horticulturae

MDPI

Article

Characterization and Phylogenetic Analyses of the Complete Chloroplast Genome Sequence in *Arachis* Species

Xiaona Yu †, Tianzhu Liang †, Yi Guo †, Yan Liang, Xiaoxia Zou, Tong Si, Yu Ni and Xiaojun Zhang *

Shandong Peanut Industry Collaborative Innovation Center/Shandong Provincial Key Laboratory of Dryland Farming Technology, College of Agronomy, Qingdao Agricultural University, Qingdao 266109, China; dtyxn@hotmail.com (X.Y.); tianzhu_liang@163.com (T.L.); gy990222@163.com (Y.G.); ly13659417096@163.com (Y.L.); xxzou@qau.edu.cn (X.Z.); nmst12@163.com (T.S.); nmniyu@qau.edu.cn (Y.N.)

* Correspondence: zhangxj@qau.edu.cn; Tel.: +86-0532-58957342; Fax: +86-0532-58957447

† These authors contributed equally to this work.

Abstract: Peanut is an important oilseed and a widely cultivated crop worldwide. Knowledge of the phylogenetic relationships and information on the chloroplast genomes of wild and cultivated peanuts is crucial for the evolution of peanuts. In this study, we sequenced and assembled 14 complete chloroplast genomes of *Arachis*. The total lengths varied from 156,287 bp to 156, 402 bp, and the average guanine–cytosine content was 36.4% in 14 *Arachis* species. A total of 85 simple sequence repeats (SSRs) loci were detected, including 3 dinucleotide and 82 polynucleotide SSRs. Based on 110 complete chloroplast genomes of *Arachis*, a phylogenetic tree was constructed, which was divided into two groups (I and II). A total of 79 different genes were identified, of which six double-copy genes (*ndhB*, *rpl2*, *rpl23*, *rps7*, *ycf1*, and *ycf2*) and one triple-copy gene (*rps12*) are present in all 14 *Arachis* species, implying that these genes may be critical for photosynthesis. The dN/dS ratios for four genes (*rps18*, *accD*, *clpP*, *ycf1*) were larger than 1, indicating that these genes are subject to positive selection. These results not only provided rich genetic resources for molecular breeding but also candidate genes for further functional gene research.

Keywords: *Arachis*; chloroplast gene; chloroplast genome; SSR; phylogeny

Citation: Yu, X.; Liang, T.; Guo, Y.; Liang, Y.; Zou, X.; Si, T.; Ni, Y.; Zhang, X. Characterization and Phylogenetic Analyses of the Complete Chloroplast Genome Sequence in *Arachis* Species. *Horticulturae* **2024**, *10*, 464. https://doi.org/10.3390/horticulturae10050464

Academic Editor: Adriana F. Sestras

Received: 25 March 2024
Revised: 28 April 2024
Accepted: 29 April 2024
Published: 1 May 2024

1. Introduction

The chloroplast has two layers of cell membrane and performs photosynthesis [1,2]. In an ancient entrapment event, cyanobacteria were entrapped and engulfed by early eukaryotic cells, which then became endosymbionts [3,4], and during the long evolutionary history of green plants, there was a massive transfer of cyanobacterial genes into the nuclear genome [5]. The study of the chloroplast genome began in the 1950s, with the first detection of DNA and RNA in the chloroplasts of some higher plants [6]. The first structural features of the chloroplast were revealed in 1986 when the complete sequence of chloroplast DNA from liverworts and tobacco [7,8] was published. After that, more and more chloroplast sequences of economic crops such as soybean [9], rice [10], wheat [11], maize [12], sesame [13], and cotton [14–16] were released. The analysis of the complete chloroplast sequence of plants can effectively support the study of the origin and evolution of plants.

To date, next generation sequencing (NGS) has developed very rapidly, resulting in tremendous improvements and enhancement in cost reduction, high throughput, capability, and applications [17,18]. High-throughput DNA sequencing technologies have increased the amount of genomic data available, and genome sequences are widely used to determine evolutionary patterns and phylogenetic relationships [19]. Chloroplast genomes, which have an independent genetic system, are maternally inherited in most angiosperms and do not exhibit meiotic recombination, making them suitable for studies of phylogenetics, population genetics, molecular evolution, and genome evolution [2,20–27]. The chloroplast

genome is 120~160 kb in size and contains 100~120 highly conserved genes [28], which contain two single-copy regions (LSC and SSC) separated by two copies of inverted repeat (IR) regions [29]. The chloroplast genome has highly conserved gene content, slow molecular evolution, and a low recombination rate, making it an ideal material for species authentication and phylogenetic studies [30–32].

Cultivated peanut (*Arachis hypogaea* L.) is an important oil and economic crop widely cultivated in tropical and subtropical regions (annual production of ~46 million tons) [33]. Cultivated peanut is an allotetraploid (2n = 4x = 40) resulting from a cross between the two wild diploids *A. duranensis* (AA genome) and *A. ipaensis* (BB genome) [34–37]. Peanut has a relatively complex evolution, and genomic analysis suggests that the lineage has been affected by at least three polyploidizations since the origin of eudicots [38]. Genomic in situ hybridization suggests that *A. monticola* may be the direct wild ancestor of *A. hypogaea* [34]. After long-term artificial domestication and historical selection, the cultivated peanut has a relatively narrow genetic base [39]. Studying the genetic relationships between cultivated and wild peanuts is important not only to understand the evolution of peanuts and effectively utilize the abundant resources of wild species, but also to transfer excellent genes of wild peanuts into cultivated peanuts, which provide a theoretical basis for molecular breeding.

The complete chloroplast genome sequence of different Arachis species has been published and is an important reference for phylogenetic and comparative analyses [40]. The chloroplast genome sequence of the four major peanut varieties (var. *hypogaea*, var. *hirsuta*, var. *fastigiate*, and var. *vulgaris*) showed that the gene contents and orders were highly conserved [41]. Through the six peanut varieties chloroplast genome, it was found that they have a single genetic origin and that *A. monticola* was the immediate tetraploid ancestor from which *A. hypogaea* emerged during domestication [42]. The reported chloroplast genome offered a wealth of genetic information for the improvement of peanuts and also contributed to a better understanding of the evolutionary relationships between wild and cultivated plants [29,43]. In this study, we assembled 14 chloroplast genomes of *Arachis*, including both cultivated and wild peanut species. Through comparative analysis with 96 other peanut chloroplast genomes available in NCBI, we aim to gain insights into the genetic diversity of Arachis and identify the potential maternal genome progenitors of cultivated peanuts. This study enriches the genetic information of the chloroplast genome of *Arachis* and provides a theoretical basis for species classification.

2. Result

2.1. Basic Characteristics of the Acquired Arachis Chloroplast Genomes

A total of 14 sequenced chloroplast genomes of *Arachis* species showed a typical quadripartite structure (Table 1 and Figure 1), and the total lengths varying from 156,287 bp (Xiaohongmao) to 156, 402 bp (Ba-1) (Figure 1, Table 1). The 14 chloroplast genomes of *Arachis* differ only slightly in the number of genes and total proteins, which are all between 88 and 91 (Table 2). The guanine–cytosine (GC) content of the chloroplast genomes of all 14 *Arachis* species was 36.4% and revealed a high extent of similarity (Table 2). A total of 85 simple sequence repeats (SSR) loci were detected in the chloroplast genomes of all 14 *Arachis* species, including 3 dinucleotide and 82 polynucleotide SSRs. We found that the content of A/T in the SSRs was higher than that of C/G (Table S1).

Table 1. The *Arachis* species information has been analyzed in the present study.

