranked non-protein amino acid-accumulating clade (NPAAA clade) [10,11]. Yet, its subtribe classification remains elusive, with varying nomenclature across systems. J. Lackey's chemotaxonomic revision following Bentham's schema in the 'Genera Plantarum' presents a dichotomy in the distribution of canavanine among subtribes, proposing a seven-subtribe structure: Cajaninae,

Kennediinae, Diocleinae, Phaseolinae, Ophrestinae, Glycininae, and Erythrinae [12]. The USDA's Germplasm Resources Information Network (GRIN) recognizes the absence of Erythrinae and the addition of Clitoriinae. Within this framework, certain taxa remain unassigned to any subtribe. Flora Reipublicae Popularis Sinicae (FRPS) offers yet another perspective, placing genera, like *Apios*, *Butea*, *Cochlianthus*, *Erythrina*, *Mucuna*, *Spatholobus*, among others, into the subtribe Erythrinae based on morphological characteristics [13], whereas these genera are categorized within undetermined subtribes in the GRIN. This taxonomic ambiguity underscores the imperative for refined investigations to elucidate the phylogenetic intricacies within the Phaseoleae tribe. Similarly uncertain is the tribe's position within the Leguminosae family, notably its alleged association with the Desmodieae and Psoraleae tribes in the Indigoferoid/Millettoid clade [9,14]. These insights suggest that Phaseoleae's classification is intermingled with other tribes, like Millettieae and Abreae, indicating a dispersed phylogenetic identity lacking clear delineation. A deeper understanding of the molecular characteristics of each subtribe will enhance the classification of Phaseoleae and provide more information for phylogenetic reconstruction.

*Amphicarpaea edgeworthii* is an annual widely distributed species that attracts considerable attention due to its three types of flowers (fruits), subterranean cleistogamous, aerial cleistogamous, and aerial chasmogamous, and serves as a model plant for studying complex flowering patterns and reproductive strategy [15–18]. It grows in the pool of the forests and geographic areas of grasslands in mountainous areas. In recent years, there has been an in-depth exploration of the nuclear genome of *A. edgeworthii* [19–21]. However, its chloroplast genome has yet to be fully analyzed, and there are unresolved relationships within the subtribe Glycininae. In a previous phylogenetic study of *A. ferruginea*, the genera *Amphicarpaea* and *Pueraria* were identified as sister taxa, forming a polyphyletic relationship with the genera *Glycine* and *Mucuna*, and *Glycine* and *Spatholobus* were clustered as clades [22]. This contradicts the existing classification system. *Amphicarpaea* and *Glycine* are classified under the Glycininae subtribe, while *Spatholobus* and *Mucuna*, as previously mentioned, are placed outside, and the subtribal allocation of these taxa is still incompletely defined. In a tree constructed using the *matK* gene, *Amphicarpaea*, *Pueraria*, and *Glycine* within the Glycininae subtribe exhibit a polytomy relationship [8]. In another RPS16 intron sequence tree, *Amphicarpaea*, *Glycine*, *Pueraria*, and *Teramnus* form three subclades, while *Amphicarpaea* is sister to a clade consisting of *Glycine* and *Teramnus*, and it is noteworthy that *Pueraria* is not monophyletic internally [23]. The Glycininae subtribe holds substantial economic importance and prominence. Among its most important members is soybean (*Glycine max*), whose seeds are rich in protein and serve as raw materials for various soy products and oil extraction. They can be utilized in the production of health supplements and pharmaceuticals [24,25]. Additionally, these plants exhibit robust nitrogen-fixing capabilities, contributing to soil improvement and promoting sustainable agricultural development [26]. Therefore, analyzing the plastome of this subtribe, determining its phylogenetic relationships with disputed genera, and elucidating its position within the Phaseoleae tribe are meaningful endeavors.

The progress of high-throughput sequencing technologies in the past two decades has enhanced the efficiency and quality of plastid genome sequencing. The plastome is a vital genomic region in plants, characterized by maternal inheritance and circular DNA structure [27,28]. It contains several essential genes involved in critical biological processes, such as photosynthesis, plastid protein import, fatty acid biosynthesis, and proteolysis [29–31]. The plastome has proved to be a valuable tool not only for establishing plant phylogenetic relationships, developing DNA barcodes, and creating molecular markers but also for studying regulatory mechanisms of photosynthesis [32–37]. Currently, the National Center for Biotechnology Information (NCBI) website has published over 1300 complete plastomes within the legume family, spanning more than 350 species. The species with the highest number of sequence records are *Medicago minima*, *Pueraria montana*, and *Trifolium pratense*, all belonging to the Papilionoideae subfamily. These complete chloroplast datasets have been applied to explore phylogenetic relationships at various scales

within the Leguminosae family, including the family level (Leguminosae), subfamily level (Papilionoideae), evolutionary branch level (Millettoid/Phaseoloid clade), and genus and subgenus [14,38–41]. Previous studies did not sample and discuss the tribe and subtribe levels within the unresolved Millettoid/Phaseoloid clade, particularly focusing on the ambiguous boundaries of Phaseoleae [14].

Here, we generated the latest complete plastome of *A. edgeworthii* and conducted the first analysis of its architecture in comparison with six other species within Glycininae. We performed a phylogenetic analysis of a total of thirty-five species within Papilionaceae, including seven subtribes and five genera awaiting subtribal assignment, and discussed the distribution patterns of boundary genes among different subtribes and provided recommendations for the internal classification of Phaseoleae.

## 2. Results and Discussion

### 2.1. The Overall Structure and General Features of the *A. edgeworthii* Chloroplast Genome

The plastome of *A. edgeworthii* has a quadripartite, circular topology with a length of 148,650 base pairs (bp) (Figure 1). The plastome consists of a pair of inverted repeats (IRb and IRa), each 23,504 bp, a Small Single-Copy (SSC) region of 17,967 bp, and a Large Single-Copy (LSC) region of 83,675 bp. A total of 49,351 bp make up the genome's non-coding region, which comprises introns and intergenic spacers, while the remaining 75,693 bp are coding (CDS). The GC content of the LSC and SSC regions is 32.9% and 28.7%, respectively, whereas in the inverted repeats IRa and IRb, it is 42.3% for both. Thus, IRs have a larger proportion of GC than the SSC and LSC regions (Table 1). The GC percentage at the first, second, and third positions in the CDS sequence are 44.43%, 36.74%, and 26.69%, respectively.

The annotation of *A. edgeworthii* plastome revealed a total of 128 genes (84 protein-coding genes, 36 tRNAs genes, and eight rRNAs genes); 86 genes are present in the LSC (66 protein-coding genes and 20 tRNA genes) and 12 genes are present in the SSC (11 protein-coding genes and one tRNA), while the remaining 14 genes (six tRNAs, four rRNAs, and four protein-coding genes) are in the IRa and IRb regions (Table 2).

The result of relatively synonymous codon usage (RSCU) was that 64 codons were employed (Figures 2 and 3). Of these, 31 codons have RSCU values lower than 1, and 31 codons have RSCU values higher than 1. A total of 96.8% of codons with high RSCU values have Cytosine (C) or Guanine (G) endings, and 93.5% of codons with lower RSCU values have Thymine (T) or Adenine (A) endings. This pattern of the third codon usage was also observed in other species of legumes [42,43]. AUG and UGG are codons without bias (i.e., with RSCU values = 1), while the termination codon, UAA, has a value of 1.9125.

**Table 1.** Nucleotide composition in the *A. edgeworthii* plastome.

Region	A (%)	T (%)	G (%)	C (%)	GC (%)	Total (bp)	Proportion in Genome (%)
Genome	32.3	32.3	17.8	17.6	35.4	148,650	100
CDS	31.7	32.3	19.3	16.7	36.0	75,693	50.92
tRNA	34.0	32.6	18.4	15.2	33.5	22,185	14.92
rRNA	21.6	24.9	29.9	23.6	53.6	2543	1.71
Cis-spliced intron	26.2	18.9	31.5	23.5	54.9	9060	6.09
Non-coding region	35.1	34.7	15.1	15.1	30.2	49,351	33.20
LSC	33.6	33.5	16.8	16.0	32.9	83,675	56.29
SSC	35.5	35.8	13.5	15.2	28.7	17,967	12.10
IRA	28.6	29.1	21.9	20.4	42.3	23,504	15.82
IRB	28.6	29.1	21.9	20.4	42.3	23,504	15.82

