

RESEARCH

Open Access



Comprehensive analysis of chloroplast genome evolution in Poaceae: codon usage patterns, selection pressures, and phylogenomic relationships

Xiaohe Geng¹, Ying Xue¹, Haodi Wang¹, Hao Tu¹, Rui Jing¹, Zixiong Li¹ and Yunpeng Gai^{1*}

Abstract

Background The Poaceae is one of the world's most ecologically and economically important plant families, encompassing major food crops and forage grasses. Despite its agricultural significance, systematic analyses of chloroplast genome evolution across this diverse family remain limited. To understand adaptive evolution in Poaceae species and their phylogenetic relationships, analysis of codon usage and evolutionary pressure in chloroplast genomes is essential. We conducted a comprehensive comparative analysis of 48 Poaceae chloroplast genomes to elucidate codon usage mechanisms and phylogenetic relationships.

Results Our analysis revealed consistent codon usage bias across all species. Notably, 93.1% of preferred codons are A/U-ending, indicating the presence of strong compositional constraints. Chloroplast genomes ranged from 129,095 bp to 148,617 bp, with GC content ranging between 38.16% and 39.04%. The average GC content was 38.5%, with GC₃ content (30.51%) significantly lower than GC₁ (47.34%) and GC₂ (39.64%) contents. ENC-GC3s plots, PR2 analysis, and neutrality plot analyses consistently demonstrated that natural selection, rather than mutational pressure, predominantly drives codon usage patterns. Simple sequence repeats (SSRs) analysis showed non-random distribution, with the highest density in intergenic regions (52.6%), followed by introns (28.2%) and coding sequences (19.2%). Mononucleotide and octanucleotide repeats constituted the predominant SSR types. Ka/Ks analysis revealed that 94.67% of chloroplast genes undergo purifying selection (Ka/Ks < 0.5), with photosystem and ribosomal protein genes under the most stringent constraints. Phylogenetic analysis confirmed robust evolutionary relationships, with genome synteny strongly correlating with taxonomic classification.

Conclusion This study provides the most comprehensive analysis of Poaceae chloroplast genome evolution to date, revealing selection-driven codon optimization that balances functional conservation with adaptive flexibility. The predominance of purifying selection reflects the critical importance of chloroplast gene function, while SSR distribution patterns suggest roles in genome plasticity and regulatory functions. These findings deepen our understanding of grass evolution and provide valuable resources for molecular marker development, phylogenetic studies, and crop improvement programs addressing global food security challenges.

*Correspondence:

Yunpeng Gai
gaiyunpeng@bjfu.edu.cn

Full list of author information is available at the end of the article



© The Author(s) 2026. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

Keywords Chloroplast genomics, Poaceae, Codon usage bias, Natural selection, Purifying selection, Evolutionary genomics, Phylogenomics, Adaptive evolution, Simple sequence repeats, Comparative genomics

Background

The Poaceae is the fifth-largest family of flowering plants, comprising approximately 11,506 species distributed across 768 genera, 12 subfamilies, 52 tribes, and 90 subtribes [1, 2]. Phylogenetic analyses estimate the crown age of Poaceae at approximately 101 million years ago, with molecular dating ranging from 59 to 129 million years ago [3]. The family has undergone extensive diversification, resulting in remarkable species diversity [4], a global distribution across all climatic zones, and significant ecological roles in terrestrial ecosystems [5, 6]. Poaceae includes the world's most economically important staple crops, such as rice, wheat, and maize, as well as numerous high-quality forage grasses and species used in ecological restoration [7, 8]. The agricultural significance of the Poaceae family extends beyond staple grain production to include diverse forage systems that support global livestock industries. Intermediate wheatgrass (*Thinopyrum intermedium*) exhibits excellent drought resistance and soil aggregation properties in arid regions, and can increase yield by 40–50% when intercropped with alfalfa (*Medicago sativa*) [9, 10]. Given the challenges to food security brought by global climate change and population growth, in-depth research on the genetic basis and adaptive evolutionary mechanisms of Poaceae has become increasingly important [11].

Eragrostis, a globally distributed genus of Poaceae, comprises approximately 350 species that are native to Eurasia and Africa and extends into regions such as China, India, and Myanmar [12–15]. This genus is characterized by rapid growth, a deep root system, and high drought tolerance, which makes it valuable for soil conservation and desertification control [16]. As a predominantly C_4 genus, *Eragrostis* exhibits superior photosynthetic efficiency and biomass production compared to that of C_3 grasses. Besides, its young leaves are rich in protein and fiber, rendering it a premium forage resource highly valued in livestock feed systems [17]. Thus, given the family's role as a primary source of staple crops, forage grasses, and vital ecosystem stabilizers, research on Poaceae holds practical significance for global food security, ecological conservation, and sustainable agriculture.

Chloroplasts, as semi-autonomous organelles within plant cells, perform multiple critical functions, including photosynthesis, amino acid synthesis, nucleotide production, and fatty acid formation [18–20]. According to the widely accepted endosymbiotic theory, chloroplasts originated from an ancient cyanobacterium that established endosymbiotic relationships with eukaryotic host cells through billions of years of co-evolution [21–24].