	Species	Variety Name	Genome Type	Ploidy
Domesticated varieties				
	A. hypogaea	Luhua11	AABB	4X
	A. hypogaea var. *fastigiata*	Silihong	AABB	4X
	A. hypogaea var. *vulgaris*	Baisha1016	AABB	4X
	A. hypogaea var. *peruviana*	Yunnanqicai	AABB	4X
	A. hypogaea var. *aequatoriana*	Chidao1	AABB	4X
	A. hypogaea cv. *tifrunner*	Tifrunner	AABB	4X
	A. hypogaea var. *hypogaea*	Xiaohongmao	AABB	4X
Wild allotetraploid species				
	A. monticola	Monticola	AABB	4X
Wild diploid species				
	A. batizocoi	Ba-1	KK	2X
	A. ipaensis	Ip-1	BB	2X
	A. duranensis	Ad-1	AA	2X
	A. duranensis	Ad-2	AA	2X
	A.stenosperma	St-1	AA	2X
	A. correntina	Correntina	AA	2X

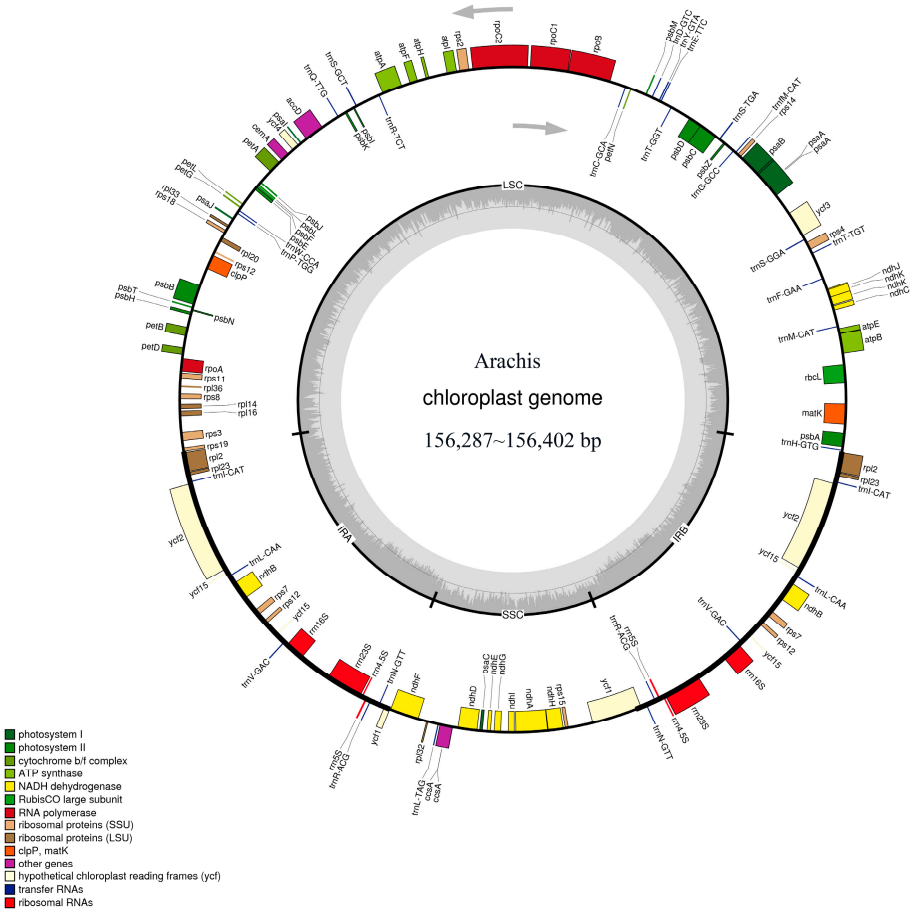

Figure 1. Map of the *Arachis* chloroplast genomes. Genes outside the outer circle are transcribed counterclockwise, while genes inside the circle are transcribed clockwise. Genes belonging to different functional groups are color-coded. The gray area in the inner circle indicates the GC content of the chloroplast genome. The four regions of a chloroplast genome are also indicated in the inner circle: the two inverted repeat regions (Ira, IRb, SSC, and LSC).

Table 2. Details of the complete chloroplast genomes of 14 *Arachis* species.

Species	Variety Name	Raw Reads	Genome Size (bp)	Gene Number	GC Content (%)	Total Protein	LSC (bp)	SSC (bp)	IR (bp)	rRNA	tRNA
M1	Luhua11	934	156,359	88	36.4	88	85,910	18,787	25,831	8	43
M2	Silihong	1006	156,391	88	36.4	88	85,913	18,794	25,842	8	43
M3	Baisha1016	897	156,355	88	36.4	88	85,906	18,769	25,840	8	43
M4	Yunnanqicai	842	156,395	88	36.4	88	85,918	18,789	25,844	8	43
M5	Tifrunner	930	156,395	88	36.4	88	85,924	18,803	25,834	8	43
M6	Monticola	1053	156,395	88	36.4	88	85,924	18,803	25,834	8	43
M7	Ip-1	936	156,399	90	36.4	90	85,938	18,793	25,834	8	43
M8	Ad-1	1323	156,343	91	36.4	91	85,902	18,805	25,818	8	29
M9	Ba-1	1807	156,402	91	36.4	91	85,922	18,812	25,834	8	29
M10	St-1	1797	156,303	91	36.4	91	85,853	18,804	25,823	8	29
M11	Ad-2	2122	156,359	91	36.4	91	85,953	18,760	25,823	8	29
M12	Correntina	1934	156,373	91	36.4	91	85,930	18,797	25,823	8	29
M13	Chidao1	1553	156,373	91	36.4	91	85,930	18,797	25,823	8	29
M14	Xiaohongmao	1789	156,287	91	36.4	91	85,843	18,798	25,823	8	29

2.2. Phylogenetic Analysis

A total of 110 complete chloroplast genomes from 19 different *Arachis* varieties were utilized to construct a phylogenetic tree using the maximum likelihood method (Figure 2; Table S4). The phylogenetic tree comprised 14 genomes obtained in this study and an additional 96 genomes obtained from NCBI. The resulting phylogenetic tree showed two major groups, Group I and Group II, which encompassed a total of 39 and 71 *Arachis* chloroplast genomes, respectively. The cultivated peanuts with genome type AABB and the species with genome type AA (*A. duranensis*) are mainly distributed in Group I. Baisha1016 belongs to *A. hypogaea* var. *vulgaris*; Silihong belongs to *A. hypogaea* var. *fastigiata*; Yunnanqicai belongs to *A. hypogaea* var. *peruviana*; Chidao1 belongs to *A. hypogaea* var. *aequatoriana*; Xiaohongmao and Tifrunner belong to *A. hypogaea* var. *hypogaea*. They are belonging to *A. hypogaea*, which gathers with *A. hypogaea* in Group I. While species in Group II had different genome types, including AABB, BB, KK, EE, etc. For example, *A. duranensis*, *A. monticola*, and the cultivated peanut are close to each other in clades of Group II. *A. hoehnei* (BB) showed a close relationship with *A. cardenasii* (AA) and *A. diogoi* (AA), while *A. ipaensis*, another possible diploid ancestor of the cultivated peanut, is grouped closely with *A. batizocoi* (KK). Other varieties are summarized in a large clade, and it is impossible to draw a clear boundary between them (Figure 2, Table S6).