**Table 2.** Genes present in the plastome of *A. edgeworthii*.

Category	Group of Genes	Name of Genes
RNA genes	Ribosomal RNA genes (rRNA)	<i>rrn5</i> <sup>a</sup> , <i>rrn4.5</i> <sup>a</sup> , <i>rrn16</i> <sup>a</sup> , <i>rrn23</i> <sup>a</sup>
	Transfer RNA genes (tRNA)	<i>trnC-GCA</i> , <i>trnD-GUC</i> , <i>trnE-UUC</i> , <i>trnF-GAA</i> , <i>trnG-CAU</i> , <i>trnG-GCC</i> , <i>trnG-UCC</i> , <i>trnH-GUG</i> , <i>trnI-GAU</i> <sup>a</sup> , <i>trnK-UUU</i> , <i>trnL-CAA</i> <sup>a</sup> , <i>trnL-UAA</i> , <i>trnL-UAG</i> , <i>trnM-CAU</i> , <i>trnN-GUU</i> <sup>a</sup> , <i>trnP-UGG</i> , <i>trnQ-UUG</i> , <i>trnR-ACG</i> <sup>a</sup> , <i>trnR-UCU</i> , <i>trnS-GCU</i> , <i>trnS-GGA</i> , <i>trnS-UGA</i> , <i>trnT-GGU</i> , <i>trnT-UGU</i> , <i>trnV-GAC</i> <sup>a</sup> , <i>trnV-UAC</i> , <i>trnW-CCA</i> , <i>trnY-GUA</i> , <i>trnA-UGC</i> <sup>a</sup>
Ribosomal proteins	Small subunit of ribosome	<i>rps11</i> , <i>rps14</i> , <i>rps15</i> , <i>rps16</i> <sup>+</sup> , <i>rps18</i> , <i>rps2</i> , <i>rps3</i> , <i>rps4</i> , <i>rps7</i> <sup>a</sup> , <i>rps8</i> , <i>rps12</i> <sup>+,a</sup> , <i>rps19</i>
Transcription	Large subunit of ribosome	<i>rpl14</i> , <i>rpl16</i> , <i>rpl2</i> <sup>+</sup> , <i>rpl20</i> , <i>rpl22</i> , <i>rpl23</i> <sup>a</sup> , <i>rpl32</i> , <i>rpl33</i> , <i>rpl36</i>
	DNA-dependent RNA polymerase	<i>rpoA</i> , <i>rpoB</i> , <i>rpoC1</i> <sup>+</sup> , <i>rpoC2</i>
Protein genes Other genes	Photosystem I	<i>psaA</i> , <i>psaB</i> , <i>psaC</i> , <i>psaI</i> , <i>psaJ</i> , <i>pafI</i> <sup>++</sup> , <i>pafII</i>
	Photosystem II	<i>psbA</i> , <i>psbB</i> , <i>psbC</i> , <i>psbD</i> , <i>psbE</i> , <i>psbF</i> , <i>psbH</i> , <i>psbI</i> , <i>psbJ</i> , <i>psbK</i> , <i>psbL</i> , <i>psbM</i> , <i>psbT</i> , <i>psbZ</i>
	Subunit of cytochrome	<i>petA</i> , <i>petB</i> <sup>+</sup> , <i>petD</i> <sup>+</sup> , <i>petG</i> , <i>petL</i> , <i>petN</i>
	Subunit of ATP synthase	<i>atpA</i> , <i>atpB</i> , <i>atpE</i> , <i>atpF</i> <sup>+</sup> , <i>atpH</i> , <i>atpI</i>
	Chloroplast envelope membrane protein	<i>cemA</i>
	NADH dehydrogenase	<i>ndhA</i> <sup>+</sup> , <i>ndhC</i> , <i>ndhD</i> , <i>ndhE</i> , <i>ndhF</i> , <i>ndhG</i> , <i>ndhH</i> , <i>ndhI</i> , <i>ndhJ</i> , <i>ndhK</i>
	Large subunit of Rubisco	<i>rbcl</i>
	Subunit acetyl-coA carboxylase	<i>accD</i>
	ATP-dependent protease subunit P	<i>clpP1</i> <sup>++</sup>
	Maturase	<i>matK</i>
C-type cytochrome synthesis	<i>ccsA</i>	
Component of the TIC complex	<i>ycf1</i> <sup>a</sup>	
Hypothetical proteins	<i>ycf2</i> <sup>a</sup>	
Translation initiation factor	<i>infA</i>	
N-terminal nucleophile aminohydrolases (Ntn hydrolases) superfamily protein	<i>pbf1</i>	

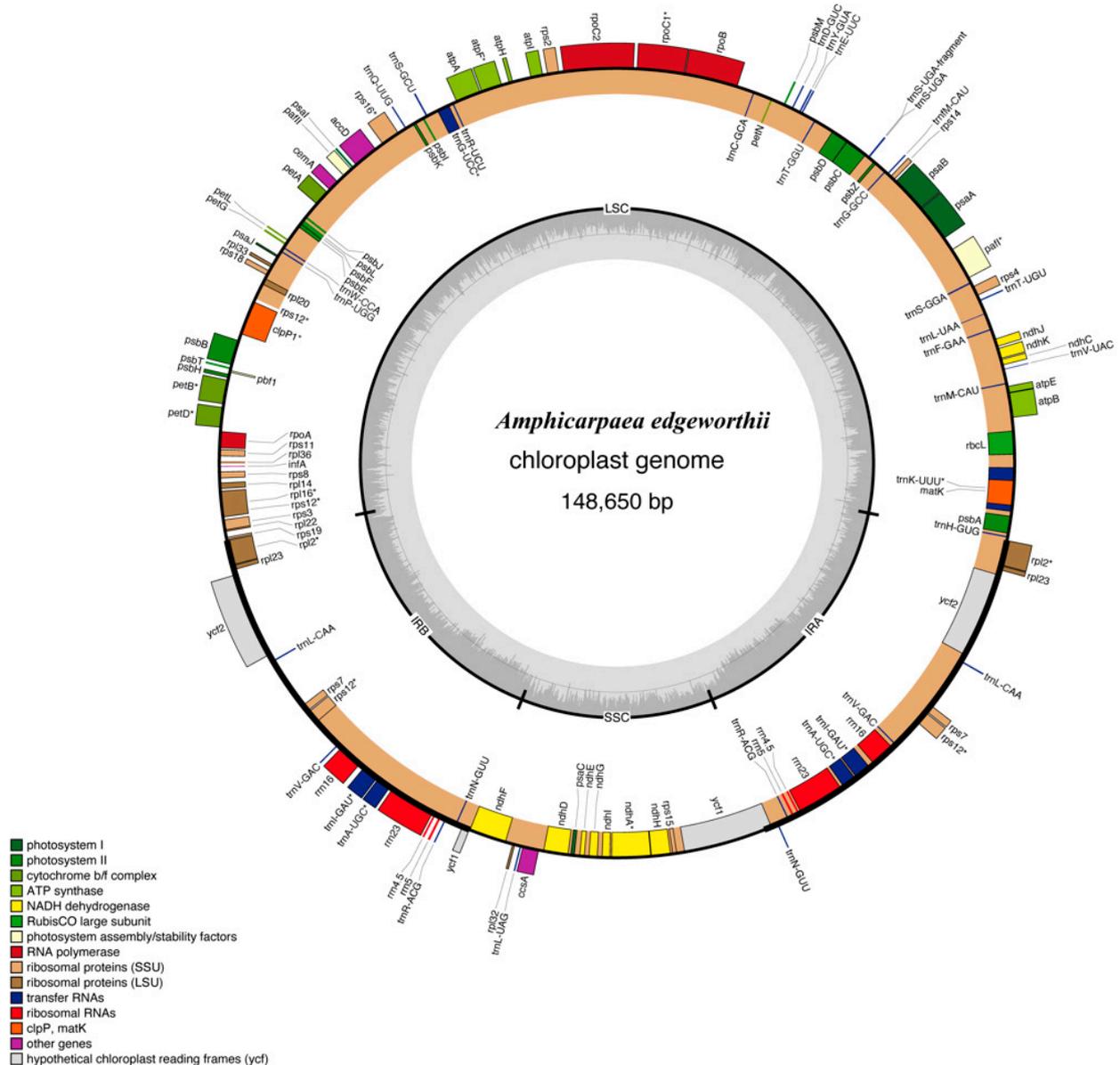
<sup>a</sup> Duplicated genes, <sup>+</sup> genes with one intron, and <sup>++</sup> genes with two introns.

In the plastome of *A. edgeworthii*, there are coding genes that contain introns (Table 3). Introns are reported to exist in some of the protein-coding and tRNAs genes of the plastome other angiosperms. Out of the 128 coding genes, 19 are characterized by one or two introns. Of these nineteen genes, six are tRNAs and thirteen are protein-coding genes.

Ten of the intron-containing genes are located in the LSC, one gene is in the SSC, and the remaining four are in the inverted repeat regions. ATP-dependent protease subunit p gene (*clpP1*) and photosystem I-related gene (*pafI*) possess two introns, while the remaining seventeen genes have only one. The tRNA gene, *trnK-UUU*, is the gene with the longest intron due to the inclusion of *matK* within its sequence.

**Table 3.** Genes with introns in the *A. edgeworthii* plastome and length of introns and exons.

Gene	Location	Exon I (bp)	Intron I (bp)	Exon II (bp)	Intron II (bp)	Exon III (bp)
ndhA	SSC	551	1273	541		
trnA-UGC	IR	38	807	35		
trnI-GAU	IR	42	944	37		
rps12	IR	114	535	232		
rpl2	IR	391	719	434		
rpl16	LSC	9	1176	399		
petD	LSC	8	729	475		
petB	LSC	6	812	642		
clpP1	LSC	71	708	292	790	228
rps16	LSC	40	879	230		



**Figure 1.** The structure of the *A. edgeworthii* plastome. Transcription occurs clockwise for genes located inside the circles and counterclockwise for those located outside the circles. The functional genes are indicated by colorful bars. The GC and AT contents of the inner circle are denoted by the dark gray and light gray colors, respectively. Genes with an asterisk \* indicate the presence of introns.

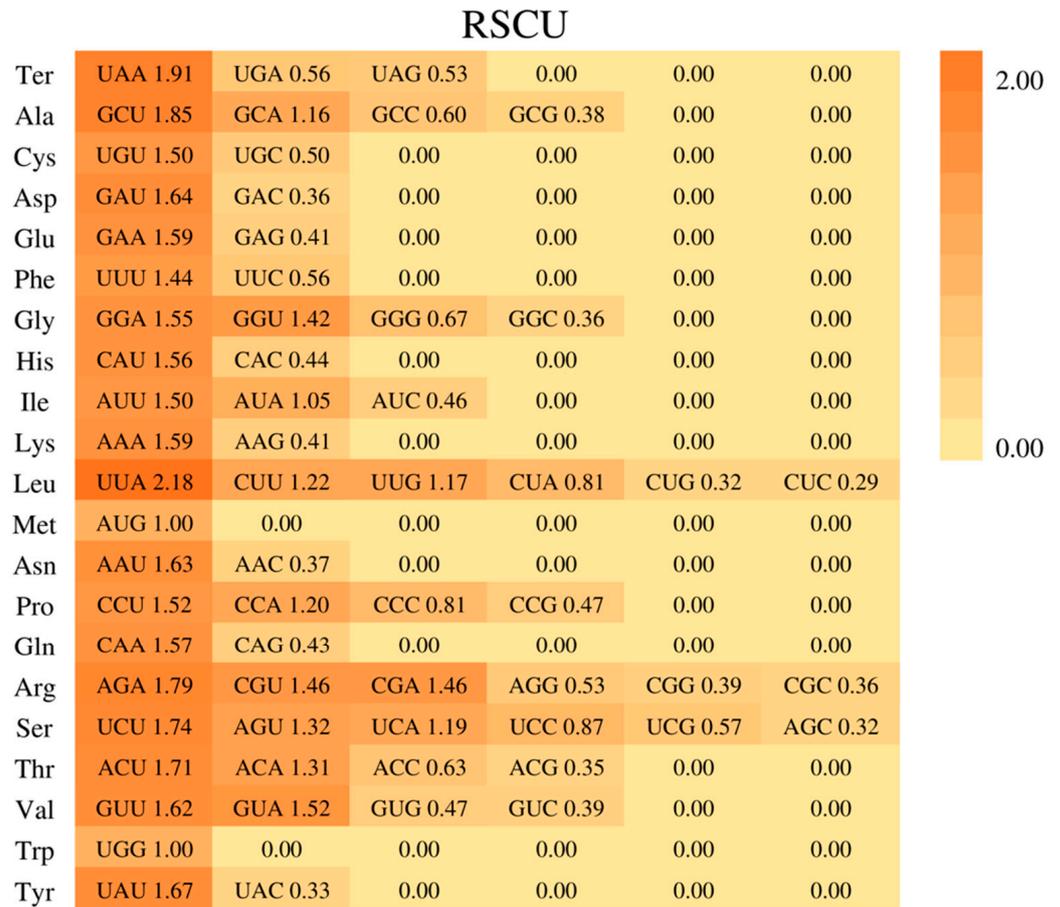


Figure 2. A heatmap of the RSCU values in the *A. edgeworthii* plastome.