Chloroplast genomes typically exhibit a highly conserved circular quadripartite structure comprising two inverted repeat regions (IRA and IRB), a small single-copy region (SSC), and a large single-copy region (LSC), which remain highly conserved among angiosperms [19, 25, 26]. Chloroplast genomes are 120–180 kb long and possess 120–130 genes, including the essential components of the photosynthetic apparatus: photosystem core proteins, ATP synthase subunits, NADH-dehydrogenase complexes, and ribosomal proteins [27–29]. The relatively small size, conserved genomic architecture, and maternal inheritance characteristics of chloroplast genomes make them ideal molecular markers for phylogenetic analysis, species identification, and evolutionary studies [30, 31]. In recent years, the rapid development of high-throughput sequencing technologies has greatly facilitated chloroplast genomics research, making large-scale comparative genomic analyses possible [4, 32–34]. In Poaceae, chloroplast genome analysis offers a dual advantage for evolutionary biology. Its conserved genomic architecture facilitates robust phylogenetic inference across deep taxonomic scales, while lineage-specific features, such as structural variations and codon usage patterns, provide insights into ecological diversification [35–37].

Codon usage bias (CUB) represents a fundamental aspect of molecular evolution, indicating the non-random selection of synonymous codons during protein synthesis [38, 39]. In the standard genetic code, except for methionine and tryptophan, the remaining 18 amino acids are encoded by 2–6 synonymous codons, and the usage frequencies of these synonymous codons in genomes often exhibit non-random distributions [40–44]. The formation of codon usage patterns is influenced by multiple evolutionary forces, including natural selection, genetic drift, mutational pressure, and gene conversion, as described by the widely-accepted mutation-selection-drift balance model [39, 44–46]. CUB significantly influences gene expression and cellular function, affecting RNA processing, protein translation, and protein folding [47–49]. In chloroplast genomes, due to their unique genetic systems and functional limitations, codon usage patterns often exhibit characteristics that differ from those of nuclear genomes [50, 51]. The translational selection theory suggests that selection pressure for translational efficiency and accuracy is the main driving force shaping codon usage patterns, and this pressure is significant in highly expressed genes [52, 53]. Previous studies have shown that optimal codons typically correspond to the most abundant tRNA molecules in cells, thus ensuring efficient protein synthesis [54–56]. Furthermore, CUB is closely

related to factors such as gene expression levels, protein folding efficiency, and mRNA stability [49, 57, 58]. In plant chloroplasts, due to the high expression requirements of photosynthesis-related genes, codon usage patterns often exhibit strong translational selection signals [50, 51, 59]. Therefore, investigating codon usage patterns in chloroplast genomes is essential for understanding the complex interplay of natural selection, mutational pressure, and translational efficiency that shapes gene expression and drives adaptive evolution in plants. This is particularly important within the context of chloroplast-specific functional constraints and maternal inheritance systems.

Although Poaceae is central to global ecosystems and agriculture, systematic studies on chloroplast genome codon usage patterns in this family remain limited. Most existing research has concentrated on single-species genomic analyses, exemplified by studies on switchgrass (*Panicum virgatum*) [60, 61] and perennial ryegrass (*Lolium perenne*) [62, 63] that investigated chloroplast genome structure, identified SSR markers, and performed phylogenetic analysis to resolve interspecific relationships. While these studies provided species-specific genomic details, including insights into CUB and SSR distribution, they lack cross-species comparative insights, thus hindering understanding of the overall evolutionary patterns and adaptive strategies of Poaceae. Recent chloroplast genome phylogenies have revised relationships among Bambusoideae, Ehrhartoideae, and Pooideae, challenging traditional taxonomy [64], but large-scale studies on CUB across Poaceae species are still lacking. With the increasing availability of chloroplast genome sequences, systematic comparative analyses are now feasible. For *Eragrostis*, as important forage and ecological restoration plants, in-depth study of their chloroplast codon usage patterns will clarify C_4 plant adaptive evolution. Thus, multi-species comparisons of Poaceae chloroplast codon usage can fill research gaps and inform evolutionary biology and breeding.

Selection pressure analysis is a core method for understanding gene evolution dynamics and adaptive processes. By comparing the ratio of synonymous substitution rates (dS) to nonsynonymous substitution rates (dN), commonly referred to as the Ka/Ks ratio, researchers can identify the selective forces acting on protein-coding genes [65, 66]. When Ka/Ks < 1, genes exhibit purifying selection, whereby deleterious mutations are selectively eliminated to preserve protein function; when Ka/Ks = 1, neutral evolution predominates; and when Ka/Ks > 1, positive selection is indicated, whereby beneficial mutations are selectively retained [66–68]. In chloroplast genomes, most protein-coding genes typically exhibit strong purifying selection, as they encode essential components of photosynthesis, cellular respiration,

and other fundamental metabolic processes [69–71]. However, selection pressure varies considerably among gene categories, species, and environmental contexts, reflecting the dynamic balance between functional constraints and adaptive responses to environmental pressures. Recent comparative genomic studies have shown that even in highly conserved chloroplast genomes, specific gene sites may undergo positive selection, particularly those related to environmental adaptability [72, 73]. For example, in plants inhabiting extreme environments, specific sites of photosystem-related genes frequently exhibit positive selection signals, potentially reflecting adaptations to unique photosynthetic efficiency and stress tolerance mechanisms [74–77]. Simple sequence repeats (SSRs), as important molecular markers, show distribution patterns in chloroplast genomes that reflect evolutionary constraints and adaptive flexibility, thereby providing insights into genomic evolution and population genetics [78, 79]. Therefore, comprehensive studies combining selection pressure analysis, SSR distribution characteristics, and codon usage patterns will provide a comprehensive perspective for understanding the evolutionary dynamics of Poaceae chloroplast genomes.

This study provides a detailed investigation of codon usage patterns, selection pressures, and phylogenetic relationships in Poaceae chloroplast genomes using 48 species. These representative species were selected to cover all 12 recognized Poaceae subfamilies and include both C_3 and C_4 photosynthetic types. By integrating codon usage analysis (ENC), selection-pressure analysis (Ka/Ks), and SSR profiling, we systematically dissect the evolutionary forces shaping codon organization, protein-coding sequence divergence, and genome structural variation across Poaceae, providing a robust foundation for downstream functional genomics and breeding.