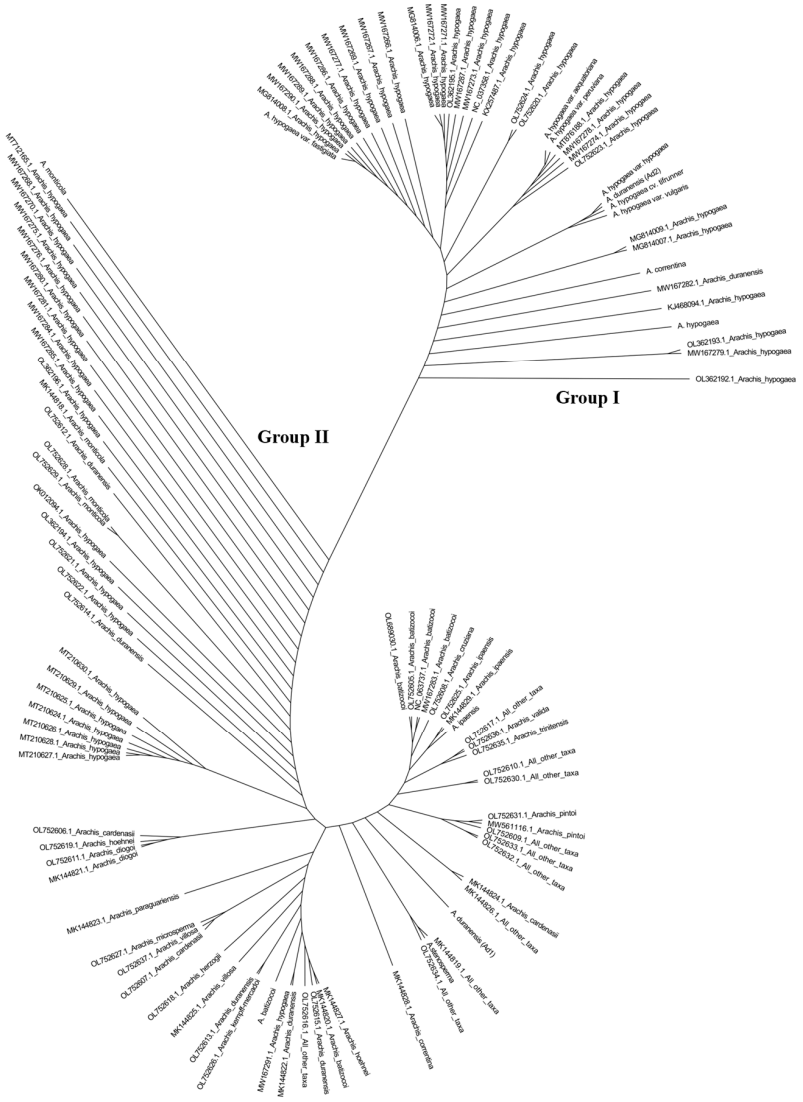

Figure 2. The phylogenetic tree of the 110 *Arachis* complete chloroplast genomes.

2.3. Information on Conserved and Variable Genes in the Arachis Chloroplast Genome

On the analysis of the conservation and variability of chloroplast-related genes, a total of 130 genes (16 genes have 2 copies in the chloroplast genome) were recorded, including 73 conserved genes, 17 synonymous mutation genes, and 40 amino acid mutation genes, which included photosynthesis genes, translation-related genes, RNA genes, etc. Most genes were found in all *Arachis* but *lhbA* was present only in Silihong, Baisha1016, Yunnanqicai, Tifrunner, and Monticola; the *psbZ* gene was present only in Luhua11, Ip-1, Ad-1, Ba-1, St-1, Ad-2, Correntina, Chidao1, and Xiaohongmao; *psbB* was present in Luhua11, Ip-1, Ad-1, Ba-1, St-1, Ad-2, Correntina, Chidao1, and Xiaohongmao; and *orf42* and *orf56* were present only in Luhua11, Silihong, Baisha1016, Yunnanqicai, Tifrunner, Monticola, and Ip-1; *petL* and *petN* are present in Ad-1, Ba-1, St-1, Ad-2, Correntina, Chidao1, and Xiaohongmao; *psaA* has a deletion in Ad-1, Ba-1, St-1, Ad-2, Correntina, Chidao1, and Xiaohongmao. *Ycf68* is present in Luhua11, Silihong, Baisha1016, Yunnanqicai, Tifrunner, Monticola, Ip-1, and *ycf15* is present in Ad-1, Ba-1, St-1, Ad-2, Correntina, Chidao1, and Xiaohongmao, but it not conserved. In addition, we found that 12 conserved genes (*rps12*, *rpl23*, *rrn16*, *rrn5*, *rrn4.5*, *trnAUGC*, *trnICAU*, *trnIGAU*, *trnLUAA*, *trnNGUU*, *trnRACG*, *trnVGAC*), 1 synonymous mutation gene (*rps7*), and 3 amino acid mutation genes (*rpl2*, *rrn23*, *ycf2*) had 2 copies in the chloroplast gene. *A. ipaensis* has synonymous mutations in *petA* and *A. stenosperma* has amino acid mutations. The result showed the diversity of *Arachis* chloroplast genes (Table 3 and Table S3).

Table 3. List of conserved and variable genes in *Arachis* chloroplast genomes.

Gene Categories	Conserved Gene	Synonymous Mutations	Amino Acid Mutations
Photosystem I	*psaC*, *psaI*, *psaJ*	*psaB*	*psaA*
Photosystem II	*psbC*, *psbE*, *psbF*, *psbH*, *psbI*, *psbJ*, *psbK*, *psbL*, *psbM*, *psbN*, *psbZ*, *psi_psbT*	*psbA*, *psbB*, *psbD*, *psbT*	
RuBisCO large subunit		*rbcL*	
Cytochrome b/f complex	*petL*, *petN*	*petB*, *petG*	*petA*, *petD*
c-type cytochrome			*ccsA*
ATP synthase	*atpB*, *atpE*,	*atpI*	*atpA*, *atpF*, *atpH*
NADH dehydrogenase	*ndhI*, *ndhJ*	*ndhD*, *ndhE*	*ndhA*, *ndhB*, *ndhC*, *ndhF*, *ndhG*, *ndhH*, *ndhK*
Assembly/stability of photosystem I			*ycf3*, *ycf4*
RNA polymerase genes			*rpoA*, *rpoB*, *rpoC1*, *rpoC2*
Ribosomal protein	*rps12* *, *rps14*, *rps18*, *rpl23* *, *rpl32*, *rpl33*	*rps4*, *rps7* *, *rps11*, *rpl14*, *rpl20*,	*rps2*, *rps3*, *rps8*, *rps15*, *rps19*, *rpl2* *, *rpl16*, *rpl36*
Ribosomal RNA	*rrn16* *, *rrn5* *, *rrn4.5* *		*rrn23* *
Transfer RNA	$trnA_{UGC}$ *, $trnC_{GCA}$, $trnD_{GUC}$, $trnE_{UUC}$, $trnF_{GAA}$, $trnG_{GCC}$, $trnG_{UCC}$, $trnH_{GUG}$, $trnI_{CAU}$ *, $trnI_{GAU}$ *, $trnK_{UUU}$, $trnL_{CAA}$, $trnL_{UAA}$ *, $trnL_{UAG}$, $trnM_{CAU}$, $trnN_{GUU}$ *, $trnQ_{UUG}$, $trnR_{ACG}$ *, $trnR_{UCU}$, $trnS_{GCU}$, $trnS_{GGA}$, $trnS_{UGA}$, $trnT_{GGU}$, $trnT_{UGU}$, $trnV_{GAC}$ *, $trnW_{CCA}$, $trnY_{GUA}$		$trnM_{CAU}$, $trnP_{UGG}$, $trnW_{CCA}$
Acetyl-CoA carboxylase subunit			*accD*
Proteolysis subunit			*clpP*
Carbon metabolism	*cemA*		
Maturase			*matK*
Conserved reading frames	*ycf68*, *orf56*, *orf42*		*ycf1*, *ycf2* *

Note: * These genes have 2 copies in chloroplast genome.