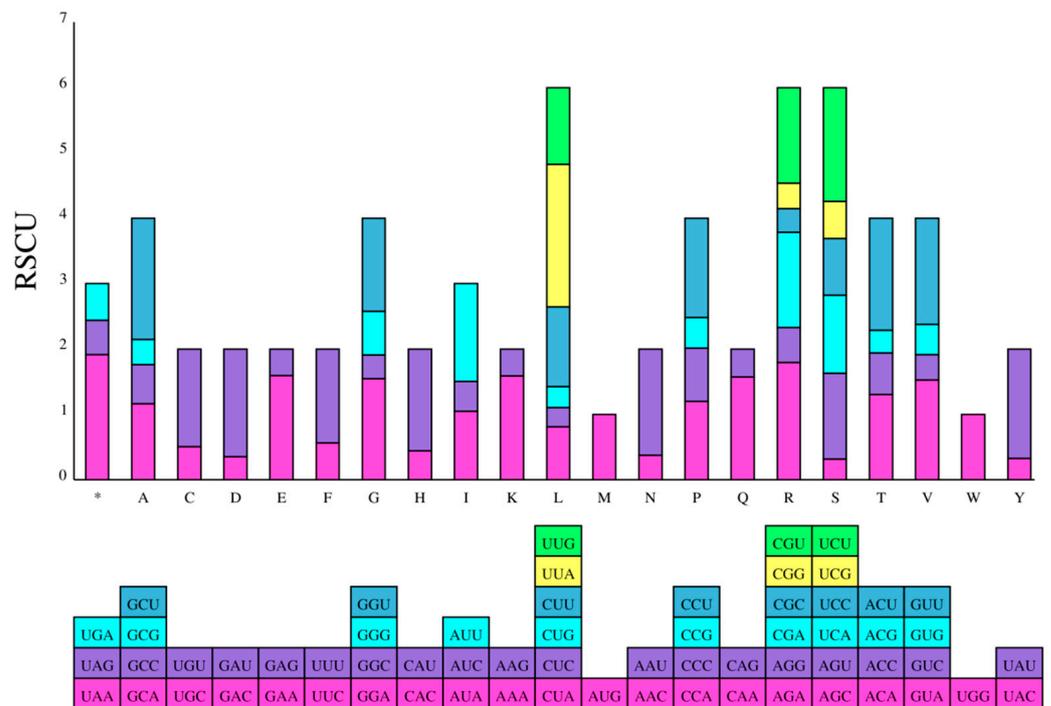
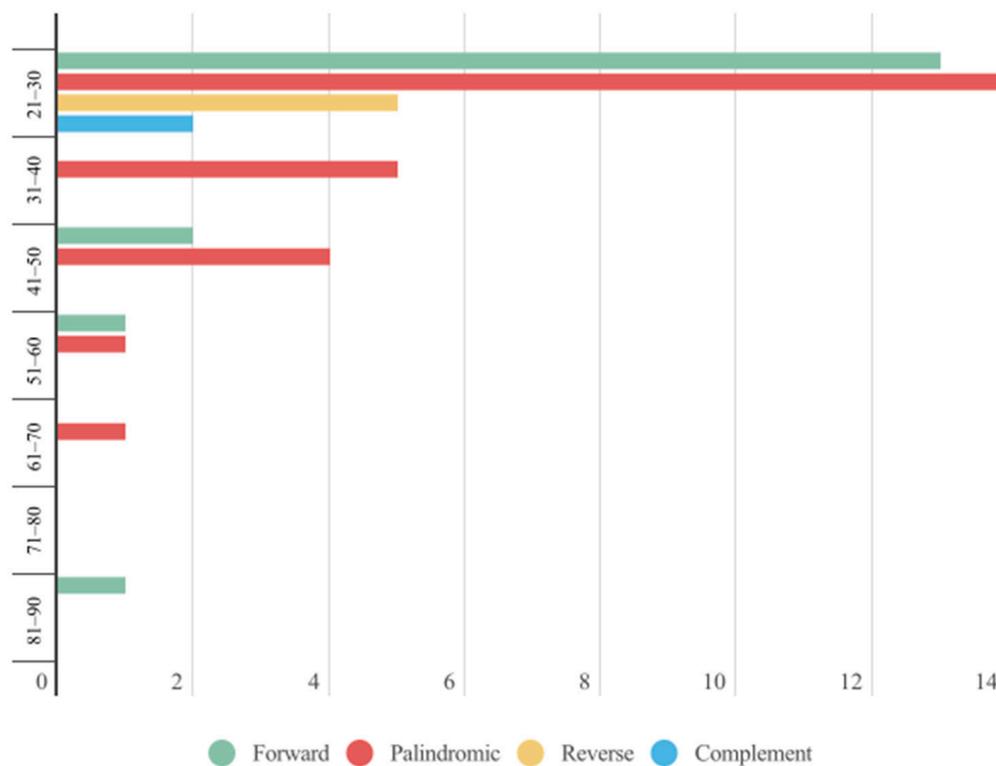


Figure 3. The stacked bar chart of RSCU values, with amino acids on the x-axis and the frequency of each codon on the y-axis. For each amino acid (column), represent each codon encoding it with a specific color. The \* columns denote stop codons.

## 2.2. Repeat Analyses

A comprehensive statistical analysis was conducted on the dispersed repetitive sequences in the plastome of *A. edgeworthii*. The analysis identified four types of dispersed repeat sequences with lengths greater than 20 bp, namely forward (F), palindromic (P), reverse (R), and complement (C) (Figure 4). Notably, P-type repeats were the most frequent, with a total count of 24. Interestingly, R and C types were only detected within the length range of 21–30 bp, while F-type repeats were exclusively identified within the length range of 81–90 bp, suggesting that they may play a unique role in the structural and functional organization of the plastid genome. These findings offer new insights into the nature and distribution of dispersed repeats in the plastome of *A. edgeworthii* and their potential impact on plastome evolution and function. Specifically, these dispersed repeats could induce DNA recombination, mutation, and gene transfer, ultimately contributing to the complexity and diversity of plastomes.

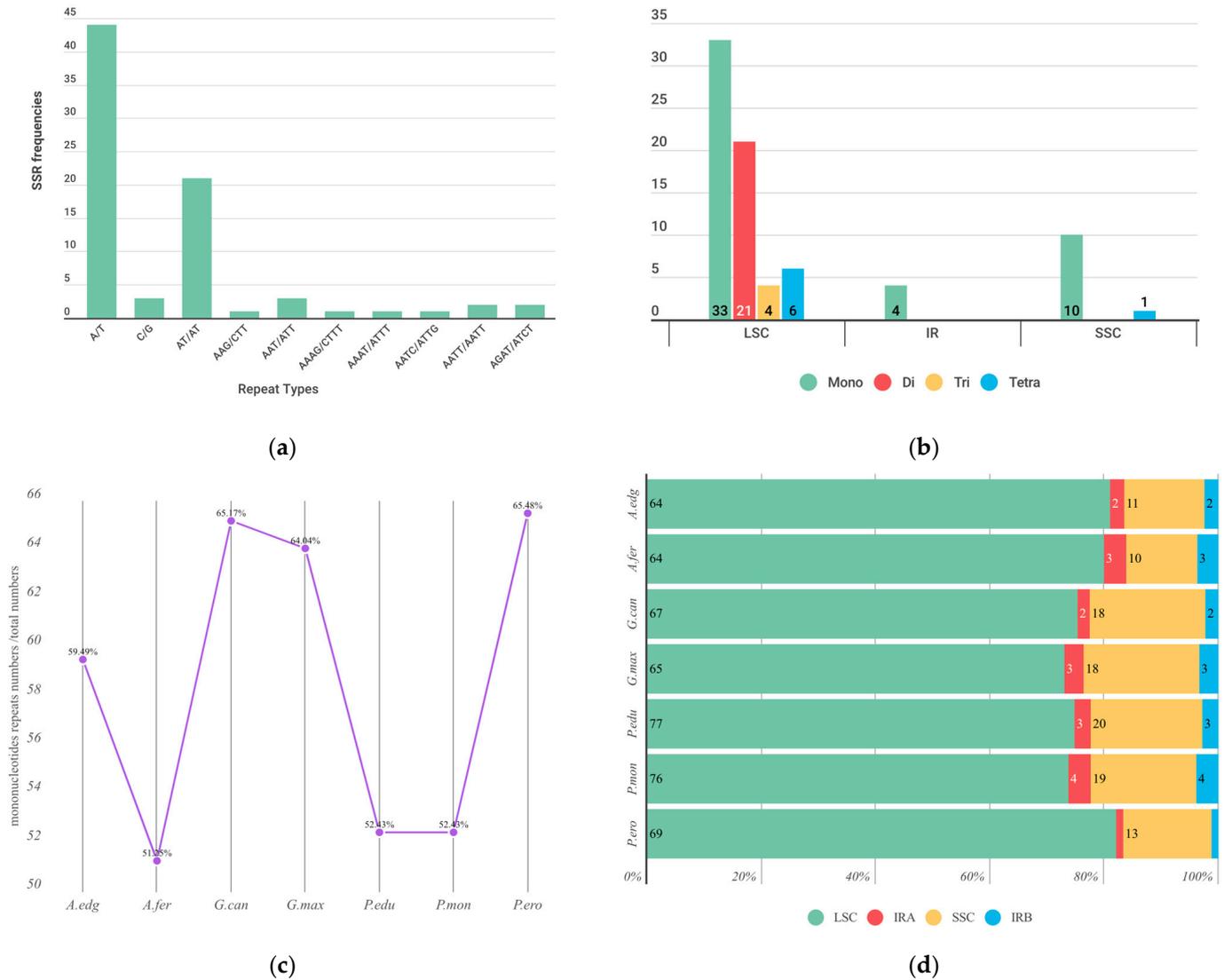


**Figure 4.** Bar chart showing the distribution of four scattered repetitive sequences across length intervals in the plastome of *A. edgeworthii*.