Materials and methods

Plant materials and data collection

Fresh leaves of *Eragrostis nigra* were collected from wild populations in the Muli area of the Qinghai-Tibet Plateau, China (101°16′48.74″E, 27°55′43.81″N, 2806 m) in July 2024 by Haodi Wang, and a voucher specimen (accession number: BJFU-CXY-36838841) was deposited at the School of Grassland Science, Beijing Forestry University. Field sampling was conducted under permission from the “Grassland Pest Census Project of Muli Tibetan Autonomous County Forestry and Grassland Bureau”. Genomic DNA was extracted using the modified CTAB method. High-throughput sequencing was conducted using the DNBSEQ-T7 platform, generating paired-end sequencing reads of 150 bp length. The raw reads were processed to obtain clean data for subsequent analysis. For comparative studies, chloroplast genome sequences of other closely related species were obtained from the National

Center for Biotechnology Information (NCBI) database (<https://www.ncbi.nlm.nih.gov/>). These additional datasets encompassed 47 Poaceae species with complete chloroplast genomes, covering all 12 subfamilies, and provided a comprehensive reference for phylogenetic and comparative genomic analysis.

Chloroplast genome assembly and annotation

The chloroplast genome of *E. nigra* was assembled using a dual-pipeline approach combining GetOrganelle v1.7.7 [80] and NOVOPlasty v4.3.4 [81] to ensure assembly accuracy and completeness. The annotation process employed the Chloë v0.1.0 database (Chloroplasts of Land Plants), a curated reference library optimized for photosynthetic organelles in land plant lineages, utilizing default parameters to maintain consistency with standard chloroplast genome annotation [82].

The annotation process involved initial automated gene predictions and functional assignments through GeSeq (<https://chlorobox.mpimp-golm.mpg.de/geseq.html>) [83], utilizing conserved chloroplast genes from related Poaceae species as reference sequences. Subsequent manual verification refined gene boundaries, corrected potential start/stop codon misannotations, and optimized exon-intron junctions, particularly for genes with complex splicing patterns, such as *ndhA* and *rpl2*. This combined approach of automated homology mapping and manual verification ensures accurate identification of protein-coding genes, rRNA operons, and tRNA genes, while distinguishing between functional sequences and pseudogenes.

Codon usage analysis

Relative Synonymous Codon Usage (RSCU) is a statistical metric that quantifies the frequency of synonymous codon usage in relation to the expected frequency, assuming equal usage of all synonymous codons. It is calculated as the ratio of the observed frequency of codon usage to the expected frequency [84]. An RSCU value of 1 indicates neutral usage, while values below 1 suggest lower frequency usage compared to other synonymous codons [85]. RSCU values were calculated using the formula in Python (v3.12.2) and visualized in R (v4.2.3).

$$RSCU_{ij} = \frac{x_{ij}}{\sum_j n_i} n_i \quad (1)$$

In the formula, $RSCU_{ij}$ represents the relative usage of the j -th synonymous codon encoding the i -th amino acid; x_{ij} represents the count of this codon in the target genome; n_i represents the total number of synonymous codons corresponding to the i -th amino acid; and $\sum_j x_{ij}$ represents the total observed count of all synonymous codons for this amino acid.

Simple sequence repeat analysis

Simple sequence repeats (SSRs) are widely used genetic markers consisting of 1 to 6 nucleotides repeated in tandem [86]. MISA-web was used to facilitate the analysis of SSRs in the chloroplast genomes of Poaceae species [87]. The MISA-web parameters were specified as follows: mononucleotides (10), dinucleotides (5), trinucleotides (4), tetranucleotides (3), pentanucleotides (3), hexanucleotides (3), heptanucleotides (3), octanucleotides (2), nonanucleotides (2), and decanucleotides (2). Data visualization was performed using R language for generating bar graphs and bubble plots.

ENC-GC_{3s} plot analysis

The Effective Number of Codons (ENC) measures the number of codons utilized in a gene, with values ranging from 20 to 61. A value of 20 indicates exclusive use of one codon per amino acid, while 61 represents uniform usage of all codons. ENC values below 35 indicate strong codon bias, whereas values greater than or equal to 35 suggest relatively weak codon bias [88, 89]. The ENC plot evaluates differences between actual and theoretical effective codon values, facilitating discussion of mutation and natural selection effects on codon usage patterns in the chloroplast genome [89]. The calculation formula is as follows:

$$ENC = 2 + \frac{9}{F_2} + \frac{1}{F_3} + \frac{5}{F_4} + \frac{3}{F_6} \quad (2)$$

In the formula, F_2 , F_3 , F_4 , and F_6 represent the average frequencies of families with 2, 3, 4, and 6 synonymous codons encoding the same amino acid, respectively. The ENC values were calculated and visualized using R. The ENC-GC_{3s} plot was constructed with GC_{3s} values on the horizontal axis and ENC values on the vertical axis. Data points of varying colors represent the actual ENC values across different species, while the dotted line indicates the expected ENC curve. When actual ENC values are distributed near the expected curve, this suggests that mutation pressure is the primary factor influencing codon usage bias formation. Conversely, if actual ENC values fall significantly below or deviate substantially from the expected curve, this indicates that natural selection or other factors predominantly determine codon usage bias [90].