2.4. Arachis Chloroplast Genomes Have Diversity

We find that *petB* has a deletion mutation in the 5′ terminal; *rpoC1* in Luhua11, Ip-1, Ad-1, Ba-1, St-1, Ad-2, and Correntina has a deletion in the 5′ terminal in; *rps12* in Luhua11, Silihong, Baisha1016, Yunnanqicai, Tifrunner, Monticola, and Ip-1 has a deletion in the 5′ terminal; *rps18* in Ad-1, Ba-1, St-1, Ad-2, Correntina, Chidao1, and Xiaohongmao has a deletion in the 3′ terminal. The sequence result showed that *psbT* might have a deletion in Ad-1, Ba-1, St-1, Ad-2, Correntina, Chidao1, and Xiaohongmao because there are two TAG in the 5′ terminal; *rpl2* lost three nucleic acids in 391–393 loci. *Rpl2* of Ad-1 lost one long fragment after 391 loci. *Ycf1* in Xiaohongmao lost only a partial fragment, only 1221 bp. Ip-1 lost three nucleic acids in 3007–3009 loci, only 3159 bp; Ad-1, Ba-1, St-1, Ad-2, Correntina, Chidao1, and Xiaohongmao has lost 1403 bp in the 3′ terminal compared with Luhua11, Silihong, Baisha1016, Yunnanqicai, Tifrunner, and Monticola. *Ycf2* In Luhua11, Baisha1016, and Tifrunner, there is 285 bp longer than others in the 5′ terminal. Luhua11 lost one fragment from 5302 to 5319, and Baisha1016 lost one fragment from 5293 to 5310. *Ycf3* in Ad-2 and Xiaohongmao lost 270 bp (total 501 bp) in the 3′ terminal (Table 4).

Table 4. The positive selection sites of *Arachis* chloroplast genes.

Gene Categories	Gene	M1 vs. M2	M7 vs. M8
Photosystem I	*psaA*	None	16 L
Cytochrome b/f complex	*petA*	None	176 V
	petD	None	137 P
c-type cytochrome	*ccsA*	None	61 I, 121 I, 284 F
ATP synthase	*atpA*	391 S	391 S
	atpF	1 S, 2 F, 3 S, 4 F, 5 G, 6 F, 7 N, 8 T, 9 D, 10 I, 11 L, 12 A	1 S, 2 F, 3 S, 4 F, 5 G, 6 F, 7 N, 8 T, 9 D, 10 I, 11 L, 12 A
	atpH	11 V	11 V
NADH dehydrogenase	*ndhA*	185 L	185 L
	ndhB	4 E, 5 M, 6 A, 7 L, 8 T, 10 F, 11 L, 13 F, 14 Y, 15 N, 16 S, 20 P, 21 D, 22 Y, 24 G	4 E, 5 M, 6 A, 7 L, 8 T, 10 F, 11 L, 13 F, 14 Y, 15 N, 16 S, 20 P, 21 D, 22 Y, 24 G
	ndhF	21 L, 186 L, 332 M, 476 Y, 490 N, 582 L, 586 S, 601 Q, 689 F	21 L, 186 L, 332 M, 476 Y, 490 N, 582 L, 586 S, 601 Q, 689 F
	ndhG	30 T, 166 A	30 T, 166 A
	ndhH	292 I, 301 P	292 I, 301 P
	ndhK	2 S, 6 L, 8 P, 10 P, 11 K, 12 Y, 13 V, 15 A, 16 M, 18 A, 19 C, 22 T, 25 M, 26 F, 29 D, 30 S, 31 Y, 33 P, 34 G, 35 C, 36 P, 37 P, 41 A, 44 D, 48 T, 51 K, 52 K, 53 Y, 54 K, 55 K	1 P, 2 S, 6 L, 8 P, 10 P, 11 K, 12 Y, 13 V, 15 A, 16 M, 18 A, 19 C, 20 T, 21 I, 22 T, 24 G, 25 M, 26 F, 27 S, 29 D, 30 S, 31 Y, 32 L, 33 P, 34 G, 35 C, 36 P, 37 P, 38 K, 40 E, 41 A, 44 D, 45 A, 47 T, 48 T, 51 K, 52 K, 53 Y, 54 K, 55 K
Assembly/stability of photosystem I	*ycf3*	40 R, 41 D, 43 M, 77 N	40 R, 77 N
	ycf4	None	3 W, 118 I
RNA polymerase genes	*rpoA*	None	111 N, 133 T, 234 A, 269 L
	rpoB	None	44 L, 210 D, 646 F
	rpoC1	1 F, 3 I, 4 D, 5 P, 6 L, 9 S, 11 P, 12 N, 449 K	1 F, 3 I, 4 D, 5 P, 6 L, 9 S, 11 P, 12 N, 449 K
	rpoC2	430 L, 469 P, 634 P, 660 E, 675 L, 697 K, 773 L, 824 H, 912 K, 996 S, 998 E, 1000 L, 1001 K, 1002 G, 1003 K, 1004 L, 1013 L, 1014 K, 1015 K, 1017 C, 1193 I, 1335 K	430 L, 469 P, 634 P, 660 E, 675 L, 697 K, 773 L, 824 H, 912 K, 996 S, 998 E, 1000 L, 1001 K, 1002 G, 1003 K, 1004 L, 1013 L, 1014 K, 1015 K, 1017 C, 1193 I, 1335 K

Table 4. *Cont.*

Gene Categories	Gene	M1 vs. M2	M7 vs. M8
Ribosomal protein	*rpl2*	131 N, 133 G, 134 V, 135 N, 138 E, 139 G, 140 R, 141 A, 143 I, 144 K, 146 A, 147 T	131 N, 133 G, 134 V, 135 N, 138 E, 139 G, 140 R, 141 A, 143 I, 144 K, 146 A, 147 T
	rpl16	104 M, 126 Q	104 M, 126 Q
	rpl36	24 L	24 L
	rps2	None	198 N
	rps3	153 Q	153 Q
	rps8	None	28 C
	rps15	18 N	18 N
	rps19	1 K, 2 K	1 K, 2 K
Acetyl-CoA carboxylase subunit	*accD*	4 G, 50 P, 119 L, 199 G, 286 M, 399 N	4 G, 50 P, 119 L, 199 G, 286 M, 399 N
Proteolysis subunit	*clpP*	1 I, 99 R	1 I, 99 R
Conserved reading frames	*ycf1*	162 K, 181 F, 226 V, 233 D, 241 F, 242 K, 257 H, 264 I, 301 A, 309 K, 349 T, 362 S, 371 Q, 379 S, 405 L, 406 S, 407 N	162 K, 181 F, 226 V, 233 D, 241 F, 242 K, 257 H, 264 I, 301 A, 309 K, 349 T, 362 S, 371 Q, 379 S, 405 L, 406 S, 407 N
	ycf2	1293 K	169 W, 531 S, 532 E, 538 N, 751 H, 1206 R, 1223 L, 1292 W, 1293 K, 1294 T

2.5. The Selective Pressure of Arachis Chloroplast Genes Using Codeml

A total of 53 chloroplast genes were selected to test the selection pressure. The dN/dS ratios for 49 genes are less than 1, indicating functional conservation of the gene during evolution and suggesting that these genes were subject to purifying selection. The dN/dS ratios for 4 genes (*rps18*, *accD*, *clpP*, *ycf1*) are greater than 1, indicating that these genes were subject to positive selection, suggesting that these genes evolved at a high rate and thus may play a crucial role in the evolution of Arachis species (Table S2). In addition, 36 genes have Ts/Tv ratios above 1, indicating that they are more frequent than transversions, and 17 genes have Ts/Tv ratios below 1, indicating that transversions are more frequent than transitions (Table S2).

2.6. The Replication of Chloroplast Genes

Gene replication occurred in 14 *Arachis*, with six double-copy genes (*ndhB*, *rpl2*, *rpl23*, *rps7*, *ycf1*, and *ycf2*) and one triple-copy gene (*rps12*) present in all 14 *Arachis*. The copy number of many chloroplast genes varied between individuals of the different *Arachis*. *NdhK* and *psaA* were present in two copies in *Ad-1*; *orf42*, *orf56*, and *ycf68* were present in three copies in Luhua11, Silihong, Baisha1016, Yunnanqicai, Tifrunner, Monticola, and Ip-1; *ndhK* and *psaA* occurred in two copies in Ad-1, Ba-1, St-1, Ad-2, *Correntina*, Chidao1, and Xiaohongmao; ycf15 occurred in two copies in Ip-1 but in four copies in Ad-1, Ba-1, St-1, Ad-2, *Correntina*, Chidao1, and Xiaohongmao (Table S3). The result shows that these genes may play a crucial role in photosynthesis.