SSRs, also known as microsatellites, are valuable genetic markers for various applications in plant and animal breeding, conservation biology, and population genetics. Analyzing the distribution and diversity of SSRs in the plastome can lead to the development of SSR markers capable of distinguishing between different plant populations, species, or varieties based on their unique genetic fingerprints. SSR sequences in the plastomes of *A. edgeworthii* were identified using the MISA program. In this study, a total of 79 microsatellites were discovered. Mononucleotides were the most frequent SSRs, comprising approximately 59.49% of the total SSRs, with the majority being composed of A/T. Among dinucleotides, only AT/AT was found, while trinucleotides were represented by AAG/CTT and AAT/ATT. The tetranucleotides included AAAG/CTTT, AAAT/ATTT, AATC/ATTG, AATT/AATT, and AGAT/ATCT. No pentanucleotides or hexanucleotides were discovered. In terms of quantity, SSRs are mainly distributed in the LSC and SSC regions of the plastome. The LSC region harbors the most diverse types of SSRs, including mononucleotides, dinucleotides, trinucleotides, and tetranucleotides. Each of the IR regions

contains two mononucleotides. The SSC region includes ten mononucleotides and one tetranucleotide.

We also conducted SSR analysis on other species within Glycininae (Figure 5). Mononucleotides accounted for the highest proportion among the SSRs of seven species. *Glycine canescens* had the highest proportion of mononucleotides, followed by *Pachyrhizus erosus* and *G. max*, with all three species having mononucleotide ratios of around 65%. The lowest was observed in *A. ferruginea*, at 51%. The distribution pattern of SSRs was similar among the seven species, mainly located in the LSC region, with the SSRs in *P. erosus*'s LSC region accounting for the highest proportion, exceeding 80%.



**Figure 5.** Single sequence repeat (SSR) analysis of the plastome in Glycininae. (a) Frequency of different SSR motifs in different repeat types in *A. edgeworthii* plastome; (b) distribution of SSR in LSC, SSC, and IR regions; (c) proportion of single-nucleotide SSRs in seven Glycininae species; (d) proportion of single-nucleotide SSRs in seven Glycininae species.

The telomer restriction fragment (TRF) analysis report indicates the presence of satellite DNA and minisatellite DNA, excluding SSRs ranging from 1–6 bp in length. Tandem repeat sequences in the plastomes were identified using the program TRF (percent matches  $\geq 95\%$ , score  $> 90$ ). This analysis revealed 24 tandem repeats. Overall, the period size ranged between 10 and 27 bp, and the number of copies aligned with the consensus pattern was between 1.9 and 4.1. The highest number of occurrences was observed for a period size of 27 bp, followed by 14 bp.



We also compared the JLB, JSB, JSA, and JLA boundaries (Figure 8). The results showed some similarities and variations among the compared plastomes. The length of the seven plastomes ranged from 148,650 bp (*A. edgeworthii*) to 153,471 bp (*Pueraria edulis*). The boundaries of JSA, JLB, and JLA are very conservative, with the main differences being reflected in the boundary of JSB. In *A. edgeworthii*, *A. ferruginea*, and *G. max*, the *ycf1* gene spans from the IRb region to the SSC region, whilst in *G. canescens*, this position hosts the *ndhF* gene. In *P. edulis*, *P. montana*, and *P. erosus*, the *trnN* gene is located within the IRB region 500 to 800 bp away from the JSB region. Their boundaries show very small degrees of contraction and expansion, with differences between species not exceeding 300 bp. Overall, the plastome structure in the subtribe Glycininae appears to be stable and relatively homogeneous. The lack of large-scale genome rearrangements points to close evolutionary relationships between the subtribe species and a relatively recent origin of the clade.

#### 2.4. Divergence of Protein-Coding Gene Sequences

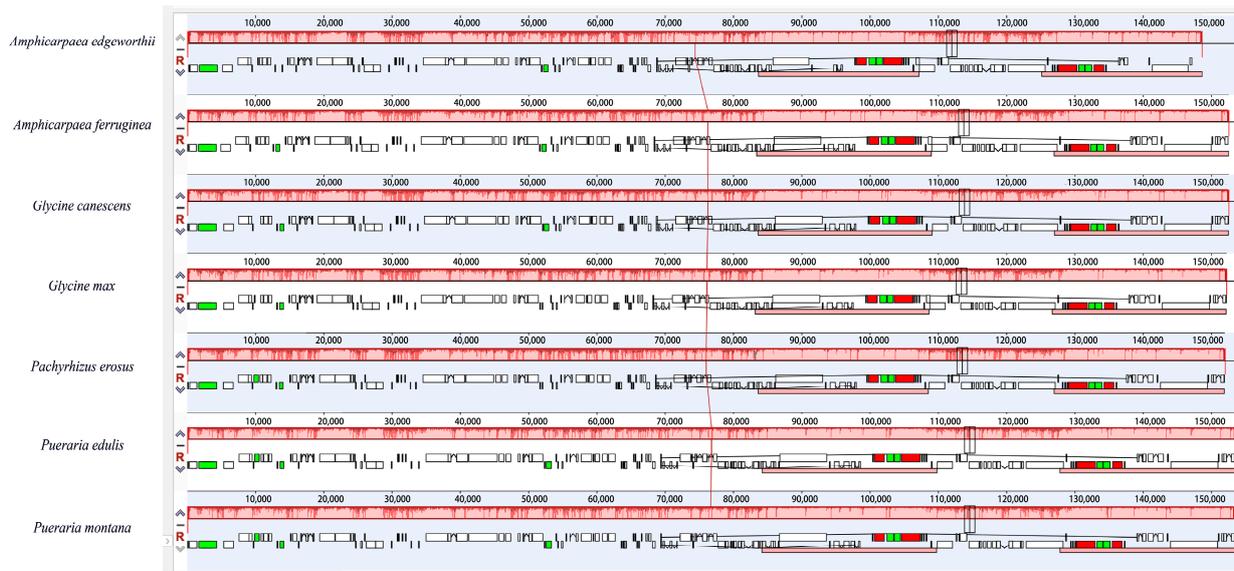
We performed manual curation on the annotated genome to eliminate any gene annotations that might have been duplicated. This work resulted in the identification of 79 single-copy orthologous genes from the genome sequences of 35 species. We calculated the level of nucleotide diversity (Pi) for each of these genes separately (Figure 9). The calculated pi values ranged from 0.00066 to 0.08757. Genes with high Pi values (>0.06) in the plastome include *matK*, *rps15*, *clpP1*, *ndhF*, *rpl32*, *ccsA*, *rpl20*, *cemA*, and *rpoC2*, with *matK* having the highest Pi value. Among these, *matK* encodes maturase, a protein that splices Group II introns, whereas *ndhF* is involved in the electron transfer chain of photosynthesis. Three other high-Pi genes code for ribosomal proteins (*rps15*, *rpl32*, *rpl20*), and the remaining ones also have significant functions. *ccsA* plays a role in plant response to oxidative stress, *cemA* is involved in the synthesis and maintenance of the cell wall, and *rpoC2* is a critical component of RNA polymerase. These genes may be associated with environmental changes, which are helpful for understanding the interaction between the Phaseoleae tribe and the environment during evolution.

#### 2.5. Phylogenetic Analysis

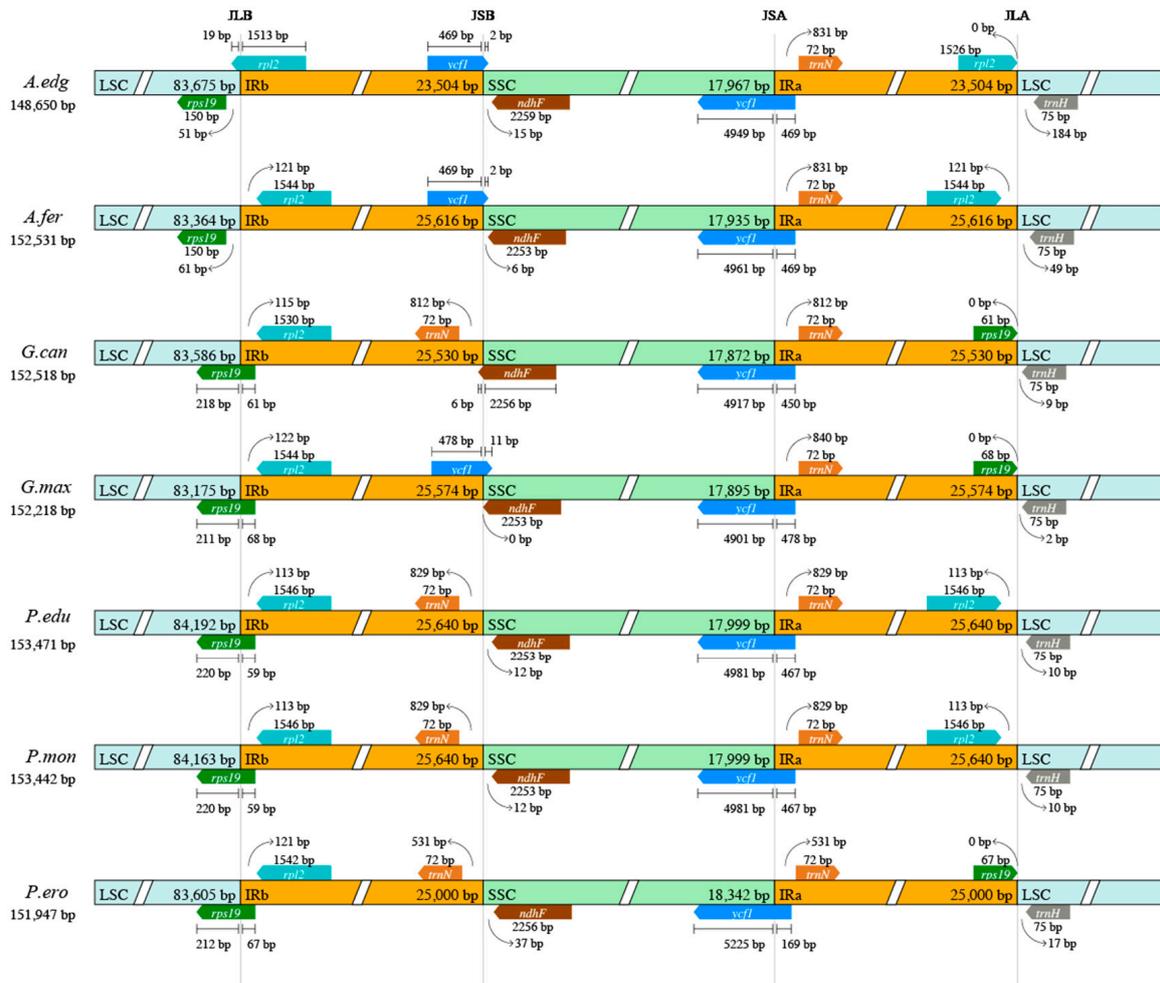
Past taxonomic disagreements have centered around whether the Erythrinae and Kennediinae should be considered independent subtribes within the subtribal classification system of Phaseoleae. Our study addressed this issue by analyzing a representative sample of the Phaseoleae tribe, comprising seven subtribes, including Erythrinae and Kennediinae.