Parity rule 2 (PR2) bias analysis

The PR2 plot elucidates whether mutation and selection pressure influence nucleotide composition. The plot's center coordinates (0.5, 0.5) denote A=T and G=C. A central point indicates that nucleotide composition is unaffected by mutation or natural selection [91]. Similar ratios of G and C (or A and T) suggest that mutation

exclusively affects gene codon usage preference. Conversely, substantial differences in these ratios indicate that natural selection or other factors predominantly influence gene codon usage preference [92]. In this study, we calculated $G_3/(G_3 + C_3)$ and $A_3/(A_3 + T_3)$ values for the study species. For each gene, the frequencies of G and C at the third codon position of synonymous codons (G_3 , C_3) and A and T (A_3 , T_3) were summed, respectively. The PR2 plot displays $G_3/(G_3 + C_3)$ on the horizontal axis and $A_3/(A_3 + T_3)$ on the vertical axis. Species are distinguished by different colors, with point sizes corresponding to ENC values.

Neutrality plot analysis

Neutrality plot analysis examines how mutational pressure and natural selection influence codon usage [93]. In this analysis, GC_{12} was calculated as the average of GC_1 and GC_2 . A scatter plot was generated with GC_3 on the horizontal axis and GC_{12} on the vertical axis, including the regression coefficient calculation. The data points were color-coded according to ENC values, and the plot included a red dashed standard curve and a blue dashed fitted regression line. A significant correlation between GC_{12} and GC_3 , combined with a regression line slope approaching or equaling 1, indicates that natural selection plays a dominant role in shaping codon usage [94].

Phylogenetic analysis

Phylogenetic relationships were inferred using a concatenated dataset containing all protein-coding genes (PCGs) from the complete chloroplast genomes. A total of 65 species were included in the phylogenetic analysis, comprising 48 Poaceae species representing all 12 recognized subfamilies (Anomochlooideae, Pharoideae, Puelioideae, Bambusoideae, Oryzoideae, Pooideae, Aristidoideae, Panicoideae, Chloridoideae, Danthonioideae, Arundinoideae, and Micrairoideae) and 17 outgroup taxa selected from closely related families within the order Poales. The outgroup taxa included 12 species from Bromeliaceae (*Brocchinia paniculata*, *B. micrantha*, *B. acuminata*, *Bakerantha lundelliorum*, *Navia igneosicola*, *Guzmania conifera*, *Tillandsia purpurea*, *T. usneoides*, *Pitcairnia atrovirens*, *Cryptanthus acaulis*, *Billbergia horrida*, and *Aechmea fasciata*) and five species from Typhaceae (*Sparganium angustifolium*, *S. subglobosum*, *Rohrbachia minima*, *Typha przewalskii*, and *T. orientalis*). We used MAFFT v7.490 [95] with the ‘—auto’ parameter to align the chloroplast genomes. Subsequently, we used trimAl v1.5 [96] with the ‘-automated1’ parameter to optimize the aligned sequences and automatically determine optimal trimming parameters. The trimming process systematically eliminated regions containing ambiguous alignments and gaps, thus

enhancing the phylogenetic signal while reducing dataset noise. The phylogenetic tree was constructed using the approximately maximum-likelihood method implemented in FastTree v2.1 with default parameters. FastTree employs the JTT (Jones-Taylor-Thornton) model for amino acid evolution and estimates the rate of evolution for each site using the ‘CAT’ approximation. Branch support values were estimated using the approximate likelihood-ratio test (aLRT) with Shimodaira-Hasegawa-like interpretation, which provides rapid and reliable assessment of branch reliability equivalent to 1,000 bootstrap pseudoreplicates. The resulting phylogenetic tree illustrated the evolutionary relationships among species. Tree visualization and annotation were performed using ggtree and the ape package in R (v4.2.3).

Selection pressure analysis

Calculation of nonsynonymous (K_a) and synonymous (K_s) substitution rates provides a crucial method for assessing sequence divergence of protein-coding orthologs across species or taxonomic lineages with uncertain evolutionary history. Orthologous gene pairs were identified based on gene name matching across the 48 Poaceae chloroplast genomes. For each orthologous gene pair, coding sequences (CDS) were extracted from GenBank-format annotation files and aligned using MAFFT v7.490 with the ‘—auto’ parameter for automatic strategy selection. Aligned sequences were converted to AXT format for K_a/K_s calculation. K_a/K_s ratios were computed using the maximum likelihood method in KaKs Calculator v2.0 (<https://sourceforge.net/projects/kakscalculator2/>) [97]. A total of 2,401 pairwise gene comparisons were performed across 11 functional gene categories. A K_a/K_s ratio exceeding 1 indicates a higher frequency of nonsynonymous substitutions compared to synonymous substitutions, typically suggesting positive selection or adaptive evolutionary pressure. Under these conditions, amino acid-altering mutations provide selective advantages, driving their fixation in populations [98]. A ratio of 1 indicates comparable rates of nonsynonymous and synonymous substitutions, suggesting neutral evolution where mutations occur randomly without significant natural selection influence. Genes in this evolutionary state typically serve essential biological functions requiring protein stability or functional consistency [72]. A ratio below 1 indicates higher synonymous substitution rates compared to nonsynonymous substitutions, suggesting purifying selection, or negative selection. Such genes likely maintain conserved functions, where nonsynonymous mutations could potentially disrupt protein structure or activity, leading to their elimination through natural selection [67].

Selection pressure categories were defined as: strong purifying selection ($Ka/Ks < 0.5$), relaxed purifying selection ($0.5 \leq Ka/Ks < 1.0$), and positive selection ($Ka/Ks > 1.0$). Genes were classified into functional categories based on standard chloroplast gene nomenclature: Photosystem I (*psaA-psaI*), Photosystem II (*psbA-psbZ*), cytochrome b/f complex (*petA-petN*), ATP synthase (*atpA-atpI*), NADH dehydrogenase (*ndhA-ndhK*), RuBisCO large subunit (*rbcL*), RNA polymerase (*rpoA-rpoC2*), large ribosomal proteins (*rpl2-rpl36*), small ribosomal proteins (*rps2-rps19*), hypothetical reading frames (*ycf* genes), and other protein-coding genes (*matK*, *cemA*, *ccsA*, *infA*, *clpP*). Visualization of Ka/Ks analysis results was performed using R v4.2.3 with ggplot2, viridis, and patchwork packages.

