

Article

Comparative Chloroplast Genomics and Codon Usage Bias Analysis in *Hevea* Genus

Yang Yang^{1,2}, Xueyang Liu³, Lixia He^{1,4}, Zhenhua Li³, Boxuan Yuan², Fengyan Fang², Mei Wang², Aifang Li², Cheng Liu^{3,4}, Minmin He² , Shugang Hui², Wenda Wang^{3,*} and Xuchu Wang^{1,2,*}

- ¹ Plant Stress Resistance Integrated Biology Laboratory, College of Life Sciences, Hainan Normal University, Haikou 571158, China; yy15776580499@163.com (Y.Y.); helixia0605@163.com (L.H.)
- ² Key Laboratory of Plant Resource Conservation and Germplasm Innovation in Mountainous Region (Ministry of Education), Institute of Agro-Bioengineering, College of Life Sciences, Guizhou University, Guiyang 550025, China; yuanboxuan111@163.com (B.Y.); fangfengyan0913@163.com (F.F.); wangmei7102@163.com (M.W.); afangliguizhou@163.com (A.L.); hmyia@163.com (M.H.); sghui@gzu.edu.cn (S.H.)
- ³ Photosynthesis Research Center, Key Laboratory of Photobiology, Institute of Botany, Chinese Academy of Sciences, Beijing 100093, China; liuxueyang@ibcas.ac.cn (X.L.); lizhenhua@ibcas.ac.cn (Z.L.); liucheng@ibcas.ac.cn (C.L.)
- ⁴ Academician Workstation of Agricultural High-Tech Industrial Area of the Yellow River Delta, National Center of Technology Innovation for Comprehensive Utilization of Saline-Alkali Land, Dongying 257300, China
- * Correspondence: wdwang@ibcas.ac.cn (W.W.); xcwang@gzu.edu.cn (X.W.)

Abstract: Objectives: This study investigates the cpDNA sequences from six *Hevea* species, aiming to explore their genomic characteristics, gene content, and genetic relationships. The objectives include understanding the structure of these genomes, identifying potential gene rearrangements, and providing insights into genetic improvement and conservation strategies for the *Hevea* genus. Methods: cpDNA sequences from six *Hevea* species were sequenced and analyzed. Genome sizes, GC content, gene encoding potential, and structural integrity were assessed. Simple sequence repeats (SSRs) and codon usage were analyzed, with a focus on optimal codons and their frequency. Phylogenetic analysis was conducted to determine the genetic relationships within the *Hevea* genus. Results: The cpDNAs from the six species exhibited genome sizes ranging from 161,093 bp to 161,254 bp, with GC content between 35.72% and 35.75%. Each genome contained 91 to 92 protein-coding genes, with the *infA* gene consistently present. No significant gene rearrangements were detected, and SSR analysis revealed mono-repeats primarily composed of A/T bases. Codon usage analysis indicated that leucine is predominantly encoded by the UUA codon, and 31 optimal codons were identified, mainly ending in A or U. Phylogenetic analysis clarified the genetic relationships among the species. Conclusions: The study provides detailed insights into the cpDNA characteristics of *Hevea* species, highlighting stable genome structures, conserved genes, and specific patterns of codon usage. These findings are valuable for conservation efforts, genetic improvement strategies, and the sustainable use of *Hevea* germplasm.

Keywords: chloroplast genome; codon; RSCU; *Hevea*



Academic Editor: Jacqueline Batley

Received: 12 January 2025

Revised: 1 February 2025

Accepted: 4 February 2025

Published: 6 February 2025

Citation: Yang, Y.; Liu, X.; He, L.; Li, Z.; Yuan, B.; Fang, F.; Wang, M.; Li, A.; Liu, C.; He, M.; et al. Comparative Chloroplast Genomics and Codon Usage Bias Analysis in *Hevea* Genus. *Genes* **2025**, *16*, 201. <https://doi.org/10.3390/genes16020201>

Copyright: © 2025 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

Natural rubber from rubber trees is a critical industrial raw material and strategic resource [1]. In December 2022, global rubber production reached 1.397 million tons. However, due to the enormous domestic consumption, China's self-sufficiency rate for natural rubber is only 15% [2]. Given the current international situation and regional

instability caused by the global debt crisis, addressing the supply issue of natural rubber is particularly urgent [3]. After a century of traditional breeding efforts by researchers both domestically and internationally, the rubber yield of wild populations has been significantly improved [4]. Nevertheless, due to climate and regional constraints, the suitable planting areas in China, including Hainan, Yunnan, and Guangdong province, have almost reached the limit [5].

Breeding cycles for woody plants such as rubber trees are prolonged, and their genetic composition is highly complex. Chloroplast genome (cpDNA) research provides valuable insights that can facilitate the development of crops with enhanced traits. Specifically, the study of cpDNA contributes to improving photosynthetic efficiency, understanding stress tolerance mechanisms, and identifying genes linked to desirable traits. Furthermore, chloroplast genome research plays a pivotal role in taxonomy, reconstructing phylogenetic relationships, and exploring the adaptive evolution of plant species. By 2024, 34,923 plant cpDNAs, including that of *Hevea brasiliensis* RRIM600 [6], had been sequenced and cataloged in the Chloroplast Genome Information Resource (CGIR) database. The *Hevea* genus consists of 11 species, with *H. brasiliensis* (RRIM600/PR107) [7], *Hevea pauciflora* [8], and *Hevea camargoana* [9] having their cpDNAs fully sequenced. Various *Hevea* varieties are extensively cultivated in China's rubber-growing regions, contributing significantly to socio-economic benefits and fostering the rapid and sustainable growth of the natural rubber industry [10]. This research aims to investigate the variations in gene content, intra-species diversity, and structural differences in the cpDNAs among different *Hevea* genus. Our findings highlight the importance of chloroplast genome data in understanding the evolution, phylogeny, and genetic diversity of the *Hevea* genus. This information provides valuable insights into the genetic resources of *Hevea* species, offering a theoretical foundation for future conservation, genetic improvement, and breeding programs.

2. Materials and Methods

2.1. Sequence Retrieval and Filtering

The cpDNAs of six *Hevea* species were downloaded from the NCBI GenBank database (Table 1) [11]. Utilizing Geneious 9.0.2 software [12], we performed statistical and comparative analyses on gene composition and structure, focusing on GC content and sequence lengths of LSC, IRb, SSC, and IRa regions. Data analysis and visualization were carried out using Excel 2019 and OriginPro 2021 [13].

Table 1. Synopsis of the cpDNAs of *Hevea* genus.

Species	GenBank Number	Genome Size (bp)	Genes	rRNA	tRNA	GC%
<i>H. benthamiana</i>	MT333859	161,124	91	8	36	35.72
<i>H. brasiliensis</i>	NC015308	161,191	92	8	37	35.73
<i>H. camargoana</i>	MN781109	161,291	91	8	36	35.72
<i>H. nitida</i>	MT413435	161,124	91	8	36	35.73
<i>H. pauciflora</i>	NC059798	161,123	91	8	36	35.75
<i>H. spruceana</i>	NC059799	161,093	91	8	36	35.72

2.2. IR/SC Boundary Contraction and Expansion Analysis

Using IRscope [14], a comparative study of IR/SC boundaries in the cpDNAs of *Hevea* species was conducted, creating diagrams to illustrate IR/SC region boundaries. We examined the differences at the four junction sites of chloroplast genome regions. For the analysis of six *Hevea* cpDNAs, *H. brasiliensis* served as the reference sequence [15]. All sequences were aligned using MAFFT v7.450 software [16]. Post-alignment, manual inspection, and corrections were performed to ensure data precision. Collinearity analysis

was conducted on the aligned sequences using the Mauve tool function in Geneious 9.0.2 software to deepen our understanding of the genomic structural relationships among different *Hevea* species. A Python script extracted annotation information from chloroplast genome gb files, and along with the genome sequences, these were uploaded to the mVISTA platform [17], which was used for comprehensive genome comparison to detect sequence variations. Initially, genome comparison diagrams were generated using mVISTA.

2.3. SSR Analysis

SSRs known for their significant polymorphisms [18] were analyzed, using the MISA-web tool [19] to determine the base sequences at SSR locations. REPuter was employed to examine types of repeat sequences [20]. The analysis parameters included a minimum match score of 50, and a maximum of 3 mismatches for the analysis of dispersed repeat sequences.

2.4. Phylogenetic Study of *Hevea* Chloroplast DNA

The chloroplast DNA sequences of six *Hevea* species were retrieved from the NCBI database, as detailed in Table 1. To ensure the accuracy and consistency of the sequences, we performed a BLAST analysis to validate their similarity and quality before subsequent analyses. Using Geneious 9.0.2 software, we constructed a phylogenetic tree to determine the evolutionary relationships and phylogenetic positions of the six *Hevea* species. This approach allowed us to confirm the reliability of the sequences and ensured robust phylogenetic analysis.

2.5. Analysis of Codon Usage Frequency

CodonW 1.4.2 software was used for statistical analysis of codon usage [21]. The EMBOSS platform analyzed the GC content of coding sequences (CDS) (designated as GC1, GC2, and GC3, respectively) and their averages [22]. Correlation studies were conducted using R 4.3.1. Pearson correlation coefficients were calculated using the cor function in base R, and significance testing was performed with cor.test. Results were visualized using the ggcorrplot package 0.1.4.1 [23]. The selection criteria for CDSs in *Hevea* cpDNAs were as follows: (1) CDSs must have a base count that is a multiple of three for sequence completeness. (2) The minimum coding sequence length should be 300 bp to ensure reliability. (3) High-quality sequences should contain only standard bases, avoiding non-standard bases. (4) CDSs should be free of internal stop codons to ensure proper transcription. Perl scripts were employed to filter CDSs according to these rules.

2.6. Analysis of RSCU and Its Frequency

Relative synonymous codon usage (RSCU) is a key metric for measuring codon usage bias in gene expression [24]. The formula is as follows:

$$\text{RSCU} = \frac{x_{ij}}{\sum_j^{n_i} x_{ij}} n_i$$

2.7. ENc-Plot Analysis

The effective number of codons (ENc) measures codon usage bias, with values ranging from 20 to 61 [25]. The GC3s value represents the ratio of G and C content at the third position of a codon (excluding Met and Trp) to the total number of bases in the gene [26]. The expected ENc value is calculated using

$$\text{ENc} = 2 + S + \frac{29}{S^2 + (1 - S)^2}$$

Research on optimal codons depends on ENc bias. CDSs are sorted by their ENc values, and the highest and lowest 10% of genes are chosen to form a biased gene library [27]. Codons with a ΔRSCU value ≥ 0.08 and an RSCU value > 1 are considered high-frequency codons [28]. Codons meeting both conditions are identified as optimal codons.