Our phylogenetic analysis yielded a highly resolved and well-supported evolutionary tree, clearly delineating the relationships among all the studied species (Figure 10). This tree comprises two primary sister clades: the Diocleinae clade (Clade I) and a clade consisting of the remaining species (Clade II). Within Clade II, the Erythrinae are polyphyletic, while the other five subtribes are monophyletic. The Glycininae clade, occupying a derived position, establishes a sister relationship with the Phaseolinae clade. Genes located at the junctions in Phaseoleae are *rps11*, *rps19*, *rps8*, *rps3*, *rpl2*, *rpl22*, *rpl23*, *trnN*, *ndhF*, and *ycf1*.

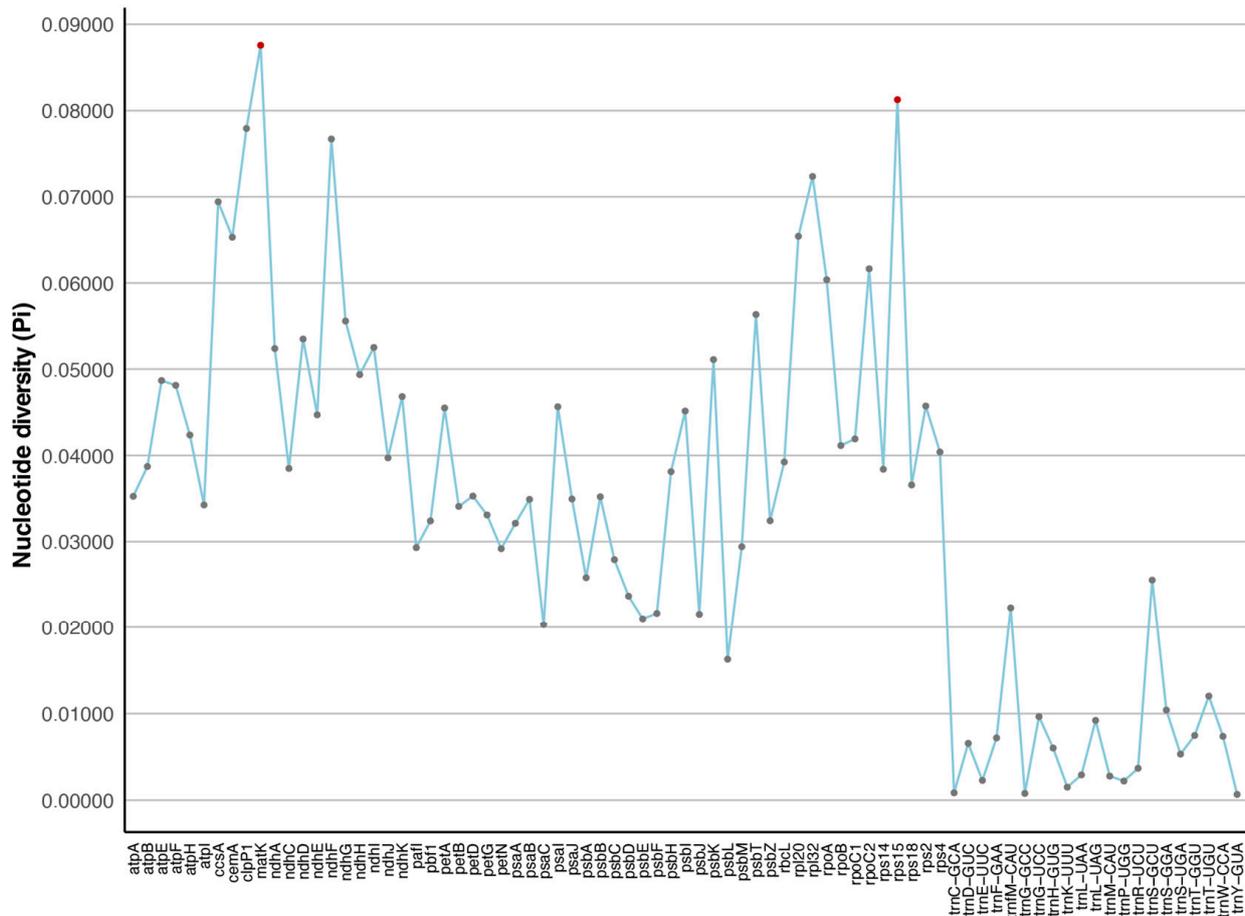
Clitorinae forms a sister relationship to the rest of the species in Clade II. The Kennediinae clade showed a relatively early diverged position compared to the remaining subtribes. Lackey once speculated on the synonymy between Kennediinae and Diocleinae, noting similarities in floral, pod, and seed attachment. However, the absence of bracteoles, the prominent aril, and geographical isolation supported the independence of Kennediinae [12]. The recent recognition of Kennediinae as an independent subtribe was mentioned in studies related to cotyledon areoles in 2008 [44]. Currently, phylogenetic evidence increasingly supports the classification of Kennediinae as a distinct subtribe within Phaseoleae.



**Figure 7.** Plastome alignment visualized using a Mauve multiple alignment plot. The color of each sequence line represents its position in the genome sequence, with similar regions having similar colors. The color boxes display the annotated features of the plastome sequences, where green corresponds to tRNAs, red corresponds to rRNAs, and white represents genes containing CDS.



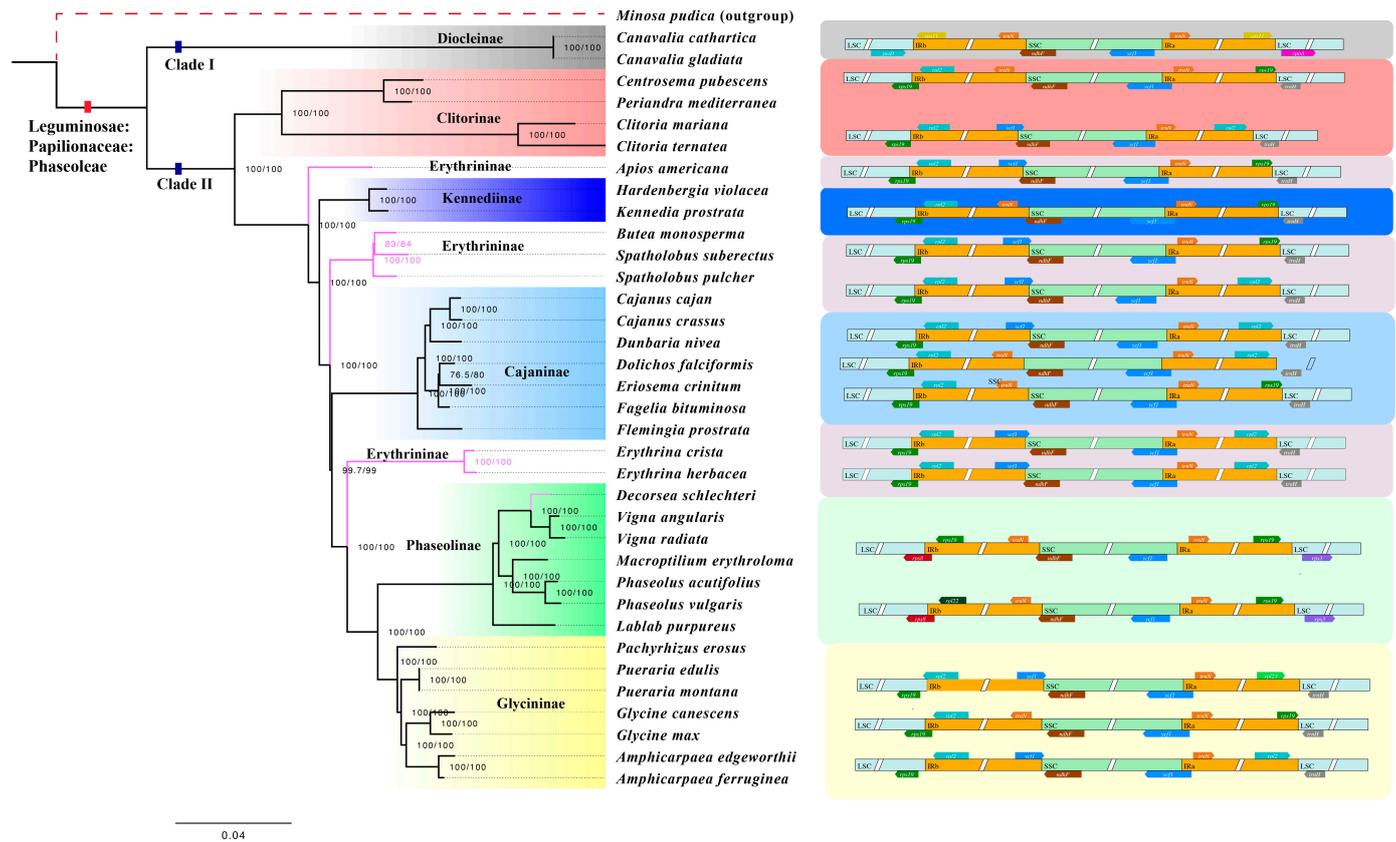
**Figure 8.** Structural variation in the junction of inverted repeat and single-copy regions among the seven plastomes of Glycininae. (JSA: junction of the SSC and the IRA; JLB: junction of the LSC and the IRb; JSB: junction of the SSC and the IRb).