Genome synteny analysis

Syntenic analysis of chloroplast genomes examines gene order, orientation, and relative positions across different species. This comparative approach helps elucidate interspecific relationships and infer evolutionary histories. Through comparative genome structure analysis and lineage mapping, synteny analysis provides fundamental data for phylogenetic studies, enabling the identification of conserved gene blocks and chromosomal rearrangements. Phylogenetic trees derived from synteny data can integrate temporal and functional information, thereby validating evolutionary hypotheses and revealing patterns of genome evolution [99]. Synteny analysis and visualization were performed using genoPlotR package (v0.8.11) in R (v4.2.3), which enables comparative visualization of multiple genomes in a linear format. Gene features were classified into 12 functional categories based on standard chloroplast nomenclature: Photosystem I (*psa* genes), Photosystem II (*psb* genes), cytochrome b6/f complex (*pet* genes), ATP synthase (*atp* genes), NADH dehydrogenase (*ndh* genes), RuBisCO large subunit (*rbcL*), large ribosomal proteins (*rpl* genes), RNA polymerase (*rpo* genes), small ribosomal proteins (*rps* genes), ribosomal RNA (*rrn* genes), transfer RNA (*trn* genes), and other genes (*accD*, *ccsA*, *cemA*, *clpP*, *matK*, *ycf*). A total of 47 pairwise comparisons were computed across the 48 chloroplast genomes to comprehensively assess structural conservation patterns.

Each horizontal track represents the complete chloroplast genome of a single species, with individual genes depicted along the genome. Genes were represented as directional arrows indicating transcriptional orientation, with distinct colors denoting different functional categories. Homologous regions between adjacent genomes were connected by colored ribbons, with color intensity reflecting sequence identity levels. A phylogenetic tree was integrated alongside the

synteny plot to illustrate evolutionary relationships among the compared species.

Results

Genome structure and nucleotide composition

The chloroplast genomes of the 48 selected Poaceae species exhibited relatively consistent base composition throughout the coding region (Table 1), with total genome lengths ranging from 129,095 bp to 148,617 bp. The total GC content varied from 38.16% to 39.04%, averaging 38.55%, substantially below 50%. The circular chloroplast genome map of *E. nigra* (Fig. 1A) was generated using CPGview (<http://www.1kmpg.cn/cpgview/>) [100]. Genes containing both introns and exons include *rps16*, *atpF*, *pafl*, *petB*, *petD*, *rpl16*, *rpl2*, *ndhB* and *ndhA* (Fig. 1B). Analysis of the inverted repeat (IR) boundaries (Fig. 1C) showed that the *rps19* gene was positioned in the IR region, near the LSC-IR boundary, across three species, while the *trnH* gene consistently localized to the center of the IR region. The *rpl22* gene was located in the LSC region, adjacent to the LSC-IR boundary. The IR-SSC boundary showed greater complexity: some species, including the studied *E. nigra*, contained both *ndhF* and *ndhH* genes at this boundary, while others possessed only *ndhH*.

The GC content of all 48 species was below 50%, indicating an A/T preference in their genomes. Analysis of GC content at different codon positions showed average contents of GC₁, GC₂, and GC₃ of 47.34%, 39.64%, and 30.51%, respectively, with GC₃ showing the lowest average content, demonstrating the strongest codon usage bias at the third position. The coding region length varied from 55,713 bp to 63,825 bp, averaging 59,796 bp. The average GC content of coding regions was 39.16%, markedly higher than that of non-coding regions at 35.11%, indicating greater GC stability in coding regions. Additionally, the average ENC values of these 48 Poaceae plants ranged from 48.028 to 50.158, all exceeding 35.

RSCU patterns

To characterize codon preferences and evolutionary constraints across Poaceae, we calculated RSCU values for 59 sense codons in 48 chloroplast genomes representing all 12 subfamilies spanning major ecological and phylogenetic diversity. Our analysis encompassed both cool-season (C₃ photosynthesis) and warm-season (C₄ photosynthesis) species.

Hierarchical clustering analysis of RSCU values revealed a strong phylogenetic signal in codon usage patterns (Fig. 2). The heatmap demonstrated distinct codon usage profiles for Pooideae, cool-season grasses, Chloridoideae, and Panicoideae, predominantly warm-season grasses, consistent with their evolutionary divergence approximately 50–60 million years ago. Ward's minimum

Table 1 Chloroplast genome information of 48 Poaceae forage and ecological species