The PR2-plot method is employed to assess the base composition at the third position of codons in amino acids [29]. The extent to which the bases cluster toward the center (where A equals T and C equals G) and their distribution direction within the plot highlight base preferences.

Neutrality analysis of GC12 and GC3 evaluates the effects of mutations and natural selection on codon usage [29]. Correspondence analysis (COA) based on RSCU was utilized to compare codon usage patterns, excluding codons for Met, Trp, and the three stop codons. This method represented the variation in codon usage through orthogonal axes, identifying the primary sources of gene variation [30].

2.8. Photosystem Protein Structure Prediction and Functional Annotation

The protein sequences of Photosystem I (PSI) and Photosystem II (PSII) in *H. brasiliensis* were used for three-dimensional (3D) structure modeling and functional prediction via SWISS-MODEL. The core subunits of PSI and PSII (e.g., PsaA, PsaB, PsaC, PsaD, PsaE, PsaF, PsaG, PsaH, PsaI, PsaJ, PsaK, PsaL, PsaM, PsaN, PsaO, PsaP, PsaQ, PsaR, PsaS, PsaT, PsaU, PsaV, PsaW, PsaX, PsaY, PsaZ, PsaAA, PsaAB, PsaAC, PsaAD, PsaAE, PsaAF, PsaAG, PsaAH, PsaAI, PsaAJ, PsaAK, PsaAL, PsaAM, PsaAN, PsaAO, PsaAP, PsaAQ, PsaAR, PsaAS, PsaAT, PsaAU, PsaAV, PsaAW, PsaAX, PsaAY, PsaAZ, PsaBA, PsaBB, PsaBC, PsaBD, PsaBE, PsaBF, PsaBG, PsaBH, PsaBI, PsaBJ, PsaBK, PsaBL, PsaBM, PsaBN, PsaBO, PsaBP, PsaBQ, PsaBR, PsaBS, PsaBT, PsaBU, PsaBV, PsaBW, PsaBX, PsaBY, PsaBZ, PsaCA, PsaCB, PsaCC, PsaCD, PsaCE, PsaCF, PsaCG, PsaCH, PsaCI, PsaCJ, PsaCK, PsaCL, PsaCM, PsaCN, PsaCO, PsaCP, PsaCQ, PsaCR, PsaCS, PsaCT, PsaCU, PsaCV, PsaCW, PsaCX, PsaCY, PsaCZ, PsaDA, PsaDB, PsaDC, PsaDD, PsaDE, PsaDF, PsaDG, PsaDH, PsaDI, PsaDJ, PsaDK, PsaDL, PsaDM, PsaDN, PsaDO, PsaDP, PsaDQ, PsaDR, PsaDS, PsaDT, PsaDU, PsaDV, PsaDW, PsaDX, PsaDY, PsaDZ, PsaEA, PsaEB, PsaEC, PsaED, PsaEE, PsaEF, PsaEG, PsaEH, PsaEI, PsaEJ, PsaEK, PsaEL, PsaEM, PsaEN, PsaEO, PsaEP, PsaEQ, PsaER, PsaES, PsaET, PsaEU, PsaEV, PsaEW, PsaEX, PsaEY, PsaEZ, PsaFA, PsaFB, PsaFC, PsaFD, PsaFE, PsaFF, PsaFG, PsaFH, PsaFI, PsaFJ, PsaFK, PsaFL, PsaFM, PsaFN, PsaFO, PsaFP, PsaFQ, PsaFR, PsaFS, PsaFT, PsaFU, PsaFV, PsaFW, PsaFX, PsaFY, PsaFZ, PsaGA, PsaGB, PsaGC, PsaGD, PsaGE, PsaGF, PsaGG, PsaGH, PsaGI, PsaGJ, PsaGK, PsaGL, PsaGM, PsaGN, PsaGO, PsaGP, PsaGQ, PsaGR, PsaGS, PsaGT, PsaGU, PsaGV, PsaGW, PsaGX, PsaGY, PsaGZ, PsaHA, PsaHB, PsaHC, PsaHD, PsaHE, PsaHF, PsaHG, PsaHH, PsaHI, PsaHJ, PsaHK, PsaHL, PsaHM, PsaHN, PsaHO, PsaHP, PsaHQ, PsaHR, PsaHS, PsaHT, PsaHU, PsaHV, PsaHW, PsaHX, PsaHY, PsaHZ, PsaIA, PsaIB, PsaIC, PsaID, PsaIE, PsaIF, PsaIG, PsaIH, PsaII, PsaIJ, PsaIK, PsaIL, PsaIM, PsaIN, PsaIO, PsaIP, PsaIQ, PsaIR, PsaIS, PsaIT, PsaIU, PsaIV, PsaIW, PsaIX, PsaIY, PsaIZ, PsaJA, PsaJB, PsaJC, PsaJD, PsaJE, PsaJF, PsaJG, PsaJH, PsaJI, PsaJJ, PsaJK, PsaJL, PsaJM, PsaJN, PsaJO, PsaJP, PsaJQ, PsaJR, PsaJS, PsaJT, PsaJU, PsaJV, PsaJW, PsaJX, PsaJY, PsaJZ, PsaKA, PsaKB, PsaKC, PsaKD, PsaKE, PsaKF, PsaKG, PsaKH, PsaKI, PsaKJ, PsaKK, PsaKL, PsaKM, PsaKN, PsaKO, PsaKP, PsaKQ, PsaKR, PsaKS, PsaKT, PsaKU, PsaKV, PsaKW, PsaKX, PsaKY, PsaKZ, PsaLA, PsaLB, PsaLC, PsaLD, PsaLE, PsaLF, PsaLG, PsaLH, PsaLI, PsaLJ, PsaLK, PsaLL, PsaLM, PsaLN, PsaLO, PsaLP, PsaLQ, PsaLR, PsaLS, PsaLT, PsaLU, PsaLV, PsaLW, PsaLX, PsaLY, PsaLZ, PsaMA, PsaMB, PsaMC, PsaMD, PsaME, PsaMF, PsaMG, PsaMH, PsaMI, PsaMJ, PsaMK, PsaML, PsaMM, PsaMN, PsaMO, PsaMP, PsaMQ, PsaMR, PsaMS, PsaMT, PsaMU, PsaMV, PsaMW, PsaMX, PsaMY, PsaMZ, PsaNA, PsaNB, PsaNC, PsaND, PsaNE, PsaNF, PsaNG, PsaNH, PsaNI, PsaNJ, PsaNK, PsaNL, PsaNM, PsaNN, PsaNO, PsaNP, PsaNQ, PsaNR, PsaNS, PsaNT, PsaNU, PsaNV, PsaNW, PsaNX, PsaNY, PsaNZ, PsaOA, PsaOB, PsaOC, PsaOD, PsaOE, PsaOF, PsaOG, PsaOH, PsaOI, PsaOJ, PsaOK, PsaOL, PsaOM, PsaON, PsaOO, PsaOP, PsaOQ, PsaOR, PsaOS, PsaOT, PsaOU, PsaOV, PsaOW, PsaOX, PsaOY, PsaOZ, PsaPA, PsaPB, PsaPC, PsaPD, PsaPE, PsaPF, PsaPG, PsaPH, PsaPI, PsaPJ, PsaPK, PsaPL, PsaPM, PsaPN, PsaPO, PsaPP, PsaPQ, PsaPR, PsaPS, PsaPT, PsaPU, PsaPV, PsaPW, PsaPX, PsaPY, PsaPZ, PsaQA, PsaQB, PsaQC, PsaQD, PsaQE, PsaQF, PsaQG, PsaQH, PsaQI, PsaQJ, PsaQK, PsaQL, PsaQM, PsaQN, PsaQO, PsaQP, PsaQQ, PsaQR, PsaQS, PsaQT, PsaQU, PsaQV, PsaQW, PsaQX, PsaQY, PsaQZ, PsaRA, PsaRB, PsaRC, PsaRD, PsaRE, PsaRF, PsaRG, PsaRH, PsaRI, PsaRJ, PsaRK, PsaRL, PsaRM, PsaRN, PsaRO, PsaRP, PsaRQ, PsaRR, PsaRS, PsaRT, PsaRU, PsaRV, PsaRW, PsaRX, PsaRY, PsaRZ, PsaSA, PsaSB, PsaSC, PsaSD, PsaSE, PsaSF, PsaSG, PsaSH, PsaSI, PsaSJ, PsaSK, PsaSL, PsaSM, PsaSN, PsaSO, PsaSP, PsaSQ, PsaSR, PsaSS, PsaST, PsaSU, PsaSV, PsaSW, PsaSX, PsaSY, PsaSZ, PsaTA, PsaTB, PsaTC, PsaTD, PsaTE, PsaTF, PsaTG, PsaTH, PsaTI, PsaTJ, PsaTK, PsaTL, PsaTM, PsaTN, PsaTO, PsaTP, PsaTQ, PsaTR, PsaTS, PsaTT, PsaTU, PsaTV, PsaTW, PsaTX, PsaTY, PsaTZ, PsaUA, PsaUB, PsaUC, PsaUD, PsaUE, PsaUF, PsaUG, PsaUH, PsaUI, PsaUJ, PsaUK, PsaUL, PsaUM, PsaUN, PsaUO, PsaUP, PsaUQ, PsaUR, PsaUS, PsaUT, PsaUU, PsaUV, PsaUW, PsaUX, PsaUY, PsaUZ, PsaVA, PsaVB, PsaVC, PsaVD, PsaVE, PsaVF, PsaVG, PsaVH, PsaVI, PsaVJ, PsaVK, PsaVL, PsaVM, PsaVN, PsaVO, PsaVP, PsaVQ, PsaVR, PsaVS, PsaVT, PsaVU, PsaVV, PsaVW, PsaVX, PsaVY, PsaVZ, PsaWA, PsaWB, PsaWC, PsaWD, PsaWE, PsaWF, PsaWG, PsaWH, PsaWI, PsaWJ, PsaWK, PsaWL, PsaWM, PsaWN, PsaWO, PsaWP, PsaWQ, PsaWR, PsaWS, PsaWT, PsaWU, PsaWV, PsaWW, PsaWX, PsaWY, PsaWZ, PsaXA, PsaXB, PsaXC, PsaXD, PsaXE, PsaXF, PsaXG, PsaXH, PsaXI, PsaXJ, PsaXK, PsaXL, PsaXM, PsaXN, PsaXO, PsaXP, PsaXQ, PsaXR, PsaXS, PsaXT, PsaXU, PsaXV, PsaXW, PsaXX, PsaXY, PsaXZ, PsaYA, PsaYB, PsaYC, PsaYD, PsaYE, PsaYF, PsaYG, PsaYH, PsaYI, PsaYJ, PsaYK, PsaYL, PsaYM, PsaYN, PsaYO, PsaYP, PsaYQ, PsaYR, PsaYS, PsaYT, PsaYU, PsaYV, PsaYW, PsaYX, PsaYY, PsaYZ, PsaZA, PsaZB, PsaZC, PsaZD, PsaZE, PsaZF, PsaZG, PsaZH, PsaZI, PsaZJ, PsaZK, PsaZL, PsaZM, PsaZN, PsaZO, PsaZP, PsaZQ, PsaZR, PsaZS, PsaZT, PsaZU, PsaZV, PsaZW, PsaZX, PsaZY, PsaZZ).