**Figure 9.** Sliding window analysis of nucleotide variability among the seven Glycininae species plastomes (window length: 600 bp; step size: 200 bp).

Regarding the polyphyletic Erythrinae, species taxonomically ascribed to this subtribe, namely *Apios*, *Butea*, *Cochlianthus*, *Erythrina*, *Mucuna*, and *Spatholobus*, are separated into four subclades (pink clades, Figure 10). Three subclades of Erythrinae form a paraphyletic group with Kennediinae, Cajaninae, and Phaseolinae. It is noteworthy that *Decorsea schlechteri* belonging to Erythrinae are embedded within Phaseolinae, showing a sister relationship with *Vigna*. *Apios americana* shares specific genes on JLB and JSB with the genus *Clitoria*, while other genes on JLA and JSA are shared with the Kennediinae subtribe, reflecting its evolutionary intermediary position. For the genus *Spatholobus*, two species exhibit boundaries consistent with the *Cajanus* genus. In contrast, for the *Butea* genus, boundary genes at JLB and LSB align with *Cajanus*, whereas JLA and JSA boundary genes match with the Kennediinae subtribe. *Erythrina*, as the representative genus, shows that Asian species are derived from a primarily African clade, with South American species being basal [45]. By comparing the junctions, the *Erythrina* genus shares some boundary genes with the *Cajanus* genus. Lackey considered that Erythrinae could potentially include various lineages originating from a Galegeae–Dalbergieae stock [12]. Despite this, Galegeae is in the IRLC clade (Inverted Repeat-Lacking Clade), and Dalbergieae is in the Dalbergioid clade. We believe that the classification based on morphology in FRPS needs to be improved, and the incertae sedis in the GRIN also needs to be revised in the light of phylogenomics.

In addition to the lingering issue of the status of Kennediinae and Erythrinae, the position of *Decorsea*, an undetermined subtribe genus traditionally placed in subtribe Phaseolinae, which exhibits a sister relationship with the genus *Vigna* in our tree (Figure 10), is also noteworthy. Its boundary genes are also in line with those of the *Vigna* genus, suggesting the possibility of incorporating *Decorsea* into the Phaseolinae subtribe.



**Figure 10.** Phylogenetic tree of Phaseoleae reconstructed using the plastomes of 35 taxa. The branch nodes' numbers represent the posterior probabilities values (PPs). Colors shaded on the branches indicate the corresponding subtribes. The pink branches on the phylogenetic tree represent members from the Erythrinae. On the right is the pattern of structural variations in the junctions of inverted repeat (IR) and single-copy (SC) regions for each subtribe, with genes at the IR/SC junctions indicated.

Within the Glycininae subtribe, *Amphicarpaea* and *Glycine* form a clade, which is sister to *Pueraria*. Our findings are similar to the tree constructed using *rps16* intron sequences by Lee et al. [23]. In their analysis, *Pueraria* exhibits non-monophyly, whereas we utilized complete genomes from two species within the *Pueraria* genus, and both remain non-dispersed.

### 3. Materials and Methods

#### 3.1. Plant Material and DNA Extraction

Leaves from healthy individuals of *A. edgeworthii* were collected and immediately frozen in liquid nitrogen for preservation. Leaves were prepared for DNA extraction with care to avoid excess mucilage. The total genomic DNA was extracted using a TIANGEN Plant Genomic DNA kit (Beijing, China) following the manufacturer's guide. Then, DNA concentration and quality were assessed by a NanoDrop 2000 spectrophotometer (Thermo Scientific, DE, USA) and 1% agarose gel electrophoresis. Qualified DNA was sent to Major-bio (Shanghai, China) for library preparation and high-throughput sequencing using Illumina Novaseq 6000 Platform (Illumina, CA, USA) with 150 bp paired-end reads.

#### 3.2. DNA Sequencing and Genome Assembly

To prepare the DNA samples, 1.0 µg of high-quality genomic DNA was sheared into fragments of approximately 350 bp using a Covaris S220 instrument (Covaris, Woburn, MA, USA). The construction of sequencing libraries was performed using the NEBNext Ultra II DNA Library Prep Kit (New England Biolabs, Ipswich, MA, USA) following the manufacturer's instructions. Subsequently, library quantification was conducted with a Qubit dsDNA HS Assay Kit (Life Technologies, Carlsbad, CA, USA), and size distribution

was assessed using an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA). After pooling, the libraries were subjected to PCR enrichment and purified using the AMPure XP system (Beckman Coulter, Brea, CA, USA). The qualified libraries based on projected data volume and effective concentration were then sequenced on an Illumina HiSeq X Ten platform (Illumina, San Diego, CA, USA).

The raw sequencing data were processed using PRINSEQlite v0.20.4 to obtain high-quality reads (5.2 GB) by filtering out low-quality reads, adapters, and ambiguous bases [46]. Plastome assembly was performed using NOVOPlasty 4.2 software, which employs a de novo assembly approach with a reference seed for the plastome [47]. The seed sequence was obtained from a closely related species and was used as a reference for the assembly process. The high-quality reads were used to generate contigs with k-mer sizes of 39, and the contigs were then merged iteratively to produce a draft plastome assembly.

### 3.3. Genome Annotation: Genes and Repetitive Elements

Gene annotation was carried out with GeSeq (Annotation of Organellar Genomes) and CPGAVAS2 using *Medicago turbinata* (NC\_068638.1) and *G. max* (CM010429.1) as reference genomes [48,49]. The 36 annotated plastomes were manually curated by reviewing gene boundaries and correcting misaligned gene fragments. Additionally, the orientation of the SSC and LSC regions of all genomes was standardized. A graphical representation of the plastome was drawn using OGDRAW software (<http://ogdraw.mpimp-golm.mpg.de/>, accessed on 25 March 2023) [50].

Simple sequence repeats (SSRs) were identified and analyzed using MISA (MicroSatellite identification) (<https://webblast.ipk-gatersleben.de/misa/>, accessed on 26 March 2023) [51]. For the detection of SSR motifs in mononucleotides, dinucleotides, trinucleotides, tetra-, penta-, and hexanucleotides, the minimum number of repeat units used were ten, five, four, and three, respectively. The online software REPuter was used to characterize the long repeat sequences in the plastome [52]. To identify dispersed repetitive elements, we employed a TRF (Tandem Repeats Finder) and performed the analysis with the default parameter [53].

### 3.4. Genome Comparison and Phylogenetic Analysis

For whole structure analysis, using the annotation of *A. edgeworthii* as a reference in the LAGAN mode, the plastomes of six other species of Glycininae (*A. ferruginea*, *G. max*, *G. canescens*, *P. montana*, *P. edulis*, *P. erosus*) were compared using the tool mVISTA [54]. The border regions of the manually curated plastomes were visualized on the online Genepioneer Cloud Platform (<http://cloud.genepioneer.com:9929>, accessed on 26 March 2023). To identify structural variations and sequence divergences between the genomes, we utilized the software progressiveMauve adopting automatic calculation of seed weights and the minimum LCB (locally collinear block) score parameter [55].

To analyze the phylogenetic relationships among 36 species, including 35 species of Phaseoleae and one outgroup species (Appendix A), we selected 35 commonly listed chloroplast genes. To construct a whole-genome phylogenetic tree, we employed MAFFT to align the sequences and used IQ-TREE's ModelFinder and ultrafast bootstrap (UFBoot) modules to build the tree [56–59]. To construct a CDS gene tree, 79 single-copy orthologous genes shared among the 36 species considered were extracted from their respective genomes. Each gene was aligned using MAFFT and concatenated to form a large matrix, which was then used to build a tree using IQ-TREE. For the whole-genome phylogenetic tree, the best-fit model based on the Bayesian information criterion (BIC) was K3Pu+F+I+I+R3, and the best-fit model of the CDS tree was TVM+F+R3. The whole-genome tree and CDS tree were visualized using FigTree [60].

## 4. Conclusions

Our comprehensive study of *A. edgeworthii* revealed a plastome with a genome size of 148,650 bp with 128 genes. The plastome is characterized by a predominance of palin-

dromic repeats and a notable presence of 79 microsatellites, mainly in the LSC and SSC regions. Comparative genomic analysis across seven Glycininae species highlighted a universally conserved plastome structure and significant nucleotide diversity. Phylogenetic reconstruction of 35 Phaseoleae species emphasized *Amphicarpaea*'s affinity with *Glycine*, positioning *Apios* as a sister lineage to other Phaseoleae species, excluding the *Clitorinae* and *Diocleinae* subtribes. Within the Glycininae subtribe, *Amphicarpaea* and *Glycine* have a sister relationship, with *Pueraria* closely related. Our findings suggest the retention of Kennediinae within the Phaseoleae tribe and advise against the independent categorization of Erythrinae. While refining the phylogenetic contours of the Phaseoleae tribe, our findings point to the need to reevaluate the current classification of the Erythrinae. This study not only enhances our understanding of the plastomic architecture and phylogenetic relationships in the Phaseoleae tribe but also lays a foundation for future research in legume evolution and crop improvement.

**Author Contributions:** Conceptualization, Z.-C.Q. and X.-L.Y.; data curation, Y.-N.X., X.-Q.W., and L.-L.D.; formal analysis, Y.-N.X. and X.-Q.W.; funding acquisition, Z.-C.Q., Y.-T.S., and X.-L.Y.; investigation, Z.-C.Q. and X.-L.Y.; methodology, L.-L.D., X.-Y.B., X.-Q.W., Y.-Q.F., and Y.-N.X.; project administration, Z.-C.Q. and X.-L.Y.; resources, Z.-C.Q., Y.-T.S., and X.-L.Y.; validation, X.-Y.B. and Y.-Q.F.; visualization, Y.-N.X.; writing—original draft, Y.-N.X. and Z.-C.Q.; writing—review and editing, Y.-N.X., Z.-C.Q., and X.-L.Y. All authors have read and agreed to the published version of the manuscript.