3. Discussion

The chloroplast genomes were highly conserved, there was no recombination, and they were compact in size and maternally inherited, which led to a better understanding of the origin and genetic resources [44–46]. In this study, all chloroplast genomes exhibited the classical quadripartite structure, which typically contains an LSC range from 85,843 bp to 85,953 bp in size, an SSC range from 18,760 bp to 18,812 bp in size, and are separated by two IR 25,818 bp to 25,844 bp. The same structure and number had also been reported in other *Arachis* species [29,39,41]. The GC content in the present study was 36.4%, which was also consistent with previous results of 36.3% to 36.4% [29,40,41,43,47]. Although the size and structure of the chloroplast genomes of Arachis are conserved, there was a wealth of genetic information, including SSRs, Indels, and SNPs in the chloroplast genomes of

Arachis. Repetitive sequences (minisatellite, microsatellite, and satellite sequences) are very common in sequence and copy number during evolution and therefore play an important role in taxonomic and phylogenetic studies. A total of 101 and 69 SSR loci were identified by Yin et al. and Wang et al., respectively [29,41], and 85 SSR loci were found in the present study, so these results can effectively provide genetic markers to elucidate the complex evolutionary history of *Arachis*.

During evolution, many ancient chloroplast genes were translocated from the chloroplast genome to the cell nucleus, but the proteins important for photosynthesis remained in the chloroplast genome [48]. On average, the chloroplast genomes of land plants have retained about 120 genes with conserved content [49]. In the present study, a total of 79 different genes were obtained and we compared and analyzed the gene content, genomic organization, and RNA editing sites of 14 representative Arachis chloroplast genomes. The results showed that the Arachis chloroplast has a relatively conservative gene content, but there are significant differences among chloroplast genes in terms of deficiency and mutation. The evolutionary rate ratio dN/dS is often used to infer selection pressure in protein-coding genes, i.e., the ratio of non-synonymous to synonymous substitution rates [50]. This ratio indicates how quickly the amino acids that make up a protein change relative to synonymous changes, and it is often used to identify protein sites subject to purifying selection ($dN/dS < 1$), neutral evolution ($dN/dS \approx 1$) or positive diversifying selection ($dN/dS > 1$) [51,52]. We tested the selection pressure for 53 chloroplast genes and found that 4 of them are above 1, of which assembly/stability of photosystem I "ycf1" occurred in three copies, implying that it may play a key role in photosynthesis, so it probably has potential research value (Table S2). In addition, we found two genes "*psbZ*" and "*psbB*" that played a crucial role in photosynthesis and are only found in wild Arachis, they might have disappeared in long-term evolution and domestication or transferred to the nucleus (Table S3). The 14 chloroplast genomes obtained in the present study not only help us to better understand the genetic and phylogenetic relationships between wild and cultivated Arachis species but also provide a wealth of genetic resources for peanut breeding.

It is generally accepted that *A. hypogaea* has been divided into two subspecies ssp. *fastigiate* and ssp. *hypogaea*, of which four botanical varieties var. *fastigiate*, var. *vulgaris*, var. *peruviana*, and var. *aequatoriana* belong to the var. *fastigiate*, and two botanical varieties var. *hypogaea*, and var. *hirsute* belong to the ssp. *Hypogaea* [53]. *Arachis* classification mainly depended on morphological characteristics and was not consistently supported by work at the molecular level when different methods or genetic markers were used [33,34]. Wild species exhibited much greater genetic diversity and provided a large pool of genetic diversity from which new allelic variation can be obtained for breeding programs [54,55]. We selected 7 different *Arachis* including var. *fastigiate* (Silihong), var. *vulgaris* (Baisha1016), var. *peruviana* (Yunnanqicai), var. *aequatoriana* (Chidao1), var. *hypogaea* (Tifrunner and Xiaohongmao) and sequential flowering intermediate type (Luhua11), and 7 various wild types including AA (Ad-1, St-1, Ad-2, Correntina), BB (Ip-1), KK (Ba-1), AABB (Monticola), then the chloroplast genomes of 14 *Arachis* were sequenced and assembled in present study. Based on our phylogenetic analysis with its genome type information, hybridization appears to play an important role in the evolutionary history of cultivated species. *A. monticola*, a wild tetraploid species, was clustered in the *Arachis* complex group II near the cultivated peanuts and may represent a transitional species that underwent the most recent hybridization event (Figure 2). Var. *fastigiate* was closely related to var. *hypogaea* in Group I, which supported a close relationship between them that differs from what we would expect based on the previous classification by high-quality SNP from genome sequencing, in which the phylogenetic tree exhibited that they were clustered into two groups [56,57]. It appears that nuclear genome sequence data are insufficient or unreliable to interpret the evolutionary relationship between allotetraploid species. In addition, we found that there are some individual species embedded in *A. hypogaea* according to the phylogenetic tree, such as *Correntina* and Ad-2, which we sequenced, although they belong to the wild species that cluster with *A. hypogaea* in Group I, which means that

there are similar genomes between wild species and cultivated species in the chloroplasts of *Arachis*, showing that wild species and cultivated species are also genetically closely related. Due to the maternally inherited and highly conservative characteristics of the chloroplast, it should be used for genetic relationships. We aligned the sequence of the chloroplast genomes of cultivated and wild peanuts and found that the identity value between Luhua11 (*A. hypogaea*) and Ad-2 (*A. duranensis*) was the highest at 41.46%. This result is consistent with the previous view that the wild diploid *A. duranensis* is one of the parents of the cultivated peanut, indicating that Ad-2 served as the maternal donor of the cultivated peanut. The 14 chloroplast genomes obtained in the present study provide a wealth of genetic resources for peanut breeding.

The genus Arachis consists of 81 species with a wide variety of genome types, including AA, BB, AABB, CC, DD, EE, EEXX, FF, HH, KK, PR, RR1, RR2, TT, and TTEE. Since not all Arachis genome types were included in this study, it is limited to understanding the origin and evolution of cultivated peanuts. Future work should attempt to collect germplasm resources of a variety of genome types from different geographic regions to provide a better understanding of the taxonomic status of the different Arachis species and the evolutionary relationships between them.

4. Materials and Methods

4.1. Plant Material and DNA Extraction

A total of 14 *Arachis* species (*A. hypogaea* (Luhua11), *A. hypogaea* var. *fastigiata* (Silihong), *A. hypogaea* var. *vulgaris* (Baisha1016), *A. hypogaea* var. *peruviana* (Yunnanqicai), *A. hypogaea* var. *aequatoriana* (Chidao1), *A. hypogaea* cv. *tifrunner* (Tifrunner), *A. hypogaea* var. *hypogaea* (Xiaohongmao), *A. monticola* (Monticola), *A. batizocoi* (Ba-1), *A. ipaensis* (Ip-1) *A. duranensis* (Ad-1), *A. duranensis* (Ad-2), *A.stenosperma* (St-1), *A. correntina* (Correntina) were grown in the greenhouse. Fresh leaves were collected for total genomic DNA isolation using the SteadyPure Plant Genomic DNA Extraction Kit (Accurate Biotechnology, Changsha China), and the DNA concentration was quantified using a NanoDrop (Thermo Scientific, Waltham, MA, USA).

4.2. Genome Assembly and Annotation

Each DNA sample was randomly fragmented, and the target amplicon fragment was repaired, then subjected to blunt end repair and phosphorylation with T4 DNA polymerase, Klenow DNA polymerase, and T4-PNK. A-tailing was then inserted at the 3′-ends. The adaptors were ligated with base "T" at the 3′ end of the DNA fragments using T4 DNA ligase. Subsequently, the qualified libraries were used for cluster preparation, and sequencing by synthesis was performed on the Illumina Hiseq platform using the Truseq v3-HS Kit (Illumina Inc., San Diego, CA, USA). Estimated genome size using K-mer statistical analysis methods and assembled with clean data from SOAPdenovo 2.04 software [58], then after paired-end from reads relationships of overlay, the assembly results were partially assembled and optimized. Finally, remove redundant segment sequences to obtain the final assembly result from GapCloser 1.12 software. Scattered repetition repeated sequences were calculated using RepeatMasker 3.30 software [59], and tandem repeat (TR) was calculated using TRF 4.04 software [60], with the result plotted using sigmaplot 13.