3. Results

3.1. Structural Attributes of the Hevea Chloroplast Genome

The analysis revealed that the lengths of the *Hevea* cpDNAs span from 161,093 bp (*H. spruceana*) to 161,254 bp (*H. camargoana*). The LSC, SSC, IRA, and IRB regions' lengths ranged from 89,079 to 89,281 bp, 18,362 to 18,376 bp, 26,810 to 26,819 bp, and 26,810 to 26,819 bp, respectively. The GC content in these genomes surpasses 15%, varying from 35.72% to 35.75%. The gene counts for rRNA and tRNA were notably consistent, with eight rRNA genes and 36 to 37 tRNA genes. The chloroplast genome comprised 91 to 92 genes coding for proteins. Within the IR region, there were eight tRNA genes, four rRNA genes, and seven genes (rps19, rpl2, ycf2, rpl23, ndhB, rps7, and rps12). Importantly, a distinctive gene, *infA*, was identified across all *Hevea* species. The *Hevea* cpDNAs were annotated with 135 to 137 genes, which included 91 to 92 protein-coding genes (Table 1).

3.2. Comparative Genomic Analysis

The cpDNAs of the *Hevea* genus show highly similar sequences. While the coding regions of these genomes are relatively consistent, the non-coding regions exhibit greater variability. Highly divergent areas are primarily those such as *ndhK-atpE*, *accD-PsaI*, *PetG-PsaJ*, *rps7-trnV-GAC*, and *ndhF-rpl32* (Figure 1). These divergent areas may be used for

The figure displays genomic tracks for six species: *H. brasiliensis*, *H. benthamiana*, *H. camargoana*, *H. nitida*, *H. pauciflora*, and *H. spruceana*. The top track shows the LSC region (green bar) and the bottom track shows the IRA (yellow), SSC (red), and IRB (yellow) regions. The tracks display gene models (blue), exons (red), UTRs (cyan), CNS (red), and mRNA (grey) for each species.

3.3. IR Region Contraction and Expansion

Figure 2. Comparison of the boundaries of LSC, SSC, and IR regions in the *Hevea* cpDNAs. Note: Genes are depicted by short bars, and the intervals and boundaries between genes are represented by base pair length. Gene extensions are indicated above the bars.

3.4. Analysis of Long Repeats and SSRs

Six types of simple sequence repeats (SSRs) were identified in the cpDNAs of *Hevea* species. The number of cpSSRs ranged from 126 in *H. brasiliensis* to 137 in *H. spruceana*. Mononucleotide repeats were the most prevalent, representing 69.05% (*H. brasiliensis*) to 71.53% (*H. spruceana*) of the total SSRs, followed by dinucleotide repeats (13.14% to 14.29%) and trinucleotide repeats (3.17% to 4.51%). The percentage of tetranucleotide repeats, varying from 8.03% to 9.52%, was higher than that of trinucleotide (3.17% to 4.51%) and pentanucleotide repeats (1.50% to 2.38%). Pentanucleotide and hexanucleotide repeats were found in all six cpDNAs analyzed, with hexanucleotide repeats being particularly uncommon, comprising only 0.73% to 1.59% (Figure 3). Most SSRs consisted of A/T repeats rather than G/C repeats. The count of these repeats varied from 98 in *H. brasiliensis* to 146 in *H. pauciflora*. Despite notable differences in the lengths of forward and palindromic repeats, many interspersed repeats were between 30 to 39 bp in length (Figure 3).

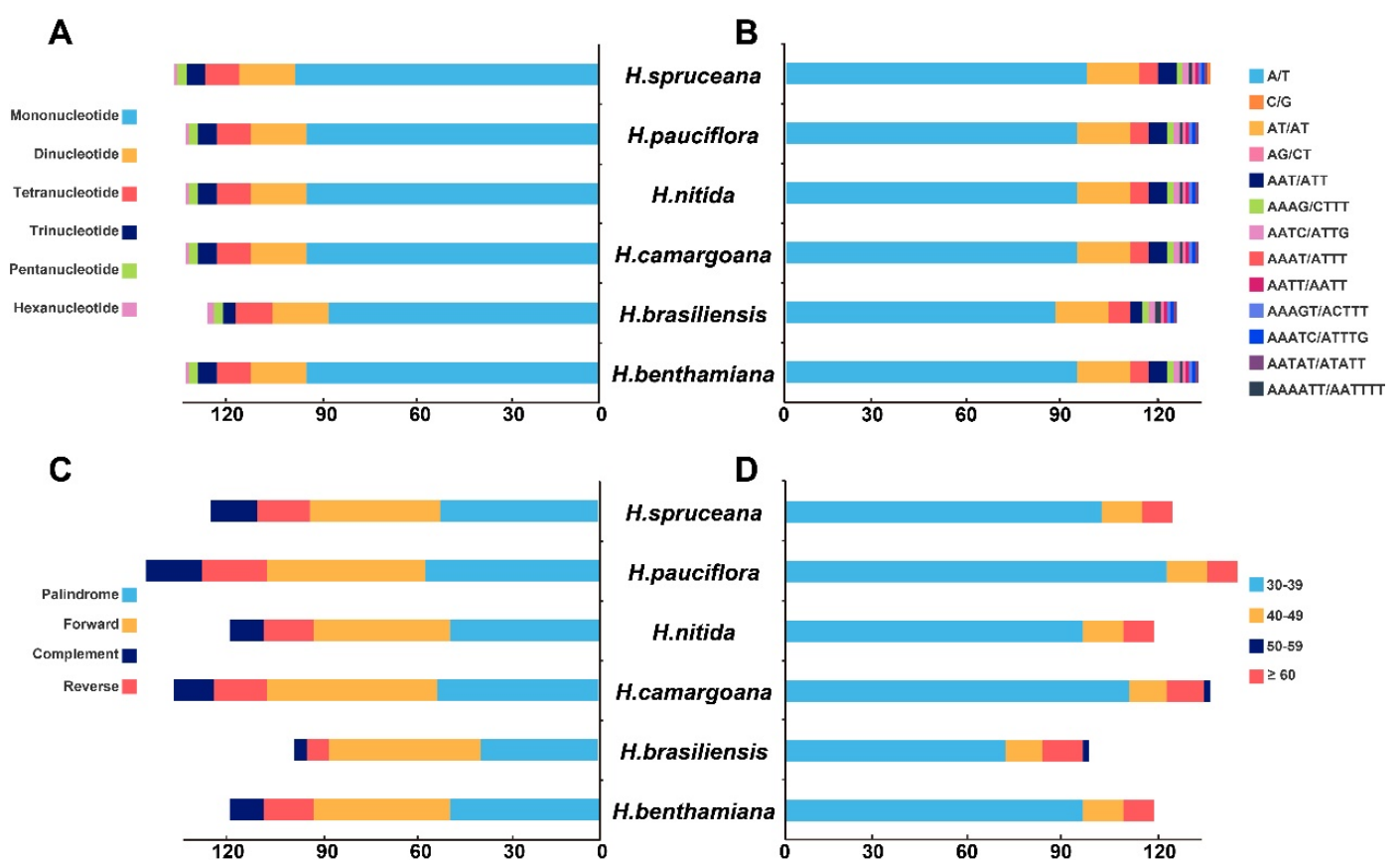


Figure 3. Quantitative analysis of SSRs and long repeat sequences in the *Hevea* genus chloroplast genome: (A) types of SSRs, (B) frequency of identified SSRs, (C) types of long repeat sequences, and (D) number of long repeat sequences by length.

3.5. Codon Usage Bias Analysis

In the analysis of *Hevea* chloroplast coding sequences, leucine (Leu) was predominantly encoded by the UUA codon, exhibiting a high RSCU value exceeding 2.00 (2.0291–2.0376). Conversely, tyrosine (Tyr) was encoded to a much lesser extent, with an RSCU value between 0.3217 and 0.3243. Met and Trp had RSCU values precisely at 1.00 (Figure 4).

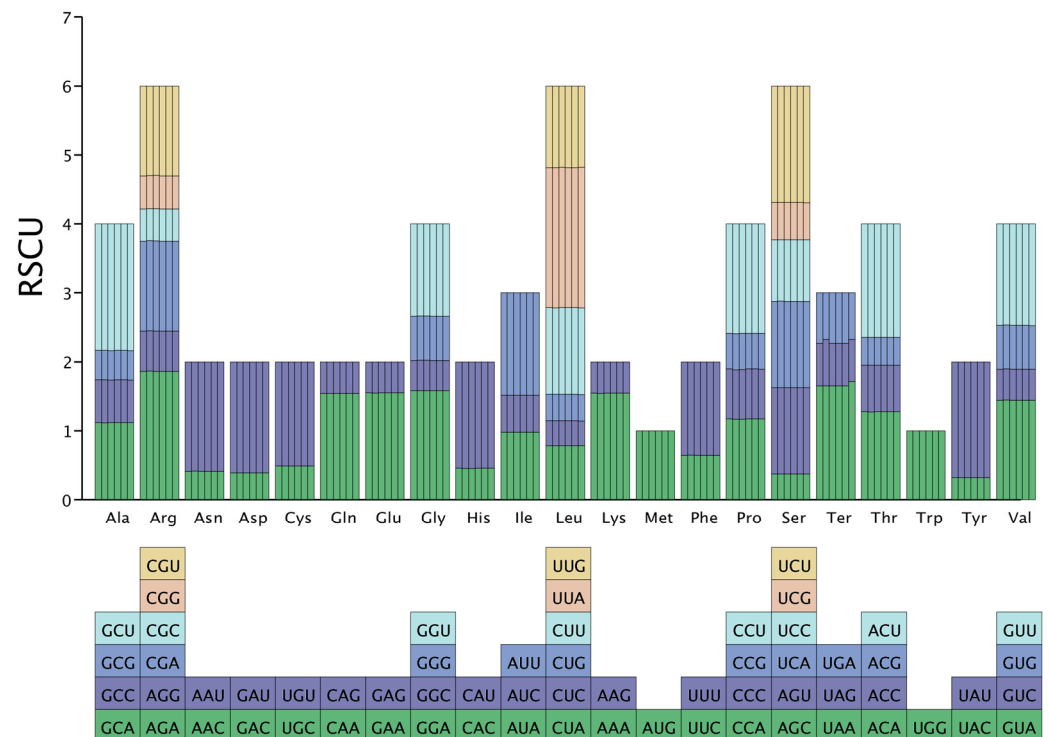


Figure 4. Relative usage of synonymous codons in the *Hevea* cpDNAs.