**Funding:** This study was funded by the Special Fund for Scientific Research of Shanghai Landscaping and City Appearance Administrative Bureau, grant numbers G222403 and G242412; the National Wild Plant Germplasm Resource Center, grant number ZWGX2202; and the Natural Science Foundation of Zhejiang Province, grant number LY21C030008.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The chloroplast genome of *A. edgeworthii* is available in the NCBI under accession number OP749930. The associated BioProject, SRA, and Bio-Sample numbers are PRJNA732525, SRR14638108, and SAMN19321915, respectively.

**Acknowledgments:** We are grateful to the reviewers for their thorough reviews and suggestions that helped to improve this paper.

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

## Appendix A

A list of 36 plastomes of Leguminosae, all sourced from the NCBI database (<https://www.ncbi.nlm.nih.gov/>). This collection includes various species in the Phaseoleae tribe, which is part of the NPAAA clade: *Glycine max*, NC\_007942.1; *Phaseolus vulgaris*, NC\_009259; *Vigna radiata*, NC\_013843.1; *Vigna angularis*, NC\_021091.1; *Glycine canescens*, NC\_021647.1; *Apios americana*, NC\_025909.1; *Pachyrhizus eros*, NC\_026682.1; *Cajanus cajan*, NC\_031429.1; *Canavalia cathartica*, NC\_047311.1; *Clitoria ternate*, NC\_047365.1; *Butea monosperma*, NC\_047384.1; *Spatholobus suberectus*, NC\_048966.1; *Spatholobus pulcher*, NC\_049094.1; *Canavalia gladiata*, NC\_050951.1; *Lablab purpureus*, NC\_054310.1; *Cajanus crassus*, NC\_057277.1; *Centrosema pubes*, NC\_057278.1; *Decorsea schlechteri*, NC\_057444.1; *Dolichos falciformis*, NC\_057447.1; *Dunbaria nivea*, NC\_057448.1; *Eriosema crinitum*, NC\_057449.1; *Erythrina crista*, NC\_057450.1; *Fagelia bituminosa*, NC\_057451.1; *Hardenbergia violacea*, NC\_057453.1; *Kennedia prostrata*, NC\_057454.1; *Pueraria montana*, NC\_060608.1; *Amphicarpaea ferruginea*, NC\_063696.1; *Pueraria edulis*, NC\_065692.1; *Flemingia prostrata*, NC\_065863.1; *Periandra mediterranea*, NC\_067529.1; *Clitoria mariana*, NC\_067531.1; *Erythrina herbacea*, NC\_067539.1; *Macroptilium erythroloma*, NC\_067541.1; and *Phaseolus acutifolius*, NC\_067543.1 and *Amphicarpaea edgeworthii*, OP749930. The species *Minosa pudica* (NC\_042921.1) was used as an outgroup.

## References

1. Vasconcelos, M.W.; Grusak, M.A.; Pinto, E.; Gomes, A.; Ferreira, H.; Balázs, B.; Centofanti, T.; Ntatsi, G.; Savvas, D.; Karkanis, A. The Biology of Legumes and Their Agronomic, Economic, and Social Impact. In *The Plant Family Fabaceae: Biology and Physiological Responses to Environmental Stresses*; Springer: Berlin/Heidelberg, Germany, 2020; pp. 3–25.
2. Ferreira, H.; Pinto, E.; Vasconcelos, M.W. Legumes as a Cornerstone of the Transition toward More Sustainable Agri-Food Systems and Diets in Europe. *Front. Sustain. Food Syst.* **2021**, *5*, 694121. [[CrossRef](#)]
3. Didinger, C.; Thompson, H.J. The Role of Pulses in Improving Human Health: A Review. *Legume Sci.* **2022**, *4*, e147. [[CrossRef](#)]
4. Aasfar, A.; Bargaz, A.; Yaakoubi, K.; Hilali, A.; Bennis, I.; Zeroual, Y.; Meftah Kadmiri, I. Nitrogen Fixing Azotobacter Species as Potential Soil Biological Enhancers for Crop Nutrition and Yield Stability. *Front. Microbiol.* **2021**, *12*, 628379. [[CrossRef](#)]
5. Ahmad, E.; Zaidi, A.; Khan, M.S.; Oves, M. *Heavy Metal Toxicity to Symbiotic Nitrogen-Fixing Microorganism and Host Legumes*; Springer: Berlin/Heidelberg, Germany, 2012; ISBN 3-7091-0729-6.
6. Gopalakrishnan, S.; Sathya, A.; Vijayabharathi, R.; Varshney, R.K.; Gowda, C.L.; Krishnamurthy, L. Plant Growth Promoting Rhizobia: Challenges and Opportunities. *3 Biotech* **2015**, *5*, 355–377. [[CrossRef](#)] [[PubMed](#)]
7. Abusaief, H.M.A.-A.; Boasoul, S.H. A Taxonomic Study of Twelve Wild Forage Species of Fabaceae. *Heliyon* **2021**, *7*, e06077. [[CrossRef](#)]
8. Cardoso, D.; de Queiroz, L.P.; Pennington, R.T.; de Lima, H.C.; Fonty, E.; Wojciechowski, M.F.; Lavin, M. Revisiting the Phylogeny of Papilionoid Legumes: New Insights from Comprehensively Sampled Early-Branching Lineages. *Am. J. Bot.* **2012**, *99*, 1991–2013. [[CrossRef](#)]
9. Group, T.L.P.W.; Bruneau, A.; Doyle, J.J.; Herendeen, P.; Hughes, C.; Kenicer, G.; Lewis, G.; Mackinder, B.; Pennington, R.T.; Sanderson, M.J.; et al. Legume Phylogeny and Classification in the 21st Century: Progress, Prospects and Lessons for Other Species-Rich Clades. *TAXON* **2013**, *62*, 217–248. [[CrossRef](#)]
10. Melgar, A.E.; Zelada, A.M. Evolutionary Analysis of Angiosperm Dehydrin Gene Family Reveals Three Orthologues Groups Associated to Specific Protein Domains. *Sci. Rep.* **2021**, *11*, 23869. [[CrossRef](#)]
11. Wojciechowski, M.F. Towards a New Classification of Leguminosae: Naming Clades Using Non-Linnaean Phylogenetic Nomenclature. *S. Afr. J. Bot.* **2013**, *89*, 85–93. [[CrossRef](#)]
12. Lackey, J.A. A Revised Classification of the Tribe Phaseoleae (Leguminosae: Papilionoideae), and Its Relation to Canavanine Distribution. *Bot. J. Linn. Soc.* **1977**, *74*, 163–178. [[CrossRef](#)]
13. Li, S. Phaseoleae. In *Flora Reipublicae Popularis Sinicae*; Wu, C.Y., Ed.; Science Press: Beijing, China, 1995; Volume 41, ISBN 7-03-004440-1.
14. Oyebanji, O.; Zhang, R.; Chen, S.-Y.; Yi, T. New Insights Into the Plastome Evolution of the Millettoid/Phaseoloid Clade (Papilionoideae, Leguminosae). *Front. Plant Sci.* **2020**, *11*, 151. [[CrossRef](#)]
15. Cheplick, G.P. Life History Evolution in Amphicarpic Plants. *Plant Species Biol.* **1994**, *9*, 119–131. [[CrossRef](#)]
16. Schnee, B.K.; Waller, D.M. Reproductive Behavior of *Amphicarpaea bracteata* (Leguminosae), an Amphicarpic Annual. *Am. J. Bot.* **1986**, *73*, 376–386. [[CrossRef](#)]
17. Cheplick, G.P. The Ecology of Amphicarpic Plants. *Trends Ecol. Evol.* **1987**, *2*, 97–101. [[CrossRef](#)] [[PubMed](#)]
18. Zhang, K.; Baskin, J.M.; Baskin, C.C.; Cheplick, G.P.; Yang, X.; Huang, Z. Amphicarpic Plants: Definition, Ecology, Geographic Distribution, Systematics, Life History, Evolution and Use in Agriculture. *Biol. Rev.* **2020**, *95*, 1442–1466. [[CrossRef](#)]
19. Song, T.; Zhou, M.; Yuan, Y.; Yu, J.; Cai, H.; Li, J.; Chen, Y.; Bai, Y.; Zhou, G.; Cui, G. Chromosome-Scale Reference Genome of *Amphicarpaea Edgeworthii*: A New Resource for Amphicarpic Plants Research and Complex Flowering Pattern. *Front. Plant Sci.* **2021**, *12*, 770660. [[CrossRef](#)] [[PubMed](#)]
20. Song, T.; Zhou, M.; Yuan, Y.; Yu, J.; Cai, H.; Li, J.; Chen, Y.; Bai, Y.; Zhou, G.; Cui, G. First High-Quality Reference Genome of *Amphicarpaea edgeworthii*. *bioRxiv* **2020**. [[CrossRef](#)]
21. Liu, Y.; Zhang, X.; Han, K.; Li, R.; Xu, G.; Han, Y.; Cui, F.; Fan, S.; Seim, I.; Fan, G.; et al. Insights into Amphicarpic Plants from the Compact Genome of the Legume *Amphicarpaea edgeworthii*. *Plant Biotechnol. J.* **2021**, *19*, 952–965. [[CrossRef](#)] [[PubMed](#)]
22. Xiao, Y.; Zhao, Z.-N.; Ping, H.-L. The Complete Plastid Genome of *Amphicarpaea ferruginea* Benth (Leguminosae), a Grass Species with Development and Utilization Prospect. *Mitochondrial DNA B Resour.* **2022**, *7*, 1221–1223. [[CrossRef](#)]
23. Lee, J.; Hymowitz, T. A Molecular Phylogenetic Study of the Subtribe Glycininae (Leguminosae) Derived from the Chloroplast DNA Rps16 Intron Sequences. *Am. J. Bot.* **2001**, *88*, 2064–2073. [[CrossRef](#)]
24. Agyei, D. Bioactive Proteins and Peptides from Soybeans. *Recent Pat. Food Nutr. Agric.* **2015**, *7*, 100–107. [[CrossRef](#)]
25. Rackis, J.J. Biological and Physiological Factors in Soybeans. *J. Am. Oil Chem. Soc.* **1974**, *51*, 161A–174A. [[CrossRef](#)]
26. Wu, D.; Zhang, W.; Xiu, L.; Sun, Y.; Gu, W.; Wang, Y.; Zhang, H.; Chen, W. Soybean Yield Response of Biochar-Regulated Soil Properties and Root Growth Strategy. *Agronomy* **2022**, *12*, 1412. [[CrossRef](#)]
27. Kirchhoff, H. Chloroplast Ultrastructure in Plants. *New Phytol.* **2019**, *223*, 565–574. [[CrossRef](#)]
28. Ozeki, H.; Umesono, K.; Inokuchi, H.; Kohchi, T.; Ohyama, K. The Chloroplast Genome of Plants: A Unique Origin. *Genome* **1989**, *31*, 169–174. [[CrossRef](#)]
29. Yang, Q.; Jiang, Y.; Wang, Y.; Han, R.; Liang, Z.; He, Q.; Jia, Q. SSR Loci Analysis in Transcriptome and Molecular Marker Development in *Polygonatum sibiricum*. *Biomed Res Int* **2022**, *2022*, 4237913. [[CrossRef](#)]
30. Wang, Q.; Chen, M.-M.; Hu, X.-F.; Wang, R.-H.; He, Q.-L. The Complete Chloroplast Genome Sequence of *Spiraea japonica* var. *acuminata* Franch. (Rosaceae). *Mitochondrial DNA B Resour.* **2022**, *7*, 275–276. [[CrossRef](#)] [[PubMed](#)]