DOGMA software [61] was used to perform component analysis of the sample genome. The identity value for the prediction of coding proteins was set to 40, and other parameters were default values to obtain the prediction results of coding genes of the sample genome and non-coding RNA. Homologous comparison methods (BLAST) were used for gene function labeling [62], and the database of generic functional annotations for prokaryotes includes the Non-Redundant Protein Database (NR), Kyoto Encyclopedia of Genes and Genomes (KEGG) [63,64], Cluster of Orthologous Groups of proteins (COG) [65,66], Gene Ontology (GO) [67,68], Swiss-Prot and TrEMBL [69].

4.3. Phylogenetic Analysis

A total of 110 complete chloroplast genomes of *Arachis* (96 from NCBI and 14 from present study) were used to construct a phylogenetic tree. To do this, an alignment was first performed and then a cutoff of 10% was set using CLC Genomics Workbench.

Maximum likelihood analysis was performed using IQ-TREE with 1000 bootstrap replicates, and the result was displayed using FigTree 1.4.4.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/horticulturae10050464/s1, Table S1. Analyses of simple sequence repeat (SSR) in the *Arachis* chloroplast genomes; Table S2. The selective pressure of peanut chloroplast gene using codeml; Table S3. The statistic of peanut chloroplast genes; Table S4. The information of reported 96 *Arachis* complete chloroplast genomes from NCBI; Table S5. GO annotations of 14 chloroplast genomes in the present study: Table S6. The sequence of 110 chloroplast genomes.

Author Contributions: X.Y.: conceptualization, methodology, software, validation, data curation, writing—review and editing, supervision. T.L.: conceptualization, methodology, software, data curation, validation, writing—original draft preparation, writing—review and editing, visualization. Y.G.: conceptualization, methodology, software, data curation, writing—original draft preparation, writing-review and editing, visualization. Y.L.: data curation, validation, writing—review and editing. Y.N.: validation, data curation. X.Z. (Xiaoxia Zou): validation, data curation, writing—review and editing, supervision. T.S.: validation, data curation, writing—review and editing. X.Z. (Xiaojun Zhang): supervision, writing—review and editing, project administration, funding acquisition. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by grants from the National Science Foundation of China (No. 32372133), the Seed Project of Shandong Province, China (No. 2020LZGC001).

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors on request.

Conflicts of Interest: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. Gao, L.L.; Hong, Z.H.; Wang, Y.; Wu, G.Z. Chloroplast proteostasis: A story of birth, life, and death. *Plant Commun.* **2023**, *4*, 100424. [CrossRef] [PubMed]
2. Daniell, H.; Lin, C.S.; Yu, M.; Chang, W.J. Chloroplast genomes: Diversity, evolution, and applications in genetic engineering. *Genome Biol.* **2016**, *17*, 134. [CrossRef] [PubMed]
3. Maréchal, E. Primary Endosymbiosis: Emergence of the Primary Chloroplast and the Chromatophore, Two Independent Events. *Methods Mol. Biol.* **2018**, *1829*, 3–16.
4. Bonnett, H.T. On the mechanism of the uptake of Vaucheria chloroplasts by carrot protoplasts treated with polyethylene glycol. *Planta* **1976**, *131*, 229–233. [CrossRef]
5. Howe, C.J.; Barbrook, A.C.; Koumandou, V.L.; Nisbet RE, R.; Symington, H.A.; Wightman, T.F. Evolution of the chloroplast genome. *Philos. Trans. R. Soc. B-Biol. Sci.* **2003**, *358*, 99–107. [CrossRef] [PubMed]
6. Chiba, Y.J.C. Cytochemical Studies on Chloroplasts I:Cytologic demonstration of nucleic acids in chloroplasts. *Cytologia* **1951**, *16*, 259–264. [CrossRef]
7. Ohyama, K.; Fukuzawa, H.; Kohchi, T.; Shirai, H.; Sano, T.; Sano, S.; Umesono, K.; Shiki, Y.; Takeuchi, M.; Chang, Z.J.N. Chloroplast gene organization deduced from complete sequence of liverwort Marchantia polymorpha chloroplast DNA. *Nature* **1986**, *322*, 572–574. [CrossRef]
8. Shinozaki, K.; Ohme, M.; Tanaka, M.; Wakasugi, T.; Hayashida, N.; Matsubayashi, T.; Zaita, N.; Chunwongse, J.; Obokata, J.; Yamaguchi-Shinozaki, K.; et al. The complete nucleotide sequence of the tobacco chloroplast genome: Its gene organization and expression. *EMBO J.* **1986**, *5*, 2043–2049. [CrossRef]
9. Saski, C.; Lee, S.B.; Daniell, H.; Wood, T.C.; Tomkins, J.; Kim, H.G.; Jansen, R.K. Complete chloroplast genome sequence of Gycine max and comparative analyses with other legume genomes. *Plant Mol. Biol.* **2005**, *59*, 309–322. [CrossRef]
10. Hiratsuka, J.; Shimada, H.; Whittier, R.; Ishibashi, T.; Sakamoto, M.; Mori, M.; Kondo, C.; Fan, J.; Zhu, W.Y.; Li, Z.F.; et al. Chloroplast genome sequence of a yellow colored rice (*Oryza sativa* L.): Insight into the genome structure and phylogeny. *Mitochondrial DNA B Resour.* **2020**, *5*, 3650–3652.
11. Skuza, L.; Androsiuk, P.; Gastineau, R.; Paukszto, Ł.; Jastrzębski, J.P.; Cembrowska-Lech, D. Molecular structure, comparative and phylogenetic analysis of the complete chloroplast genome sequences of weedy rye *Secale cereale* ssp. segetale. *Sci. Rep.* **2023**, *13*, 5412. [CrossRef] [PubMed]