3.6. Analysis of ENc-Plot

In this study, the average ENc for cpDNAs ranged from 46.47 to 46.48, with individual values spanning from 35.87 to 56.8. The genes with ENc values above 45 numbered between 43 and 45. Many *Hevea* chloroplast genes fell below the anticipated ENc standard curve, indicating that natural selection predominantly shapes codon usage bias (Figure 5). ENc-plot analysis shows that the ENc values matched the predicted values, suggesting that codon choice was driven mainly by base variations.

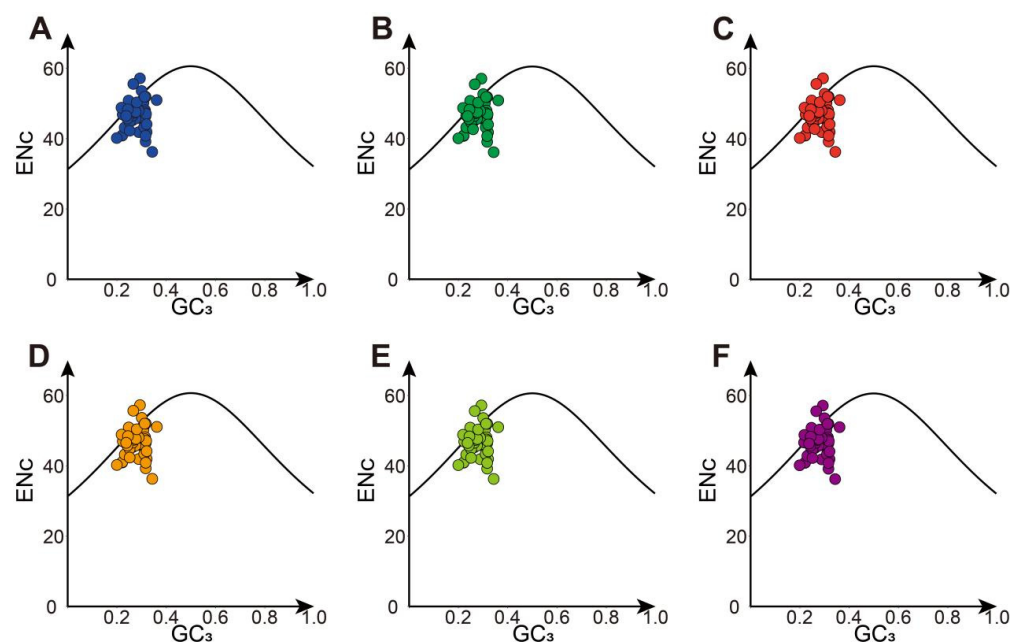


Figure 5. The relationship between the ENc and GC3 in the chloroplast genome of *Hevea* genus (ENc-plot). (A) *H. benthamiana*; (B) *H. brasiliensis*; (C) *H. camargoana*; (D) *H. nitida*; (E) *H. pauciflora*; (F) *H. spruceana*.

3.7. Analysis of PR2-Bias Plot

The overall average GC content in *Hevea* cpDNAs ranges from 37.64% to 37.80%. Specifically, GC1, GC2, and GC3 range from 28.02% to 45.32%, 26.16% to 56.12%, and 19.77% to 36.15%, respectively. These values indicate a preference for A/U bases over G/C bases in the codon selection of *Hevea* cpDNAs. Additionally, the varying GC content at GC1, GC2, and GC3 underscores the differences among codon positions (Figure 6).

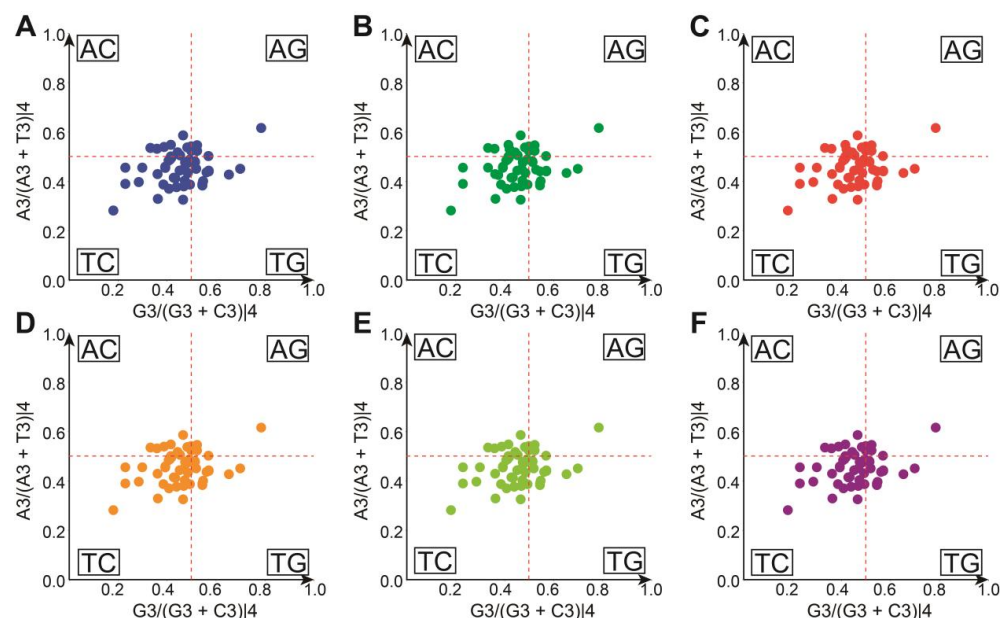


Figure 6. Codon bias analysis of *Hevea* chloroplast genome (PR2-plot). (A) *H. benthamiana*; (B) *H. brasiliensis*; (C) *H. camargoana*; (D) *H. nitida*; (E) *H. pauciflora*; (F) *H. spruceana*.

When examining the six *Hevea* chloroplast genome data sets, their coordinate points were unevenly spread across the four quadrants of the graph, primarily concentrated in the $G3/(G3 + C3) < 0.5$ and $A3/(A3 + T3) < 0.5$ regions. This suggests a higher frequency of C relative to G at the third codon position, and T being used more frequently than A. If base mutation were the sole factor in codon usage preference, the frequencies of A/T and G/C would be expected to be equal or nearly equal. PR2-plot analysis indicates that codon usage bias in *Hevea* cpDNAs is a complex process driven by the interaction of base mutations and natural selection.

3.8. Analysis of Neutrality Plot

The neutrality plot analysis of *Hevea* chloroplast genes shows that these genes are relatively concentrated, all positioned above the diagonal line. The GC12 values range from 30.32% to 54.32%, while GC3 values range from 19.77% to 36.15%. The correlation coefficients between GC12 and GC3 span from 0.1825 to 0.2697, suggesting that mutational pressure does not significantly impact codon usage bias. Additionally, all points are above the 1:1 line, indicating that GC12 is greater than GC3, which implies that the G+C content at the third codon position is low (Figure 7). The regression slope analysis indicates that the influence of mutational pressure on codon usage patterns in cpDNAs ranges from 20.00% to 25.00%, indicating that mutational pressure has a relatively minor influence, while natural selection is the main factor driving codon selection preference.

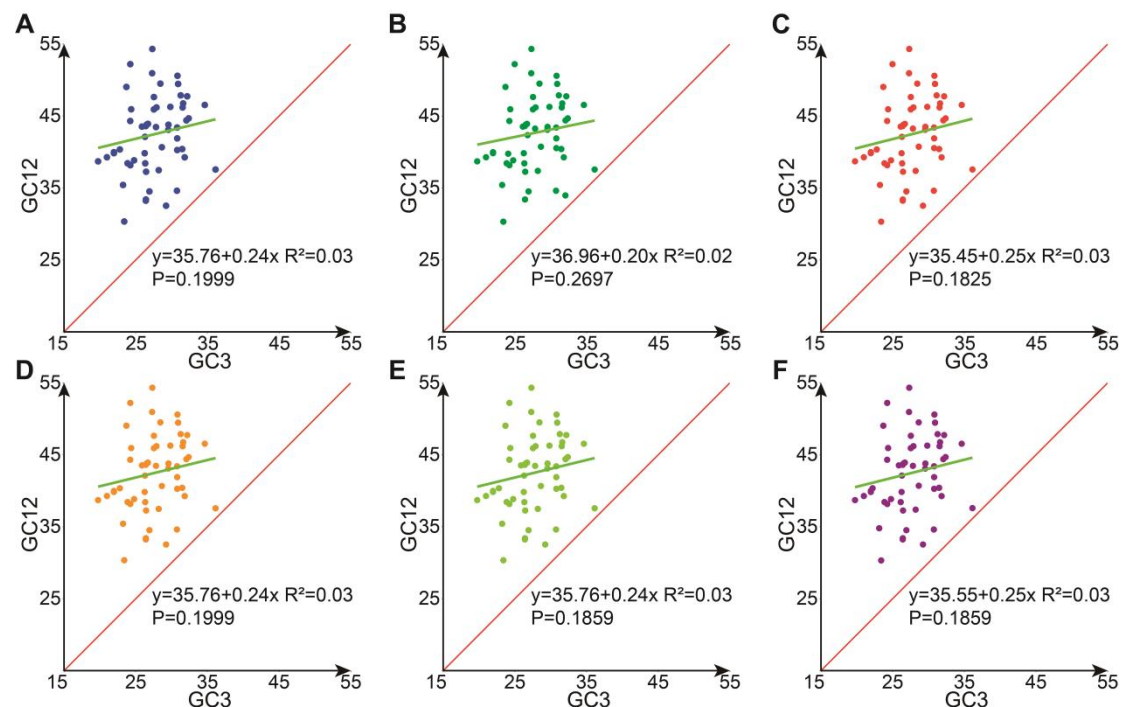


Figure 7. Codon neutrality mapping analysis of *Hevea* chloroplast genes. (A) *H. benthamiana*; (B) *H. brasiliensis*; (C) *H. camargoana*; (D) *H. nitida*; (E) *H. pauciflora*; (F) *H. spruceana*.