31. Chang, Q.; Li, Y.; Chen, X.; Yan, Y.; Xia, P. Characterization of the Complete Chloroplast Genome Sequence of *Elaeagnus henryi* Warb. Ex Diels (Elaeagnaceae). *Mitochondrial DNA B Resour.* **2022**, *7*, 1876–1878. [[CrossRef](#)]
32. Li, Q.; Chen, X.; Yang, D.; Xia, P. Genetic Relationship of *Pleione* Based on the Chloroplast Genome. *Gene* **2023**, *858*, 147203. [[CrossRef](#)]
33. Guo, L.; Wang, X.; Wang, R.; Li, P. Characterization and Comparative Analysis of Chloroplast Genomes of Medicinal Herb *Scrophularia ningpoensis* and Its Common Adulterants (Scrophulariaceae). *Int. J. Mol. Sci.* **2023**, *24*, 10034. [[CrossRef](#)]
34. Wang, R.; Gao, J.; Feng, J.; Yang, Z.; Qi, Z.; Li, P.; Fu, C. Comparative and Phylogenetic Analyses of Complete Chloroplast Genomes of *Scrophularia incisa* Complex (Scrophulariaceae). *Genes* **2022**, *13*, 1691. [[CrossRef](#)] [[PubMed](#)]
35. Zhang, M.; Chen, M.-M.; Zhang, X.-M.; Chen, S.-N.; Liang, Z.-S. The Complete Chloroplast Genome Sequence of Traditional Chinese Medicine *Uncaria Macrophylla* (Rubiaceae). *Mitochondrial DNA B Resour.* **2022**, *7*, 694–695. [[CrossRef](#)] [[PubMed](#)]
36. Xu, X.; Yao, X.; Zhang, C.; Xia, P. Characterization of the Complete Chloroplast Genome Sequence of *Cardamine lyrata* Bunge (Brassicaceae). *Mitochondrial DNA B Resour.* **2022**, *7*, 936–937. [[CrossRef](#)]
37. Chen, M.-M.; Zhang, M.; Liang, Z.-S.; He, Q.-L. Characterization and Comparative Analysis of Chloroplast Genomes in Five *Uncaria* Species Endemic to China. *Int. J. Mol. Sci.* **2022**, *23*, 11617. [[CrossRef](#)] [[PubMed](#)]
38. Zhou, S.-M.; Wang, F.; Yan, S.-Y.; Zhu, Z.-M.; Gao, X.-F.; Zhao, X.-L. Phylogenomics and Plastome Evolution of *Indigofera* (Fabaceae). *Front. Plant Sci.* **2023**, *14*, 1186598. [[CrossRef](#)] [[PubMed](#)]
39. Feng, J.; Wu, L.; Wang, Q.; Pan, Y.; Li, B.; Lin, Y.; Yao, H. Comparison Analysis Based on Complete Chloroplast Genomes and Insights into Plastid Phylogenomics of Four *Iris* Species. *BioMed Res. Int.* **2022**, *2022*, e2194021. [[CrossRef](#)]
40. Zhang, R.; Wang, Y.-H.; Jin, J.-J.; Stull, G.W.; Bruneau, A.; Cardoso, D.; De Queiroz, L.P.; Moore, M.J.; Zhang, S.-D.; Chen, S.-Y.; et al. Exploration of Plastid Phylogenomic Conflict Yields New Insights into the Deep Relationships of Leguminosae. *Syst. Biol.* **2020**, *69*, 613–622. [[CrossRef](#)]
41. Choi, I.-S.; Cardoso, D.; de Queiroz, L.P.; de Lima, H.C.; Lee, C.; Ruhlman, T.A.; Jansen, R.K.; Wojciechowski, M.F. Highly Resolved Papilionoid Legume Phylogeny Based on Plastid Phylogenomics. *Front. Plant Sci.* **2022**, *13*, 823190. [[CrossRef](#)]
42. Moghaddam, M.; Ohta, A.; Shimizu, M.; Terauchi, R.; Kazempour-Osaloo, S. The Complete Chloroplast Genome of *Onobrychis Gaubae* (Fabaceae-Papilionoideae): Comparative Analysis with Related IR-Lacking Clade Species. *BMC Plant Biol.* **2022**, *22*, 75. [[CrossRef](#)] [[PubMed](#)]
43. Xiong, Y.; Xiong, Y.; He, J.; Yu, Q.; Zhao, J.; Lei, X.; Dong, Z.; Yang, J.; Peng, Y.; Zhang, X. The Complete Chloroplast Genome of Two Important Annual Clover Species, *Trifolium Alexandrinum* and *T. Resupinatum*: Genome Structure, Comparative Analyses and Phylogenetic Relationships with Relatives in Leguminosae. *Plants* **2020**, *9*, 478. [[CrossRef](#)] [[PubMed](#)]
44. Lackey, J.A. Cotyledon Areoles in Subtribe Kennediinae (Leguminosae: Phaseoleae). *Aust. J. Bot.* **2008**, *56*, 265–271. [[CrossRef](#)]
45. Bruneau, A. Phylogenetic and Biogeographical Patterns in *Erythrina* (Leguminosae: Phaseoleae) as Inferred from Morphological and Chloroplast DNA Characters. *Syst. Bot.* **1996**, *24*, 587–605. [[CrossRef](#)]
46. Schmieder, R.; Edwards, R. Quality Control and Preprocessing of Metagenomic Datasets. *Bioinformatics* **2011**, *27*, 863–864. [[CrossRef](#)] [[PubMed](#)]
47. Dierckxsens, N.; Mardulyn, P.; Smits, G. NOVOPlasty: De Novo Assembly of Organelle Genomes from Whole Genome Data. *Nucleic Acids Res.* **2017**, *45*, e18. [[CrossRef](#)] [[PubMed](#)]
48. Tillich, M.; Lehwark, P.; Pellizzer, T.; Ulbricht-Jones, E.S.; Fischer, A.; Bock, R.; Greiner, S. GeSeq—Versatile and Accurate Annotation of Organelle Genomes. *Nucleic Acids Res.* **2017**, *45*, W6–W11. [[CrossRef](#)] [[PubMed](#)]
49. Shi, L.; Chen, H.; Jiang, M.; Wang, L.; Wu, X.; Huang, L.; Liu, C. CPGAVAS2, an Integrated Plastome Sequence Annotator and Analyzer. *Nucleic Acids Res.* **2019**, *47*, W65–W73. [[CrossRef](#)] [[PubMed](#)]
50. 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](#)]
51. 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*, 411–422. [[CrossRef](#)]
52. Kurtz, S.; Choudhuri, J.V.; Ohlebusch, E.; Schleiermacher, C.; Stoye, J.; Giegerich, R. REPuter: The Manifold Applications of Repeat Analysis on a Genomic Scale. *Nucleic Acids Res.* **2001**, *29*, 4633–4642. [[CrossRef](#)] [[PubMed](#)]
53. Benson, G. Tandem Repeats Finder: A Program to Analyze DNA Sequences. *Nucleic Acids Res.* **1999**, *27*, 573–580. [[CrossRef](#)] [[PubMed](#)]
54. Mayor, C.; Brudno, M.; Schwartz, J.R.; Poliakov, A.; Rubin, E.M.; Frazer, K.A.; Pachter, L.S.; Dubchak, I. VISTA: Visualizing Global DNA Sequence Alignments of Arbitrary Length. *Bioinformatics* **2000**, *16*, 1046–1047. [[CrossRef](#)] [[PubMed](#)]
55. Darling, A.E.; Mau, B.; Perna, N.T. progressiveMauve: Multiple Genome Alignment with Gene Gain, Loss and Rearrangement. *PLoS ONE* **2010**, *5*, e11147. [[CrossRef](#)] [[PubMed](#)]
56. Katoh, K.; Misawa, K.; Kuma, K.; Miyata, T. MAFFT: A Novel Method for Rapid Multiple Sequence Alignment Based on Fast Fourier Transform. *Nucleic Acids Res.* **2002**, *30*, 3059–3066. [[CrossRef](#)]
57. Nguyen, L.-T.; Schmidt, H.A.; von Haeseler, A.; Minh, B.Q. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Mol. Biol. Evol.* **2015**, *32*, 268–274. [[CrossRef](#)] [[PubMed](#)]
58. Hoang, D.T.; Chernomor, O.; Von Haeseler, A.; Minh, B.Q.; Vinh, L.S. UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Mol. Biol. Evol.* **2018**, *35*, 518–522. [[CrossRef](#)]

- 
59. Kalyaanamoorthy, S.; Minh, B.Q.; Wong, T.K.; Von Haeseler, A.; Jermini, L.S. ModelFinder: Fast Model Selection for Accurate Phylogenetic Estimates. *Nat. Methods* **2017**, *14*, 587–589. [[CrossRef](#)]
  60. Rambaut, A. FigTree V1. 3.1: Tree Figure Drawing Tool. Available online: <http://treebio.ed.ac.uk/software/figtree/> (accessed on 1 May 2023).

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.