Species name	Accession number	Subfamily	Total length	Total GC (%)	Average ENC	GC3s	Total gene
<i>Agropyron cristatum</i> [101]	KY126307	Pooideae	135554	38.33	48.291	26.82	131
<i>Agrostis stolonifera</i>	EF115543	Pooideae	136584	38.45	48.657	27.22	141
<i>Aristida adscensionis</i> [102]	MZ373986	Aristidoideae	137947	38.41	48.761	27.1	133
<i>Aristida congesta</i> [103]	MK570615	Aristidoideae	134072	38.33	48.832	27.04	131
<i>Arundinaria tecta</i> [104]	KC817463	Bambusoideae	139499	38.92	49.496	28.15	131
<i>Arundo donax</i>	KX109945	Arundinoideae	137093	38.63	48.484	28.52	143
<i>Avena fatua</i>	MG687307	Pooideae	135889	38.51	48.584	27.2	134
<i>Bothriochloa ischaemum</i>	MW699899	Panicoideae	138316	38.53	49.148	27.46	134
<i>Cenchrus purpureus</i>	MF594682	Panicoideae	138199	38.59	49.915	27.87	137
<i>Chikusichloa aquatica</i>	KR078265	Bambusoideae	136563	39.04	49.769	28.25	129
<i>Dactylis glomerata</i>	OM049559	Pooideae	134972	38.43	48.316	26.97	129
<i>Elymus sibiricus</i> [105]	MT610376	Pooideae	135075	38.33	48.462	26.98	122
<i>Eragrostis atrovirens</i> [106]	MW255512	Chloridoideae	134857	38.16	48.302	27.05	133
<i>Eragrostis cylindrica</i> [106]	MW255516	Chloridoideae	133777	38.25	48.492	27.04	133
<i>Eragrostis ferruginea</i> [106]	MW255517	Chloridoideae	134380	38.23	48.759	26.96	131
<i>Eragrostis nigra</i>	PV455321	Chloridoideae	134870	38.22	48.811	26.89	134
<i>Eragrostis pilosa</i> [107]	OL678516	Chloridoideae	134764	38.25	48.867	27.01	132
<i>Eriachne avenacea</i> [108]	MH036238	Micrairoideae	134857	38.62	49.395	27.75	129
<i>Eriachne pallescens</i> [108]	MH036267	Micrairoideae	134949	38.57	48.987	27.61	129
<i>Fargesia edulis</i> [109]	MH988735	Bambusoideae	139572	38.89	49.535	28.19	131
<i>Fargesia grossa</i> [109]	MH988734	Bambusoideae	139601	38.89	49.53	28.17	131
<i>Fargesia spathacea</i> [110]	JX513417	Bambusoideae	139767	38.88	49.567	28.04	130
<i>Festuca ovina</i> [111]	MN309824	Pooideae	133508	38.36	48.045	27.06	127
<i>Himalayacalamus gyirongensis</i> [109]	MH988733	Bambusoideae	139592	38.85	49.548	28.15	131
<i>Hygroryza aristata</i>	MW849262	Oryzoideae	135681	39.03	49.246	28.06	131
<i>Indocalamus wilsonii</i> [110]	JX513421	Bambusoideae	139962	38.8	49.428	28	130
<i>Leptaspis zeylanica</i>	KU666546	Pharoideae	141946	38.48	49.909	27.75	115
<i>Lolium arundinaceum</i> [112]	FJ466687	Pooideae	136048	38.39	48.903	27.41	129
<i>Lolium perenne</i> [113]	AM777385	Pooideae	135282	38.25	48.577	26.85	129
<i>Megathyrsus maximus</i>	MT679705	Panicoideae	138186	38.59	49.584	27.56	133
<i>Panicum capillare</i> [4]	KU291475	Panicoideae	134520	38.51	49.676	27.57	131
<i>Panicum incomtum</i> [114]	MH394349	Panicoideae	139999	38.68	50.158	27.91	129
<i>Panicum lycopodioides</i> [115]	MF563374	Panicoideae	138954	38.61	49.993	27.57	130
<i>Panicum pygmaeum</i> [115]	MF563373	Panicoideae	140996	38.64	49.746	27.66	130
<i>Panicum sumatrense</i>	KX756177	Panicoideae	139384	38.62	49.599	27.68	132
<i>Paspalum dilatatum</i> [4]	KU291477	Panicoideae	135950	38.67	49.531	27.83	131
<i>Pharus latifolius</i>	JN032131	Pharoideae	142077	38.37	49.528	27.75	117
<i>Phragmites australis</i>	KF730315	Arundinoideae	137561	38.65	49.175	28.17	133
<i>Poa pratensis</i>	MT295102	Pooideae	135649	38.3	49.241	27.29	131
<i>Puccinellia distans</i>	MW044608	Pooideae	135647	38.32	48.519	27.09	129
<i>Puelia olyriiformis</i> [116]	KC534841	Puelioidae	140344	38.97	49.747	28.27	134
<i>Rytidosperma pallidum</i> [36]	MF460980	Danthonioideae	132765	38.29	48.164	27	131
<i>Rytidosperma semiannulare</i>	KY432793	Danthonioideae	129095	38.31	48.028	27.21	129
<i>Stipagrostis pennata</i> [102]	MZ373985	Aristidoideae	138725	38.42	49.681	27.23	129
<i>Streptochaeta spicata</i> [117]	KU666544	Anomochlooideae	148617	38.25	49.252	27.65	129
<i>Thamnocalamus spathiflorus</i> [110]	JX513425	Bambusoideae	139778	38.84	49.446	27.95	130
<i>Yushania levigata</i> [110]	JX513426	Bambusoideae	139633	38.9	49.583	28.07	130
<i>Zizania latifolia</i> [118]	KT161956	Oryzoideae	136461	39	49.268	28.05	129

The table summarizes genome characteristics for each species, including total genome length (bp), subfamily, total GC content (%), average ENC value, GC content at the 3rd Synonymous codon position, and total number of genes