3.9. Analysis of Correspondence

Using COA based on RSCU, the RSCU of each gene was plotted against the average values on axes 1 and 2 at the origin. The genes were distributed throughout the coordinate system of axes 1 and 2, showing no significant preference, which indicates a certain uniformity in codon usage among different chloroplast genes. Further examination of codon diversity sources in the *Hevea* chloroplast genome positioned the coding sequences (CDSs) on the principal component analysis plane. COA revealed that the contribution rates of axes 1 and 2 range from 10.11% to 10.22% and 9.59% to 9.66%, respectively (Figure 8). This finding indicates that the first axis contributes the most to explaining codon usage preference but only accounts for part of the variation. Genes were grouped into five functional categories based on function: genetic system genes, ribosomal protein genes, photosystem genes, other protein genes, and conserved open reading frames. The dispersed distribution of different categories of genes on the COA plane indicates significant differences in codon usage patterns among genes with various functions. The development of codon usage bias in *Hevea* chloroplast genes is a complex process influenced by multiple factors, not just a single factor. Although the first axis plays the largest role in revealing codon usage bias, it only represents a portion of the variation, suggesting the existence of other influencing factors not accounted for.

Further codon usage analysis demonstrated that ENc and CAI were negatively correlated with axis 1, while GC3 content showed a positive correlation with ENc (Figure 9). The same species show a negative correlation between CAI and axis 1, indicating lower CAI may reduce gene expression efficiency due to specific environmental pressures. Additionally, in *H. benthamiana*, *H. nitida*, and *H. pauciflora*, this negative correlation highlights the inhibition of codon adaptation and diversity. Conversely, a positive correlation between CAI and axis 1 in *H. brasiliensis*, *H. camargoana*, and *H. spruceana* indicates different regulatory patterns. Overall, ENc and CAI are negatively correlated with axis 1, while GC3 is positively correlated with ENc and L_{aa} (Figure 9). These patterns suggest that

gene groups enhance adaptability through multiple factors, reflecting different strategies to adjust gene expression and adapt to environmental and evolutionary pressures.

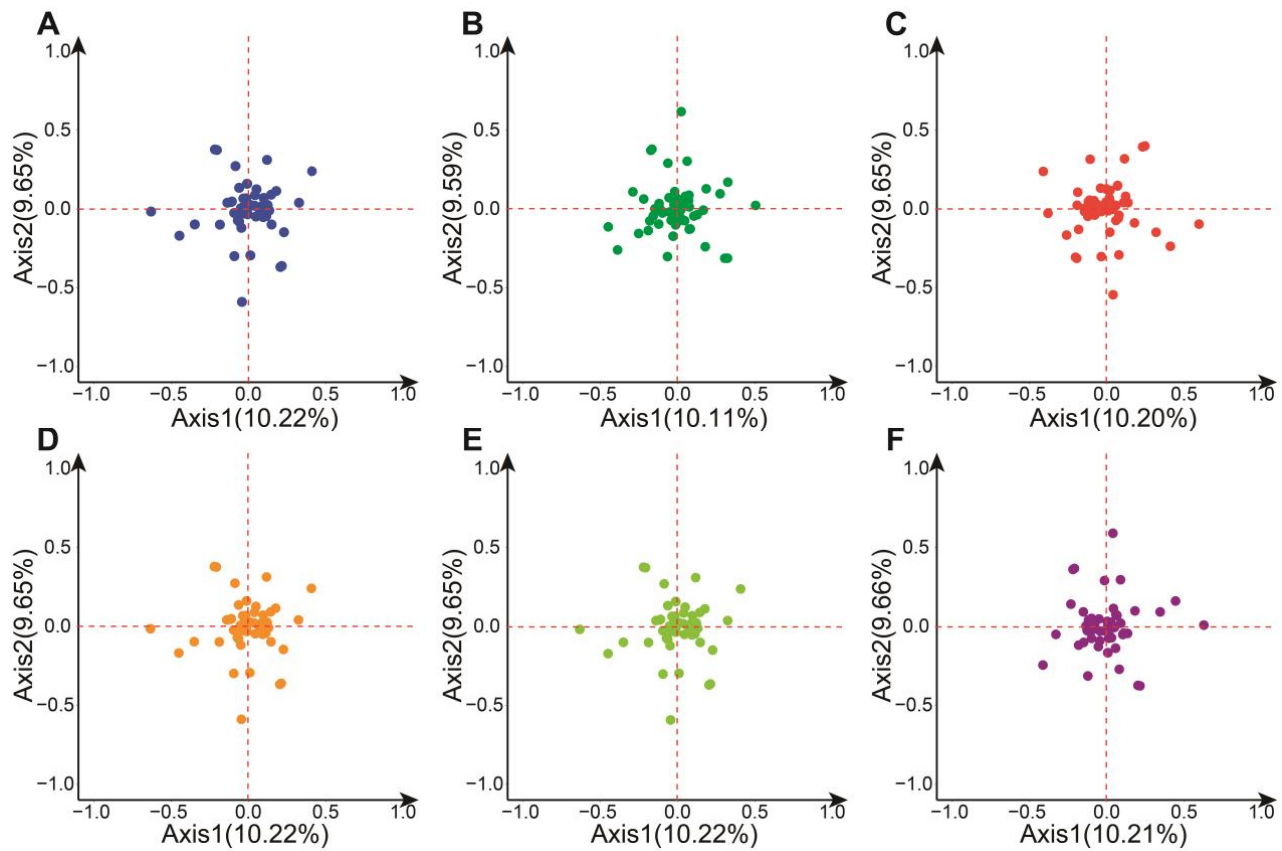


Figure 8. COA of *Hevea* chloroplast genome. (A) *H. benthamiana*; (B) *H. brasiliensis*; (C) *H. camargoana*; (D) *H. nitida*; (E) *H. pauciflora*; (F) *H. spruceana*.

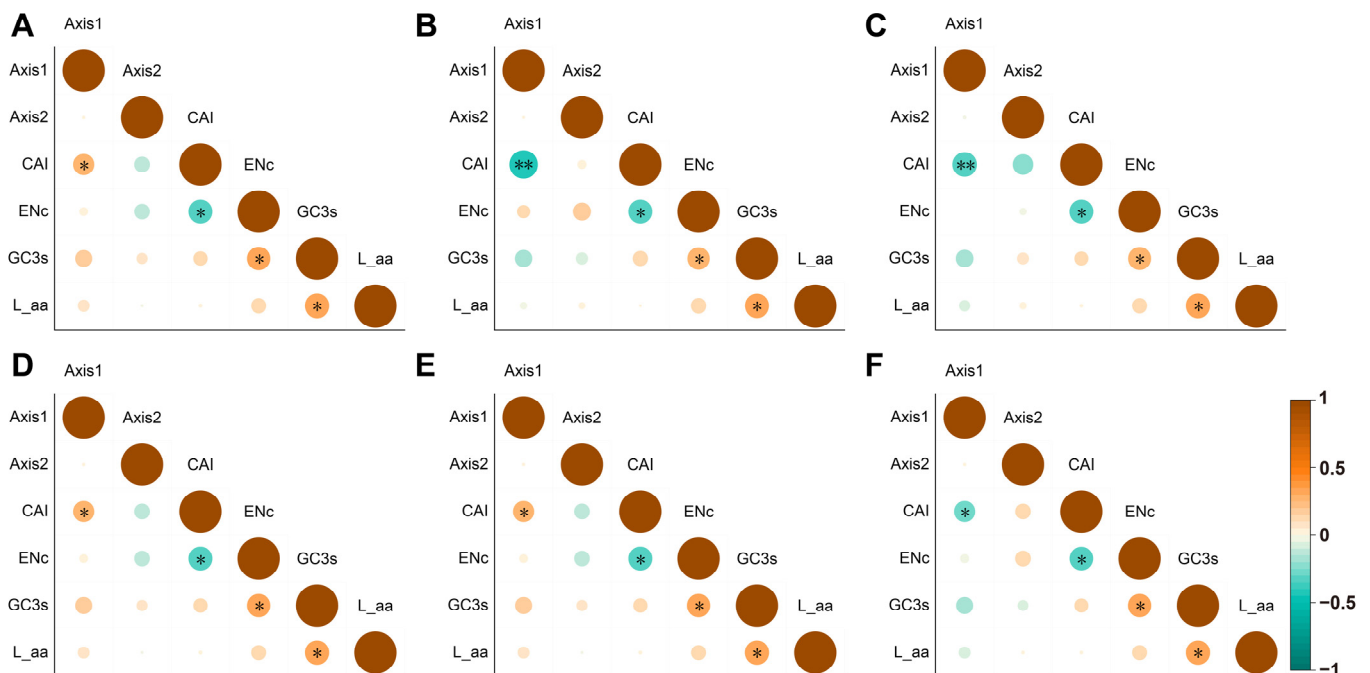


Figure 9. COA of axis 1, axis 2, and codon usage indices of the cpDNAs in six *Hevea* species. (A) *H. benthamiana*; (B) *H. brasiliensis*; (C) *H. camargoana*; (D) *H. nitida*; (E) *H. pauciflora*; (F) *H. spruceana*. * Indicates $p < 0.05$, ** indicates $p < 0.01$.

3.10. Optimal Codons

Further examination of the coding sequences of *Hevea* chloroplasts revealed 30 codons with an RSCU value not exceeding 1.00, most of which (29) end in G or C. In contrast, 29 codons have an RSCU value greater than 1.00, most of which (28) end in A or U. These findings clearly indicate a higher preference for codons ending in A or U in *Hevea* cpDNAs compared to those ending in G or C (Figure 10A). Based on the criterion of $RSCU > 1$, 30 high-frequency codons were identified in *Hevea* species, with 48.28% ending in U, 48.28% in A, and only one ending in G. These data emphasize a significant preference for codons ending in U/A in the *Hevea* chloroplast genome (Table S1). A total of 31 optimal codons were identified based on $RSCU > 1$ and $\Delta RSCU \geq 0.08$, with 22 conserved across all six species. In the *Hevea* genus, 31 optimal codons were identified. Among the six *Hevea* cpDNAs, 22 common codons were identified, including Ala (GCA), Ala (GCC), Ala (GCG), Cys (UGC), Gly (GGA), Gly (GGG), His (CAU), Ile (AUA), Lys (AAG), Leu (CUC), Leu (CUG), Asn (AAU), Pro (CCA), Pro (CCC), Arg (AGA), Arg (CGA), Arg (CGG), Ser (UCA), Ser (UCC), Ser (UCG), Thr (ACG), and Val (GUG). Among these, Ser (UCA), Ser (UCG), Asn (AAU), Ala (GCC), Gly (GGA), Val (GUG), Leu (CUG), Cys (UGC), Gly (GGG), Arg (AGA), and Pro (CCC) have $\Delta RSCU \geq 0.3$, while Leu (CUG), Cys (UGC), Gly (GGG), Arg (AGA), and Pro (CCC) have $\Delta RSCU \geq 0.5$ (Figure 10B, Table S2).