12. Maier, R.M.; Neckermann, K.; Igloi, G.L.; Kössel, H. Complete sequence of the maize chloroplast genome: Gene content, hotspots of divergence and fine tuning of genetic information by transcript editing. *J. Mol. Biol.* **1995**, *251*, 614–628. [CrossRef] [PubMed]
13. Yi, D.K.; Kim, K.J. Complete chloroplast genome sequences of important oilseed crop *Sesamum indicum* L. *PLoS ONE* **2012**, *7*, e35872. [CrossRef] [PubMed]
14. Xu, Q.; Xiong, G.; Li, P.; He, F.; Huang, Y.; Wang, K.; Li, Z.; Hua, J. Analysis of complete nucleotide sequences of 12 Gossypium chloroplast genomes: Origin and evolution of allotetraploids. *PLoS ONE* **2012**, *7*, e37128. [CrossRef] [PubMed]
15. Lee, S.B.; Kaittanis, C.; Jansen, R.K.; Hostetler, J.B.; Tallon, L.J.; Town, C.D.; Daniell, H. The complete chloroplast genome sequence of Gossypium hirsutum: Organization and phylogenetic relationships to other angiosperms. *BMC Genom.* **2006**, *7*, 61. [CrossRef] [PubMed]
16. Ibrahim, R.I.; Azuma, J.; Sakamoto, M. Complete nucleotide sequence of the cotton (*Gossypium barbadense* L.) chloroplast genome with a comparative analysis of sequences among 9 dicot plants. *Genes. Genet. Syst.* **2006**, *81*, 311–321. [CrossRef] [PubMed]
17. Alekseyev, Y.O.; Fazeli, R.; Yang, S.; Basran, R.; Maher, T.; Miller, N.S.; Remick, D. A Next-Generation Sequencing Primer-How Does It Work and What Can It Do? *Acad. Pathol.* **2018**, *5*, 2374289518766521. [CrossRef] [PubMed]
18. McCombie, W.R.; McPherson, J.D.; Mardis, E.R. Next-Generation Sequencing Technologies. *Cold Spring Harb. Perspect. Med.* **2019**, *9*, a036798. [CrossRef] [PubMed]
19. Carbonell-Caballero, J.; Alonso, R.; Ibañez, V.; Terol, J.; Talon, M.; Dopazo, J. A Phylogenetic Analysis of 34 Chloroplast Genomes Elucidates the Relationships between Wild and Domestic Species within the Genus Citrus. *Mol. Biol. Evol.* **2015**, *32*, 2015–2035. [CrossRef]
20. Zhang, Z.; Zhang, Y.; Song, M.; Guan, Y.; Ma, X. Species Identification of Dracaena Using the Complete Chloroplast Genome as a Super-Barcode. *Front. Pharmacol.* **2019**, *10*, 1441. [CrossRef]
21. Huo, Y.; Gao, L.; Liu, B.; Yang, Y.; Kong, S.; Sun, Y.; Wu, X. Complete chloroplast genome sequences of four Allium species: Comparative and phylogenetic analyses. *Sci. Rep.* **2019**, *9*, 12250. [CrossRef] [PubMed]
22. Van Binh Nguyen, V.B.N.; Vo Ngoc Linh Giang, V.N.L.G.; Waminal, N.E.; Park HyunSeung, P.H.; Kim NamHoon, K.N.; Jang WooJong, J.W.; Yang TaeJin, Y.T. Comprehensive comparative analysis of chloroplast genomes from seven Panax species and development of an authentication system based on species-unique single nucleotide polymorphism markers. *J. Ginseng Res.* **2020**, *44*, 135–144. [CrossRef] [PubMed]
23. Hu, H.; Hu, Q.; Al-Shehbaz, I.A.; Luo, X.; Zeng, T.; Guo, X.; Liu, J. Species Delimitation and Interspecific Relationships of the Genus Orychophragmus (*Brassicaceae*) Inferred from Whole Chloroplast Genomes. *Front. Plant Sci.* **2016**, *7*, 1826. [CrossRef] [PubMed]
24. Nie, Y.; Foster, C.S.P.; Zhu, T.; Yao, R.; Duchêne, D.A.; Ho, S.Y.W.; Zhong, B. Accounting for Uncertainty in the Evolutionary Timescale of Green Plants Through Clock-Partitioning and Fossil Calibration Strategies. *Syst. Biol.* **2020**, *69*, 1–16. [CrossRef]
25. Gu, X.; Li, L.; Li, S.; Shi, W.; Zhong, X.; Su, Y.; Wang, T. Adaptive evolution and co-evolution of chloroplast genomes in Pteridaceae species occupying different habitats: Overlapping residues are always highly mutated. *BMC Plant Biol.* **2023**, *23*, 511. [CrossRef] [PubMed]
26. Gao, L.Z.; Liu, Y.L.; Zhang, D.; Li, W.; Gao, J.; Liu, Y.; Li, K.; Shi, C.; Zhao, Y.; Zhao, Y.J.; et al. Evolution of Oryza chloroplast genomes promoted adaptation to diverse ecological habitats. *Commun. Biol.* **2019**, *2*, 278. [CrossRef] [PubMed]
27. Zhai, W.; Duan, X.; Zhang, R.; Guo, C.; Li, L.; Xu, G.; Shan, H.; Kong, H.; Ren, Y. Chloroplast genomic data provide new and robust insights into the phylogeny and evolution of the Ranunculaceae. *Mol. Phylogenetics Evol.* **2019**, *135*, 12–21. [CrossRef] [PubMed]
28. Olejniczak, S.A.; Łojewska, E.; Kowalczyk, T.; Sakowicz, T. Chloroplasts: State of research and practical applications of plastome sequencing. *Planta* **2016**, *244*, 517–527. [CrossRef] [PubMed]
29. Yin, D.; Wang, Y.; Zhang, X.; Ma, X.; He, X.; Zhang, J. Development of chloroplast genome resources for peanut (*Arachis hypogaea* L.) and other species of Arachis. *Sci. Rep.* **2017**, *7*, 11649. [CrossRef]
30. Teske, D.; Peters, A.; Möllers, A.; Fischer, M. Genomic Profiling: The Strengths and Limitations of Chloroplast Genome-Based Plant Variety Authentication. *J. Agric. Food Chem.* **2020**, *68*, 14323–14333. [CrossRef]
31. Wang, Y.; Yu, J.; Chen, Y.K.; Wang, Z.C. Complete Chloroplast Genome Sequence of the Endemic and Endangered Plant Dendropanax oligodontus: Genome Structure, Comparative and Phylogenetic Analysis. *Genes* **2022**, *13*, 2028. [CrossRef] [PubMed]
32. Drouin, G.; Daoud, H.; Xia, J. Relative rates of synonymous substitutions in the mitochondrial, chloroplast and nuclear genomes of seed plants. *Mol. Phylogenetics Evol.* **2008**, *49*, 827–831. [CrossRef] [PubMed]
33. Bertioli, D.J.; Cannon, S.B.; Froenicke, L.; Huang, G.; Farmer, A.D.; Cannon, E.K.; Liu, X.; Gao, D.; Clevenger, J.; Dash, S.; et al. The genome sequences of Arachis duranensis and Arachis ipaensis, the diploid ancestors of cultivated peanut. *Nat. Genet.* **2016**, *48*, 438–446. [CrossRef] [PubMed]
34. Seijo, G.; Lavia, G.I.; Fernández, A.; Krapovickas, A.; Ducasse, D.A.; Bertioli, D.J.; Moscone, E.A. Genomic relationships between the cultivated peanut (*Arachis hypogaea, Leguminosae*) and its close relatives revealed by double GISH. *Am. J. Bot.* **2007**, *94*, 1963–1971. [CrossRef] [PubMed]
35. Tang, Y.; Li, X.; Hu, C.; Qiu, X.; Li, J.; Li, X.; Zhu, H.; Wang, J.; Sui, J.; Qiao, L. Identification and characterization of transposable element AhMITE1 in the genomes of cultivated and two wild peanuts. *BMC Genom.* **2022**, *23*, 500. [CrossRef]