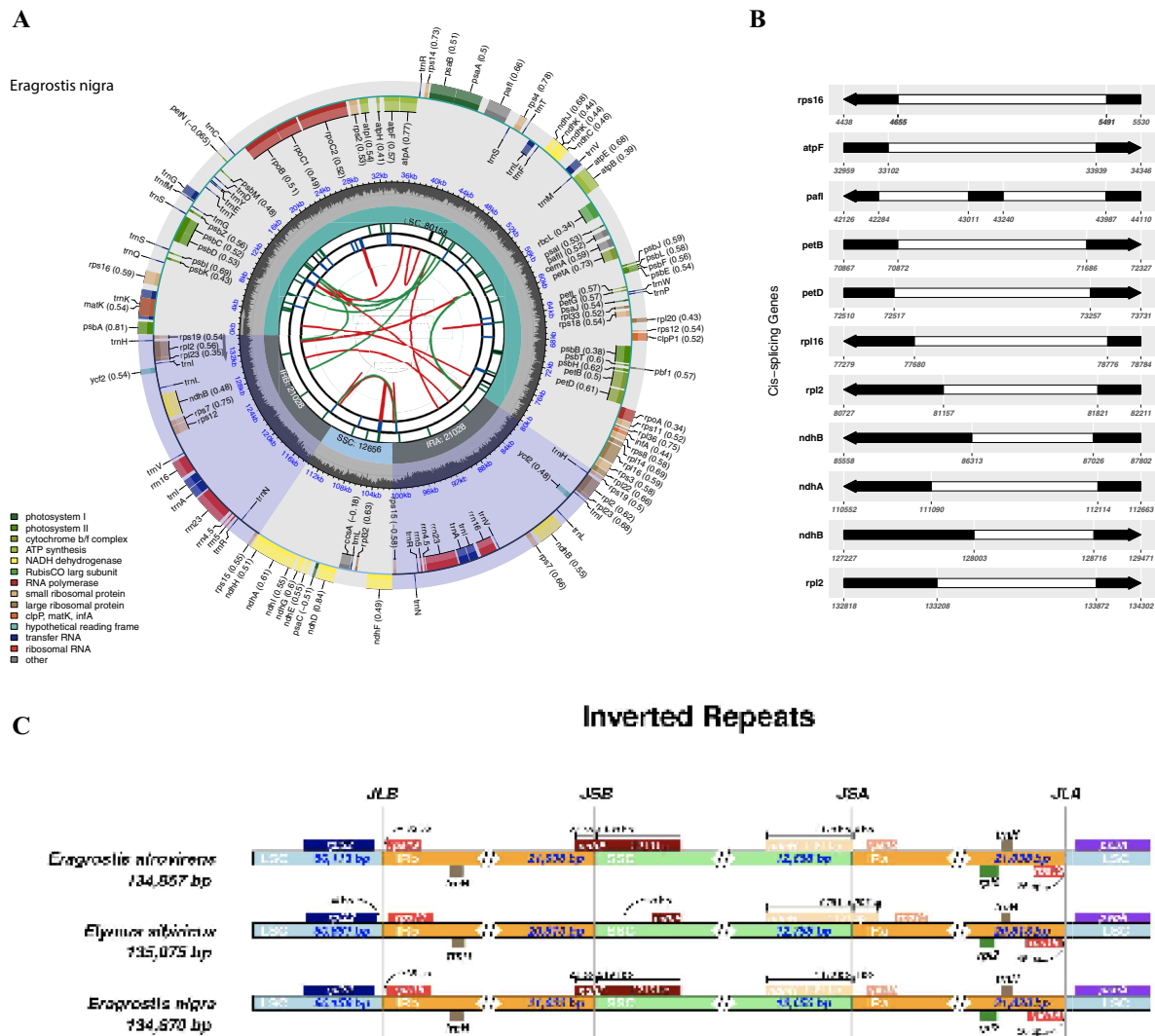


Fig. 1 Chloroplast genome structure and organization in Poaceae species. **A** Circular map of the complete chloroplast genome of *Eragrostis nigra*. **B** Detailed structures of selected chloroplast genes showing exon (black boxes) and intron (white boxes) organization. Gene names are shown on the left, with nucleotide positions marked below each gene structure. **C** Comparative analysis of IR boundary regions among three species (*E. atrovirens*, *E. nigra*, and *Elymus sibiricus*). Colored boxes represent genes at or near the IR boundaries. The four junctions between the two single-copy regions and the inverted repeats are labeled as JLB (LSC/IRb), JSB (SSC/IRb), JSA (SSC/IRa), and JLA (LSC/IRa). Total genome sizes for each species are shown on the left

variance clustering successfully grouped species according to subfamily affiliation, with bootstrap support exceeding 85% for major clades. Notable exceptions included several Bambusoideae species that clustered with Pooideae, reflecting their shared cool-season adaptation despite phylogenetic distance. Analysis of codon usage bias intensity using ENC revealed significant differences between temperature adaptation groups. Examination of RSCU frequency distribution revealed significant deviation from neutral expectations. The distribution was right-skewed with a mode at RSCU = 0.85, indicating that most codons are used less frequently than expected under unbiased conditions. Approximately 23% of codons showed under-representation (RSCU < 1.0), 45% fell within moderate usage (1.0 ≤ RSCU < 1.5), and 32%

displayed preferential usage (RSCU ≥ 1.5). A pronounced right-tail extended beyond RSCU = 3.0, reflecting a subset of strongly favored codons used at more than triple neutral expectation. Across all 48 species, universally preferred codons were strongly biased toward A/U-ending variants. The 10 most frequently used codons, namely UUA, AGA, GCU, ACU, UCU, UAU, GGA, CCU, GAU, and CAA, were predominantly A/U-terminated, aligning with the plastome-wide AT-richness. Conversely, the 10 least preferred codons, namely GCG, CAG, CGG, GAC, AGC, CCG, UAC, CUC, GGC, and CUG, were G/C-ending forms, suggesting both mutational pressure and selective constraints in tRNA recognition contribute to suppressing GC-rich codons. The prevalence of A/U-biased optimal codons, paired with under-representation