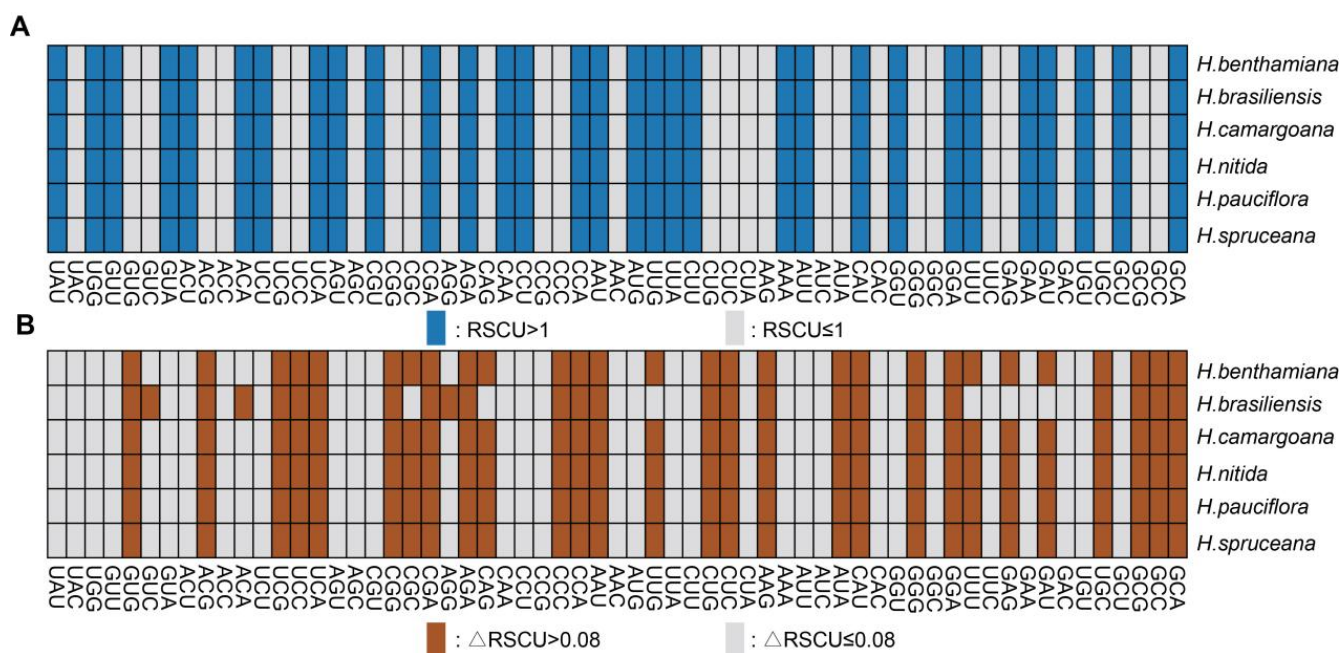


Figure 10. RSCU (A) and $\Delta RSCU$ (B) of six *Hevea* species.

3.11. Phylogenetic Analysis of *Hevea* cpDNAs

A detailed analysis of the six complete chloroplast genome datasets was performed to construct a phylogenetic tree of the *Hevea* genus. *Deutzianthus tonkinensis*, another member of the *Euphorbiaceae* family, was utilized as an outgroup to build the phylogenetic tree alongside the cpDNAs of the six *Hevea* species. *H. brasiliensis* and *H. camargoana* showed relatively close genetic relationships. In another clade, *H. spruceana*, *H. benthamiana*, *H. nitida*, and *H. pauciflora* formed a closely related group (Figure 11). Constructing this phylogenetic tree not only enhanced our understanding of the genetic diversity and relationships within the *Hevea* genus but also provided essential scientific evidence for future conservation, genetic improvement, and sustainable use of germplasm resources.

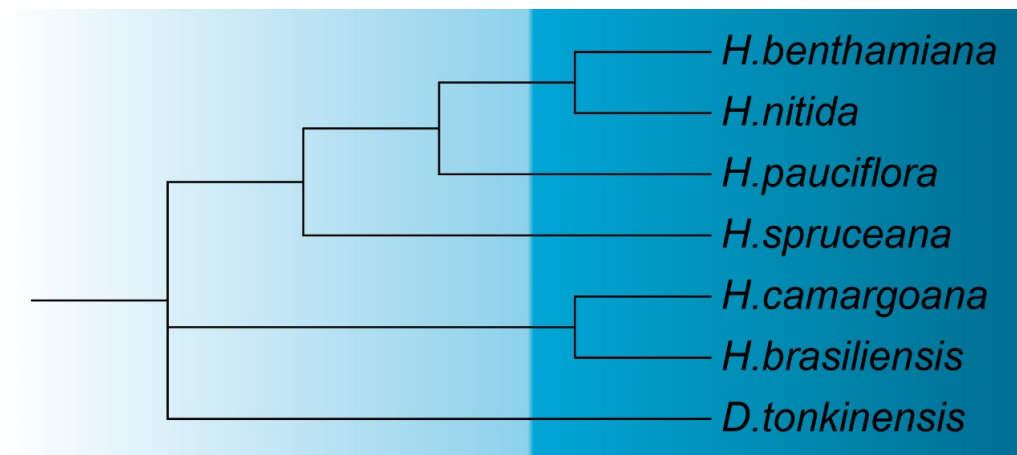


Figure 11. Phylogenetic analysis of the cpDNAs of the *Hevea* genus.

3.12. Structural and Quantitative Analysis of Photosystem Proteins

Figure 12A illustrates the predicted three-dimensional (3D) structures of Photosystem I (PSI) and Photosystem II (PSII) proteins in *H. brasiliensis*, based on SWISS-MODEL structural modeling. The core subunits of PSI (e.g., PsaA, PsaB, PsaC) and PSII (e.g., PsbA, PsbD, PsbC) were successfully identified, alongside auxiliary subunits such as PsaI (PSI) and PsbE (PSII). This analysis highlights the high level of conservation of photosystem protein structures in *H. brasiliensis*, which is consistent with observations in other plant species. Moreover, the conserved nature of PSI and PSII subunits suggests that the evolutionary dynamics of the chloroplast genome have preserved essential genes for photosynthetic function, maintaining their structural integrity and functional relevance. Figure 12B provides a quantitative analysis of the protein abundance of PSI and PSII subunits based on chloroplast proteomic data, visualized through a heatmap. The data reveal that core PSI subunits, particularly PsaA and PsaB, exhibit the highest abundance, underscoring their crucial roles in light energy capture and electron transfer. Similarly, PSII subunits PsbA and PsbD, integral to oxygen-evolving activity and electron transfer, also show consistently high abundance across chloroplast samples. These findings align with the transcriptional and translational priorities encoded in the *H. brasiliensis* chloroplast genome, where genes for core photosynthetic components (e.g., *psaA*, *psaB*, *psbA*, and *psbD*) are not only highly conserved but also encoded directly in the chloroplast genome to ensure efficient regulation and expression. Notably, the analysis suggests potential interplay between the chloroplast genome and nuclear genome in regulating photosystem assembly. While core subunits are encoded by the chloroplast genome, many auxiliary proteins and regulatory factors are encoded by the nuclear genome and imported into the chloroplast. This division of genetic responsibility reflects the co-evolution of the chloroplast and nuclear genomes to optimize photosynthetic efficiency. Furthermore, the protein abundance patterns suggest differential regulatory mechanisms between PSI and PSII, possibly reflecting distinct roles in light harvesting under varying environmental conditions. In conclusion, this integrated analysis of chloroplast proteomics and structural prediction underscores the critical roles of PSI and PSII subunits in *H. brasiliensis*, as well as the evolutionary and functional significance of the chloroplast genome in sustaining photosynthetic efficiency. Future studies may focus on how environmental stressors influence the expression and abundance of these subunits, providing further insights into the adaptive mechanisms of chloroplast function.

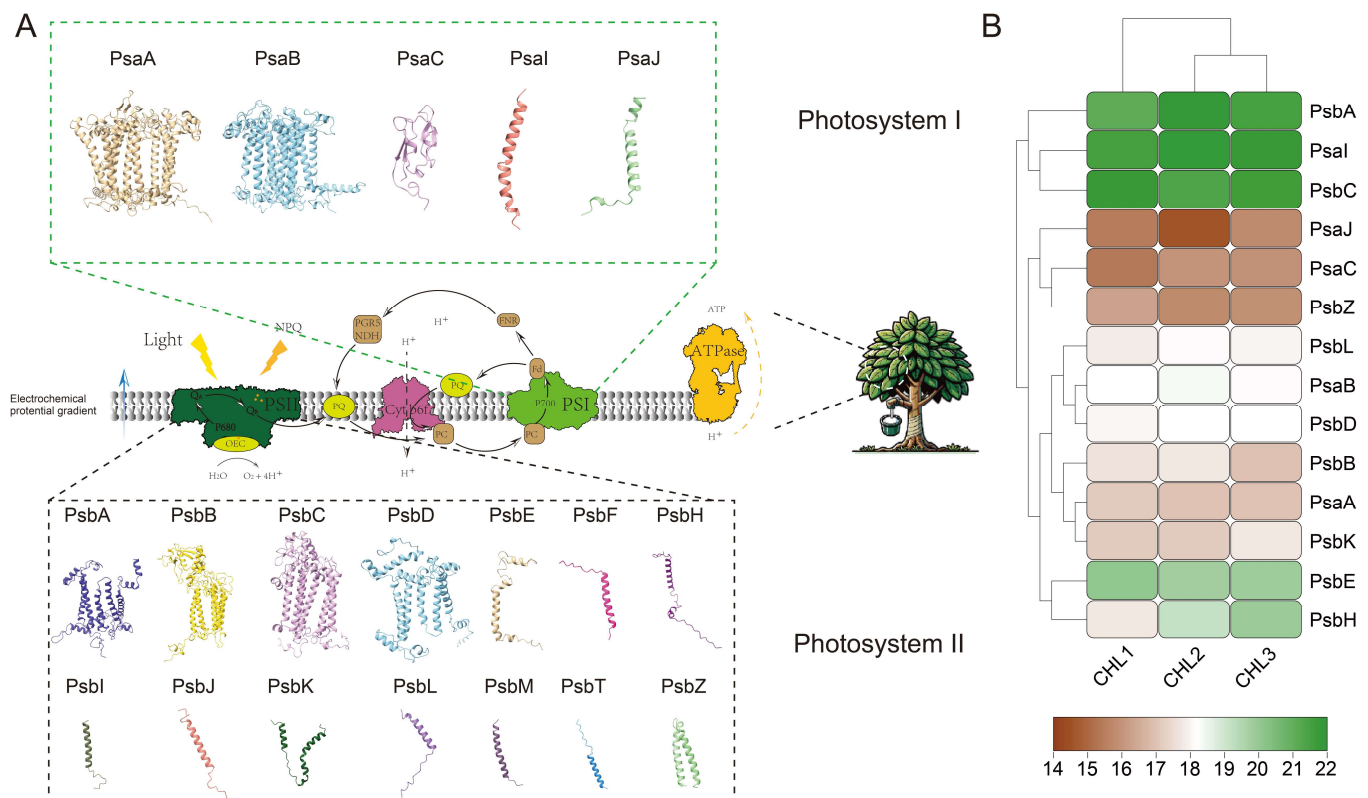


Figure 12. Three-dimensional (3D) Structure Prediction and Abundance Analysis of PSI and PSII Proteins in *H. brasiliensis*. **(A)** The 3D structures of core proteins from PSI and PSII were predicted using SWISS-MODEL. **(B)** Heatmap of protein abundance levels for major subunits of PSI and PSII based on chloroplast proteomic data. Different colors indicate variations in protein abundance, with green representing high abundance and brown representing low abundance.