36. Bertioli, D.J.; Jenkins, J.; Clevenger, J.; Dudchenko, O.; Gao, D.; Seijo, G.; Leal-Bertioli, S.C.M.; Ren, L.; Farmer, A.D.; Pandey, M.K.; et al. The genome sequence of segmental allotetraploid peanut Arachis hypogaea. *Nat. Genet.* **2019**, *51*, 877–884. [CrossRef] [PubMed]
37. Chen, X.; Lu, Q.; Liu, H.; Zhang, J.; Hong, Y.; Lan, H.; Li, H.; Wang, J.; Liu, H.; Li, S.; et al. Sequencing of Cultivated Peanut, *Arachis hypogaea*, Yields Insights into Genome Evolution and Oil Improvement. *Mol. Plant* **2019**, *12*, 920–934.e39. [CrossRef] [PubMed]
38. Chen, X.; Li, H.; Pandey, M.K.; Yang, Q.; Wang, X.; Garg, V.; Li, H.; Chi, X.; Doddamani, D.; Hong, Y.; et al. Draft genome of the peanut A-genome progenitor (*Arachis duranensis*) provides insights into geocarpy, oil biosynthesis, and allergens. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 6785–6790. [CrossRef]
39. Wang, J.; Li, Y.; Li, C.; Yan, C.; Zhao, X.; Yuan, C.; Sun, Q.; Shi, C.; Shan, S. Twelve complete chloroplast genomes of wild peanuts: Great genetic resources and a better understanding of Arachis phylogeny. *BMC Plant Biol.* **2019**, *19*, 504. [CrossRef]
40. Prabhudas, S.K.; Prayaga, S.; Madasamy, P.; Natarajan, P. Shallow Whole Genome Sequencing for the Assembly of Complete Chloroplast Genome Sequence of *Arachis hypogaea* L. *Front. Plant Sci.* **2016**, *7*, 1106. [CrossRef]
41. Wang, J.; Li, C.; Yan, C.; Zhao, X.; Shan, S. A comparative analysis of the complete chloroplast genome sequences of four peanut botanical varieties. *PeerJ* **2018**, *6*, e5349. [CrossRef] [PubMed]
42. Grabiele, M.; Chalup, L.; Robledo, G.; Seijo, G.J.P.S. Evolution, Genetic and geographic origin of domesticated peanut as evidenced by 5S rDNA and chloroplast DNA sequences. *Plant Syst. Evol.* **2012**, *298*, 1151–1165. [CrossRef]
43. Tian, X.; Shi, L.; Guo, J.; Fu, L.; Du, P.; Huang, B.; Wu, Y.; Zhang, X.; Wang, Z. Chloroplast Phylogenomic Analyses Reveal a Maternal Hybridization Event Leading to the Formation of Cultivated Peanuts. *Front. Plant Sci.* **2021**, *12*, 804568. [CrossRef] [PubMed]
44. Brock, J.R.; Mandáková, T.; McKain, M.L.M.A.; Olsen, K.M. Chloroplast phylogenomics in Camelina (*Brassicaceae*) reveals multiple origins of polyploid species and the maternal lineage of *C. sativa*. *Hortic. Res.* **2022**, *9*, uhab050. [CrossRef] [PubMed]
45. Meng, J.; Li, X.; Li, H.; Yang, J.; Wang, H.; He, J. Comparative analysis of the complete chloroplast genomes of four aconitum medicinal species. *Molecules* **2018**, *23*, 1015. [CrossRef] [PubMed]
46. Liu, L.X.; Li, R.; Worth, J.R.P.; Li, X.; Li, P.; Cameron, K.M.; Fu, C.X. The complete chloroplast genome of Chinese bayberry (*Morella rubra*, *Myricaceae*): Implications for understanding the evolution of fagales. *Front. Plant Sci.* **2017**, *8*, 968. [CrossRef] [PubMed]
47. Schwarz, E.N.; Ruhlman, T.A.; Sabir, J.S.; Hajrah, N.H.; Alharbi, N.S.; Al-Malki, A.L.; Jansen, R.K. Evolution, Plastid genome sequences of legumes reveal parallel inversions and multiple losses of rps16 in papilionoids. *J. Syst. Evol.* **2015**, *53*, 458–468. [CrossRef]
48. Dobrogojski, J.; Adamiec, M.; Luciński, R. The chloroplast genome: A review. *Acta Physiol. Plant.* **2020**, *42*, 98. [CrossRef]
49. Zhang, Y.; Tian, L.; Lu, C. Chloroplast Gene Expression: Recent Advances and Perspectives. *Plant Commun.* **2023**, *4*, 100611. [CrossRef]
50. Spielman, S.J.; Wilke, C.O. The relationship between dN/dS and scaled selection coefficients. *Mol. Biol. Evol.* **2015**, *32*, 1097–1108. [CrossRef]
51. Nielsen, R.; Yang, Z. Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. *Genetics* **1998**, *148*, 929–936. [CrossRef] [PubMed]
52. Kosakovsky Pond, S.L.; Frost, S.D. Not so different after all: A comparison of methods for detecting amino acid sites under selection. *Mol. Biol. Evol.* **2005**, *22*, 1208–1222. [CrossRef] [PubMed]
53. Singh, A.; Raina, S.N.; Rajpal, V.R.; Singh, A.K. Seed protein fraction electrophoresis in peanut (*Arachis hypogaea* L.) accessions and wild species. *Physiol. Mol. Biol. Plants Int. J. Funct. Plant Biol.* **2018**, *24*, 465–481. [CrossRef] [PubMed]
54. Zhang, H.; Mittal, N.; Leamy, L.J.; Barazani, O.; Song, B.H. Back into the wild-Apply untapped genetic diversity of wild relatives for crop improvement. *Evol. Appl.* **2017**, *10*, 5–24. [CrossRef] [PubMed]
55. Dempewolf, H.; Baute, G.; Anderson, J.; Kilian, B.; Smith, C.; Guarino, L.J.C.S. Past and Future Use of Wild Relatives in Crop Breeding. *Crop Sci.* **2017**, *57*, 1070–1082. [CrossRef]
56. Zheng, Z.; Sun, Z.; Fang, Y.; Qi, F.; Liu, H.; Miao, L.; Du, P.; Shi, L.; Gao, W.; Han, S.; et al. Genetic diversity, population structure, and botanical variety of 320 global peanut accessions revealed through tunable genotyping-by-sequencing. *Sci. Rep.* **2018**, *8*, 14500. [CrossRef] [PubMed]
57. Otyama, P.I.; Kulkarni, R.; Chamberlin, K.; Ozias, A.P.; Chu, Y.; Lincoln, L.M.; MacDonald, G.E.; Anglin, N.L.; Dash, S.; Bertioli, D.J.; et al. Genotypic characterization of the U.S. peanut core collection. *G3-Genes. Genomes Genet.* **2020**, *10*, 4013–4026. [CrossRef] [PubMed]
58. Luo, R.; Liu, B.; Xie, Y.; Li, Z.; Huang, W.; Yuan, J.; He, G.; Chen, Y.; Pan, Q.; Liu, Y.; et al. SOAPdenovo2: An empirically improved memory-efficient short-read de novo assembler. *GigaScience* **2012**, *1*, 18. [CrossRef] [PubMed]
59. Saha, S.; Bridges, S.; Magbanua, Z.V.; Peterson, D.G. Empirical comparison of ab initio repeat finding programs. *Nucleic Acids Res.* **2008**, *36*, 2284–2294. [CrossRef]
60. Behboudi, R.; Nouri-Baygi, M.; Naghibzadeh, M. RPTRF: A rapid perfect tandem repeat finder tool for DNA sequences. *Biosystems* **2023**, *226*, 104869. [CrossRef]
61. Kemena, C.; Dohmen, E.; Bornberg-Bauer, E. DOGMA: A web server for proteome and transcriptome quality assessment. *Nucleic Acids Res.* **2019**, *47*, W507–W510. [CrossRef] [PubMed]

62. Zhang, X.; Smith, D.R. An overview of online resources for intra-species detection of gene duplications. *Front. Genet.* **2022**, *13*, 1012788. [CrossRef]
63. Kanehisa, M.; Goto, S.; Hattori, M.; Aoki-Kinoshita, K.F.; Itoh, M.; Kawashima, S.; Katayama, T.; Araki, M.; Hirakawa, M. From genomics to chemical genomics: New developments in KEGG. *Nucleic Acids Res.* **2006**, *34*, D354–D357. [CrossRef] [PubMed]
64. Huckvale, E.; Moseley, H.N.B. kegg_pull: A software package for the RESTful access and pulling from the Kyoto Encyclopedia of Gene and Genomes. *BMC Bioinform.* **2023**, *24*, 78. [CrossRef]
65. Galperin, M.Y.; Wolf, Y.I.; Makarova, K.S.; Vera Alvarez, R.; Landsman, D.; Koonin, E.V. COG database update: Focus on microbial diversity, model organisms, and widespread pathogens. *Nucleic Acids Res.* **2021**, *49*, D274–D281. [CrossRef]
66. Tatusov, R.L.; Fedorova, N.D.; Jackson, J.D.; Jacobs, A.R.; Kiryutin, B.; Koonin, E.V.; Krylov, D.M.; Mazumder, R.; Mekhedov, S.L.; Nikolskaya, A.N.; et al. The COG database: An updated version includes eukaryotes. *BMC Bioinform.* **2003**, *4*, 41. [CrossRef] [PubMed]
67. Gene Ontology Consortium. Gene Ontology Consortium: Going forward. *Nucleic Acids Res.* **2015**, *43*, D1049–D1056. [CrossRef] [PubMed]
68. The Gene Ontology Consortium. The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Res.* **2019**, *47*, D330–D338. [CrossRef]
69. Magrane, M. UniProt Knowledgebase: A hub of integrated protein data. *Database J. Biol. Databases Curation* **2011**, *2011*, bar009. [CrossRef]