4. Discussion

CpDNAs within the *Hevea* genus display remarkable sequence similarity, with internal repeat regions exhibiting higher conservation compared to single-copy regions. This level of conservation is vital for preserving genomic stability and ensuring proper chloroplast functionality [31]. Similar observations have been made in other plant genera, such as the Poaceae, where conserved repeat regions are essential for chloroplast genome stability [32]. Although coding regions are relatively consistent, non-coding regions show greater variability [33]. For example, in the *Brassicaceae*, the trnH-psbA intergenic spacer is frequently employed in phylogenetic studies due to its variability [34]. Our findings, which show no gene rearrangements and complete collinearity among the six genomes, are consistent with previous research highlighting the stability of cpDNAs in plants [33]. Research on the *Fabaceae* family supports this stability, with minimal chloroplast genome rearrangements observed [35]. Our data suggest that the GC content of *Hevea* cpDNAs exceeds 35%, with genome lengths varying from 161,093 bp (*H. spruceana*) to 161,254 bp (*H. camargoana*). The lengths of the LSC, SSC, IRA, and IRB regions show slight variation among species. The overall GC content, ranging from 35.72% to 35.75%, aligns with trends seen in other plant species, where GC content is critical for genome stability and function [32,36]. In the *Asteraceae* family, similar GC content trends correlate with genome stability [37]. The highly conserved sequences, with 91 to 92 protein-coding genes, further underscores the evolutionary stability of *Hevea* cpDNAs. This finding is consistent with the *Solanaceae* family, where conserved gene numbers are essential for chloroplast functionality [38].

Our research indicates that *H. camargoana* has the longest chloroplast genome, while *H. spruceana* has the shortest. The patterns of IR region expansion and contraction among *Hevea* species suggest adaptive functions during their evolution [39]. Similar patterns have been observed in the *Orchidaceae* family, where IR expansion and contraction significantly impact genome size [40]. Expanding IR regions can enhance genome stability and increase gene numbers, contributing to functional expression and regulatory mechanisms in chloroplasts [41]. However, excessive IR expansion can lead to genomic instability, warranting further research to understand its specific impacts on *Hevea*'s physiological functions [42]. Our analysis identified six types of SSRs in *Hevea* cpDNAs, with cpSSRs numbers ranging from 126 to 137. Mononucleotide repeats are the most prevalent, consistent with findings in other plant species where SSRs are crucial for genetic diversity and adaptability [31]. For instance, in the *Rosaceae* family, SSRs have been identified as key markers for genetic diversity studies [43]. The predominance of A/T repeats over G/C repeats aligns with the high AT content observed in *Hevea* cpDNAs and other plants [33]. This trend is also seen in the *Fabaceae* family, where high AT content is associated with repeat regions [35].

Hevea cpDNAs exhibit a marked preference for codons ending in A or U, with Leu predominantly encoded by the UUA codon. This A/U-ending codon bias is prevalent in cpDNAs and can improve translational efficiency [36,44]. In the *Poaceae* family, a similar bias toward A/U-ending codons has been noted, enhancing protein synthesis efficiency [45]. ENc-plot analysis indicates that natural selection significantly influences codon usage bias, with some genes showing weak bias influenced by base variations [46,47]. This bias is likely an adaptive mechanism to optimize protein synthesis within the chloroplast's environment. In the *Asteraceae* family, similar adaptive mechanisms for codon usage bias have been proposed [37]. PR2-bias plot analysis of *Hevea* cpDNAs shows a preference for T/C at the third codon position, suggesting that codon selection patterns are influenced by factors beyond mutational pressure, such as natural selection [36]. Neutrality plot underscores the complex interplay between mutational pressure and natural selection in shaping codon usage patterns in *Hevea* cpDNAs. COA based on RSCU values indicates significant differences in codon usage patterns among *Hevea* chloroplast genes with varying functions. This dispersed distribution suggests that codon usage bias is influenced by multiple factors [46]. The identification of 31 optimal codons, most ending in A or U, further highlights the preference for these codons in *Hevea* cpDNAs. The identification of 31 optimal codons, primarily ending in A or U, provides a valuable resource for genetic engineering. Optimizing gene expression in the transgenic *Hevea* genus could be achieved by selecting codons with high translational efficiency, thereby enhancing protein synthesis in chloroplasts.

Our phylogenetic analysis of *Hevea* cpDNAs reveals a close genetic relationship between *H. brasiliensis* and *H. camargoana*, while *H. spruceana*, *H. benthamiana*, *H. nitida*, and *H. pauciflora* form another closely related cluster. This phylogenetic arrangement offers critical insights into the genetic diversity and evolutionary relationships within the *Hevea* genus, aiding future conservation and genetic enhancement initiatives [7]. Similar phylogenetic structures are observed in the *Cucurbitaceae* family, where close genetic relationships guide conservation strategies [48]. This research lays a theoretical groundwork for genomic studies and chloroplast genetic engineering in *Hevea*. By examining gene variations and codon usage patterns, we have deepened our understanding of photosynthesis characteristics and adaptability in *Hevea* species, particularly in low-temperature environments [49]. Future studies should investigate the functional implications of specific gene variations and explore the potential of genetic engineering to introduce advantageous traits into *Hevea* cpDNAs, utilizing advanced genomic and transcriptomic techniques [50]. Similar

methods have been effectively used in the Solanaceae family to enhance crop resilience and productivity [51].

5. Conclusions

We performed comparative genomic and phylogenetic analyses of chloroplasts in the *Hevea* genus. The chloroplast genome sizes of *Hevea* species range between 161,093 bp and 161,254 bp, featuring a typical four-part structure: the LSC ranging from 89,079 to 89,281 bp, the SSC spanning 18,362 to 18,376 bp, and two IRs measuring between 26,810 and 26,819 bp. A total of 135 to 137 genes were identified. The chloroplast sequences show high conservation, with significant sequence variations at the IRs/SSC junctions. Codon preference in the cpDNAs of the six *Hevea* species favors A/U endings, with ENc values all exceeding 45, suggesting a weak codon usage bias. Thirty-one optimal codons were pinpointed within the cpDNAs of these species. Phylogenetic analysis revealed that the *Hevea* genus is a monophyletic group, providing a theoretical foundation for genetic diversity studies and the development and utilization of specific genes within the *Hevea* chloroplast genome.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/genes16020201/s1>; Supplementary File S1: Table S1. RSCU values of codons in the chloroplast genomes of six *Hevea* species. Supplementary File S2: Table S2. RSCU and Δ RSCU values in the high and low expression libraries in the chloroplast genomes of six *Hevea* species.

Author Contributions: Y.Y.: Original draft, Software, Data curation, Conceptualization. L.H., B.Y. and F.F.: Review and editing, Visualization. X.L. and M.W.: Methodology, Data curation. Z.L., A.L., M.H. and S.H.: Review and editing. C.L.: Software, Visualization. W.W. and X.W.: Review and editing, Funding acquisition, Conceptualization. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Key Research and Development Program of China (No. 2021YFA1300401 and 2021YFA1300403), the Innovation Center for Academicians of Hainan Province and the Specific Research Fund of the Innovation Platform for Academicians of Hainan Province (No. YSPTZX202309), the CAS Project for Young Scientists in Basic Research (YSBR-093), and the Special Fund of Guizhou University (LJ2023-03).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original contributions presented in the study are included in the article/Supplementary Materials, further inquiries can be directed to the corresponding author.

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

References

1. Venkatachalam, P.; Geetha, N.; Sangeetha, P.; Thulaseedharan, A. Natural rubber producing plants: An overview. *Afr. J. Biotechnol.* **2013**, *12*, 12.
2. Priyadarshan, P.M. Molecular markers to devise predictive models for juvenile selection in *Hevea* rubber. *Plant Breed.* **2022**, *141*, 159–183. [[CrossRef](#)]
3. Ali, M.F.; Akber, M.A.; Smith, C.; Aziz, A.A. The dynamics of rubber production in Malaysia: Potential impacts, challenges and proposed interventions. *For. Policy Econ.* **2021**, *127*, 102449. [[CrossRef](#)]
4. Abdel-Haleem, H.; Luo, Z.; Ray, D. Genetic improvement of guayule (*Parthenium argentatum* A. Gray): An alternative rubber crop. *Adv. Plant Breed. Strateg. Ind. Food Crops* **2019**, *6*, 151–178.
5. Chen, H.; Yi, Z.-F.; Schmidt-Vogt, D.; Ahrends, A.; Beckschäfer, P.; Kleinn, C.; Ranjitkar, S.; Xu, J. Pushing the limits: The pattern and dynamics of rubber monoculture expansion in Xishuangbanna, SW China. *PLoS ONE* **2016**, *11*, e0150062. [[CrossRef](#)]

6. Tangphatsornruang, S.; Uthaisaisanwong, P.; Sangsrakru, D.; Chanprasert, J.; Yoocha, T.; Jomchai, N.; Tragoonrung, S. Characterization of the complete chloroplast genome of *Hevea brasiliensis* reveals genome rearrangement, RNA editing sites and phylogenetic relationships. *Gene* **2011**, *475*, 104–112. [\[CrossRef\]](#) [\[PubMed\]](#)
7. Feng, L.-Y.; Liu, J.; Gao, C.-W.; Wu, H.-B.; Li, G.-H.; Gao, L.-Z. Higher genomic variation in wild than cultivated rubber trees, *Hevea brasiliensis*, revealed by comparative analyses of chloroplast genomes. *Front. Ecol. Evol.* **2020**, *8*, 237. [\[CrossRef\]](#)
8. Hu, Y.-S.; Huang, H.-S.; Liu, J. Characterization of the complete chloroplast genome of *Hevea pauciflora* (Euphorbiaceae), an important wild relative of the rubber tree. *Mitochondrial DNA Part B* **2022**, *7*, 1589–1593. [\[CrossRef\]](#) [\[PubMed\]](#)
9. Niu, Y.-F.; Hu, Y.-S.; Zheng, C.; Liu, Z.-Y.; Liu, J. The complete chloroplast genome of *Hevea camargoana*. *Mitochondrial DNA Part B* **2020**, *5*, 607–608. [\[CrossRef\]](#)
10. Fu, Y.; Brookfield, H.; Guo, H.; Chen, J.; Chen, A.; Cui, J. Smallholder rubber plantation expansion and its impact on local livelihoods, land use and agrobiodiversity, a case study from Daka, Xishuangbanna, southwestern China. *Int. J. Sustain. Dev. World Ecol.* **2009**, *16*, 22–29. [\[CrossRef\]](#)
11. Schoch, C.L.; Ciufo, S.; Domrachev, M.; Hotton, C.L.; Kannan, S.; Khovanskaya, R.; Leipe, D.; Mcveigh, R.; O'Neill, K.; Robbertse, B. NCBI Taxonomy: A comprehensive update on curation, resources and tools. *Database* **2020**, *2020*, baaa062. [\[CrossRef\]](#)
12. Kearse, M.; Moir, R.; Wilson, A.; Stones-Havas, S.; Cheung, M.; Sturrock, S.; Buxton, S.; Cooper, A.; Markowitz, S.; Duran, C. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **2012**, *28*, 1647–1649. [\[CrossRef\]](#)
13. Deschenes, L.A.; David, A. Vanden Bout University of Texas, A. Origin 6.0: Scientific Data Analysis and Graphing Software Origin Lab Corporation (Formerly Microcal Software, Inc.). Commercial Price: 595. Academic price: 446. 2000. Available online: www.originlab.com (accessed on 26 June 2024).
14. Amiryousefi, A.; Hyvönen, J.; Poczai, P. IRscope: An online program to visualize the junction sites of chloroplast genomes. *Bioinformatics* **2018**, *34*, 3030–3031. [\[CrossRef\]](#)
15. Darling, A.C.; Mau, B.; Blattner, F.R.; Perna, N.T. Mauve: Multiple alignment of conserved genomic sequence with rearrangements. *Genome Res.* **2004**, *14*, 1394–1403. [\[CrossRef\]](#) [\[PubMed\]](#)
16. Rozewicki, J.; Li, S.; Amada, K.M.; Standley, D.M.; Katoh, K. MAFFT-DASH: Integrated protein sequence and structural alignment. *Nucleic Acids Res.* **2019**, *47*, W5–W10. [\[CrossRef\]](#) [\[PubMed\]](#)
17. Frazer, K.A.; Pachter, L.; Poliakov, A.; Rubin, E.M.; Dubchak, I. VISTA: Computational tools for comparative genomics. *Nucleic Acids Res.* **2004**, *32*, W273–W279. [\[CrossRef\]](#) [\[PubMed\]](#)
18. Ahmad, A.; Wang, J.-D.; Pan, Y.-B.; Sharif, R.; Gao, S.-J. Development and use of simple sequence repeats (SSRs) markers for sugarcane breeding and genetic studies. *Agronomy* **2018**, *8*, 260. [\[CrossRef\]](#)
19. Beier, S.; Thiel, T.; Münch, T.; Scholz, U.; Mascher, M. MISA-web: A web server for microsatellite prediction. *Bioinformatics* **2017**, *33*, 2583–2585. [\[CrossRef\]](#)
20. 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. [\[CrossRef\]](#)
21. Peden, J.F. Analysis of Codon Usage. Ph.D. Thesis, University of Nottingham, Nottingham, UK, 2000.
22. Rice, P.; Longden, I.; Bleasby, A. EMBOSS: The European molecular biology open software suite. *Trends Genet.* **2000**, *16*, 276–277. [\[CrossRef\]](#) [\[PubMed\]](#)
23. Ihaka, R.; Gentleman, R. R: A language for data analysis and graphics. *J. Comput. Graph. Stat.* **1996**, *5*, 299–314. [\[CrossRef\]](#)
24. Sharp, P.M.; Li, W.-H. The codon adaptation index—a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res.* **1987**, *15*, 1281–1295. [\[CrossRef\]](#)
25. Wang, L.; Xing, H.; Yuan, Y.; Wang, X.; Saeed, M.; Tao, J.; Feng, W.; Zhang, G.; Song, X.; Sun, X. Genome-wide analysis of codon usage bias in four sequenced cotton species. *PLoS ONE* **2018**, *13*, e0194372. [\[CrossRef\]](#)
26. Li, R.; Zhang, X.; Ma, X.; Guo, R.; Yan, S.; Jin, G.; Zhou, P. Analyses of Codon Usage Patterns and Codon Usage Bias in Peach (*Prunus persica*). *Mol. Plant Breed.* **2022**, *13*, 28. [\[CrossRef\]](#)
27. Feng, Z.; Jiang, Y.; Zheng, Y.; MIAO, Y.-j.; HUANG, L.-f.; LUO, G.-m. Codon use bias analysis of chloroplast genome of *Cistanche*. *Chin. Tradit. Herb. Drugs* **2023**, *54*, 1540–1550.
28. Li, Q.; Luo, Y.; Sha, A.; Xiao, W.; Xiong, Z.; Chen, X.; He, J.; Peng, L.; Zou, L. Analysis of synonymous codon usage patterns in mitochondrial genomes of nine *Amanita* species. *Front. Microbiol.* **2023**, *14*, 1134228. [\[CrossRef\]](#)
29. Paul, P.; Malakar, A.K.; Chakraborty, S. Compositional bias coupled with selection and mutation pressure drives codon usage in *Brassica campestris* genes. *Food Sci. Biotechnol.* **2018**, *27*, 725–733. [\[CrossRef\]](#)
30. RoyChoudhury, S.; Mukherjee, D. A detailed comparative analysis on the overall codon usage pattern in herpesviruses. *Virus Res.* **2010**, *148*, 31–43. [\[CrossRef\]](#)
31. Powell, W.; Morgante, M.; McDevitt, R.; Vendramin, G.; Rafalski, J. Polymorphic simple sequence repeat regions in chloroplast genomes: Applications to the population genetics of pines. *Proc. Natl. Acad. Sci. USA* **1995**, *92*, 7759–7763. [\[CrossRef\]](#) [\[PubMed\]](#)

32. Wicke, S.; Schneeweiss, G.M.; Depamphilis, C.W.; Müller, K.F.; Quandt, D. The evolution of the plastid chromosome in land plants: Gene content, gene order, gene function. *Plant Mol. Biol.* **2011**, *76*, 273–297. [[CrossRef](#)]
33. Provan, J.; Powell, W.; Hollingsworth, P.M. Chloroplast microsatellites: New tools for studies in plant ecology and evolution. *Trends Ecol. Evol.* **2001**, *16*, 142–147. [[CrossRef](#)] [[PubMed](#)]
34. Shaw, J.; Lickey, E.B.; Schilling, E.E.; Small, R.L. Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: The tortoise and the hare III. *Am. J. Bot.* **2007**, *94*, 275–288. [[CrossRef](#)]
35. Jansen, R.K.; Cai, Z.; Raubeson, L.A.; Daniell, H.; Depamphilis, C.W.; Leebens-Mack, J.; Müller, K.F.; Guisinger-Bellian, M.; Haberle, R.C.; Hansen, A.K. Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 19369–19374. [[CrossRef](#)] [[PubMed](#)]
36. Palmer, J.D. Chloroplast DNA evolution and biosystematic uses of chloroplast DNA variation. *Am. Nat.* **1987**, *130*, S6–S29. [[CrossRef](#)]
37. Timme, R.E.; Kuehl, J.V.; Boore, J.L.; Jansen, R.K. A comparative analysis of the *Lactuca* and *Helianthus* (Asteraceae) plastid genomes: Identification of divergent regions and categorization of shared repeats. *Am. J. Bot.* **2007**, *94*, 302–312. [[CrossRef](#)]
38. Daniell, H.; Kumar, S.; Dufourmantel, N. Breakthrough in chloroplast genetic engineering of agronomically important crops. *TRENDS Biotechnol.* **2005**, *23*, 238–245. [[CrossRef](#)] [[PubMed](#)]
39. Ikeda, Y.; Kato, A.; Kohjiya, S.; Nakajima, Y. *Rubber Science*; Springer: Singapore, 2018; Volume 10, pp. 978–981.
40. Liu, H.; Ye, H.; Zhang, N.; Ma, J.; Wang, J.; Hu, G.; Li, M.; Zhao, P. Comparative analyses of chloroplast genomes provide comprehensive insights into the adaptive evolution of *Paphiopedilum* (Orchidaceae). *Horticulturae* **2022**, *8*, 391. [[CrossRef](#)]
41. Guisinger, M.M.; Kuehl, J.V.; Boore, J.L.; Jansen, R.K. Genome-wide analyses of Geraniaceae plastid DNA reveal unprecedented patterns of increased nucleotide substitutions. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 18424–18429. [[CrossRef](#)] [[PubMed](#)]
42. Raubeson, L.A.; Jansen, R.K. Chloroplast DNA evidence on the ancient evolutionary split in vascular land plants. *Science* **1992**, *255*, 1697–1699. [[CrossRef](#)]
43. Qu, J.; Liu, J. A genome-wide analysis of simple sequence repeats in maize and the development of polymorphism markers from next-generation sequence data. *BMC Res. Notes* **2013**, *6*, 403. [[CrossRef](#)]
44. 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](#)]
45. Morton, B.R. Selection on the codon bias of *Chlamydomonas reinhardtii* chloroplast genes and the plant *psbA* gene. *J. Mol. Evol.* **1996**, *43*, 28–31. [[CrossRef](#)]
46. Sharp, P.M.; Emery, L.R.; Zeng, K. Forces that influence the evolution of codon bias. *Philos. Trans. R. Soc. B Biol. Sci.* **2010**, *365*, 1203–1212. [[CrossRef](#)] [[PubMed](#)]
47. Hershberg, R.; Petrov, D.A. Selection on codon bias. *Annu. Rev. Genet.* **2008**, *42*, 287–299. [[CrossRef](#)] [[PubMed](#)]
48. Cui, J.; Yang, Y.; Luo, S.; Wang, L.; Huang, R.; Wen, Q.; Han, X.; Miao, N.; Cheng, J.; Liu, Z. Whole-genome sequencing provides insights into the genetic diversity and domestication of bitter melon (*Momordica* spp.). *Hortic. Res.* **2020**, *7*, 85. [[CrossRef](#)]
49. Priyadarshan, P. Genetic diversity and erosion in *Hevea* rubber. *Genet. Divers. Eros. Plants Case Hist.* **2016**, *8*, 233–267.
50. 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](#)]
51. Ruf, S.; Hermann, M.; Berger, I.J.; Carrer, H.; Bock, R. Stable genetic transformation of tomato plastids and expression of a foreign protein in fruit. *Nat. Biotechnol.* **2001**, *19*, 870–875. [[CrossRef](#)] [[PubMed](#)]

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.