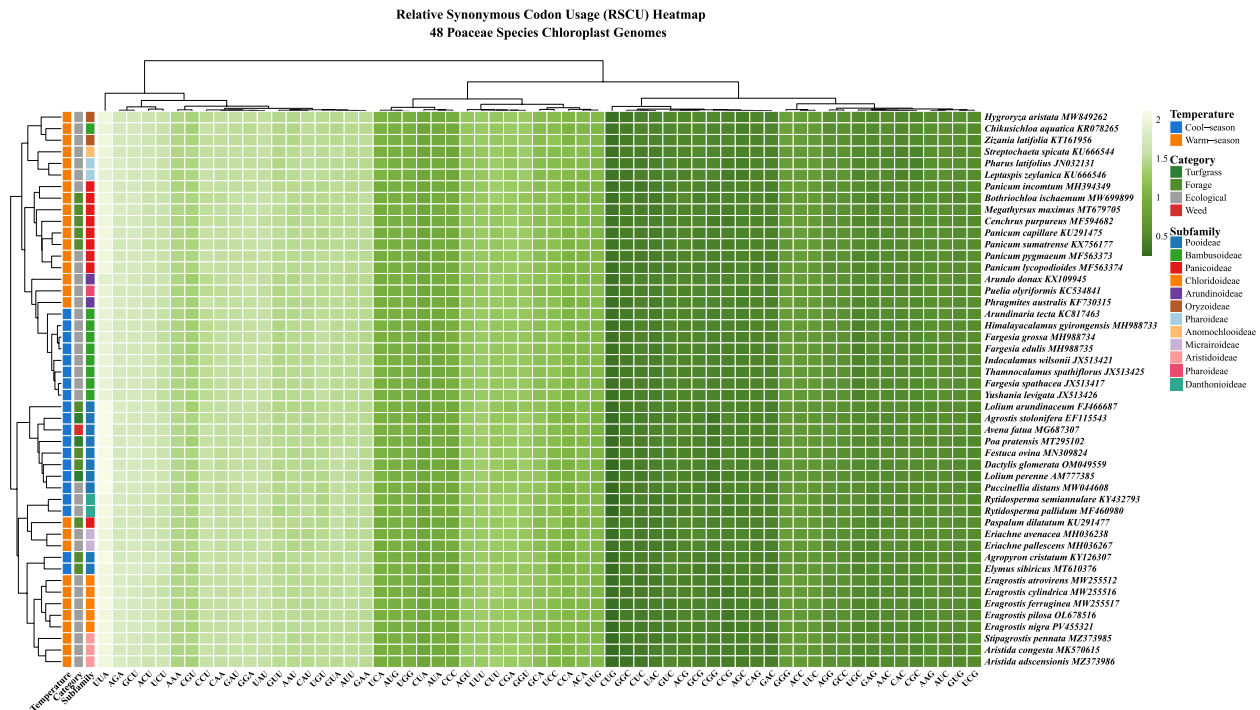


Fig. 2 Analysis of codon usage bias in 48 Poaceae chloroplast genomes reveals diverse subfamily-specific evolutionary patterns and temperature-dependent optimization strategies across multiple ecological adaptations

of G/C-ending variants, supports the role of translational selection acting in concert with nucleotide compositional bias. These patterns indicate that chloroplast-encoded proteins likely experience preferential translation efficiency when using A/U-rich codons, possibly reflecting limited tRNA diversity and resource allocation in plastids.

Distribution patterns of SSRs

SSR profiling across the chloroplast genomes of 48 Poaceae species revealed clear patterns in both genomic location and motif composition. SSRs were most frequently located in intergenic regions (IGS) (52.56%), followed by intronic regions (28.23%), with the lowest occurrence in coding sequence regions (CDS) (19.21%). The analysis demonstrated that SSRs across the 48 samples exhibited the highest distribution in intergenic regions (Fig. 3A). Additionally, 11 SSR types were present in nearly all species with 6 types appearing in all species. Mononucleotide and octanucleotide repeats were the most abundant types of simple sequence repeats (SSRs) in the complete chloroplast genomes across all studied species. Specifically, mononucleotide repeats totaled 1438, accounting for 40.97% of the total SSRs, while octanucleotide repeats reached 792, representing 22.56% of the total (Fig. 3B). Furthermore, the number of SSRs in the complete chloroplast genomes exhibited a notable range across the studied species, varying from a minimum of 45 in *Hydroryza*

aristata to a maximum of 99 in *E. nigra*, indicating inter-specific variation in SSR abundance (Fig. 3C).

ENC-GC_{3s} plot analysis

To further assess codon usage bias in Poaceae chloroplast genomes, ENC values were calculated and visualized using an ENC-GC_{3s} plot. The mean ENC value was 49.14, indicating an overall weak CUB across the 48 species. ENC and GC_{3s} distributions showed similar overall trends, with data points forming concentrated clusters rather than broad dispersion, suggesting comparable evolutionary constraints among lineages. The majority of these points were distributed outside or above the expected ENC curve, while few points appeared on the curve (Fig. 4). This deviation from the theoretical neutrality line indicates that codon usage patterns are not solely determined by mutational pressure. Instead, the observed upward shift of most points implies that natural selection plays the dominant role in shaping chloroplast codon preference in Poaceae, while mutation pressure exerts a comparatively minor influence.

Parity rule 2 bias analysis

To evaluate the relative influence of mutation and natural selection on codon usage patterns, a PR2 plot was generated to compare the proportions of A vs T and G vs C at the third codon position (Fig. 5). The vertical distribution showed that most data points clustered in the lower half of the plot, near the central dashed line, indicating a

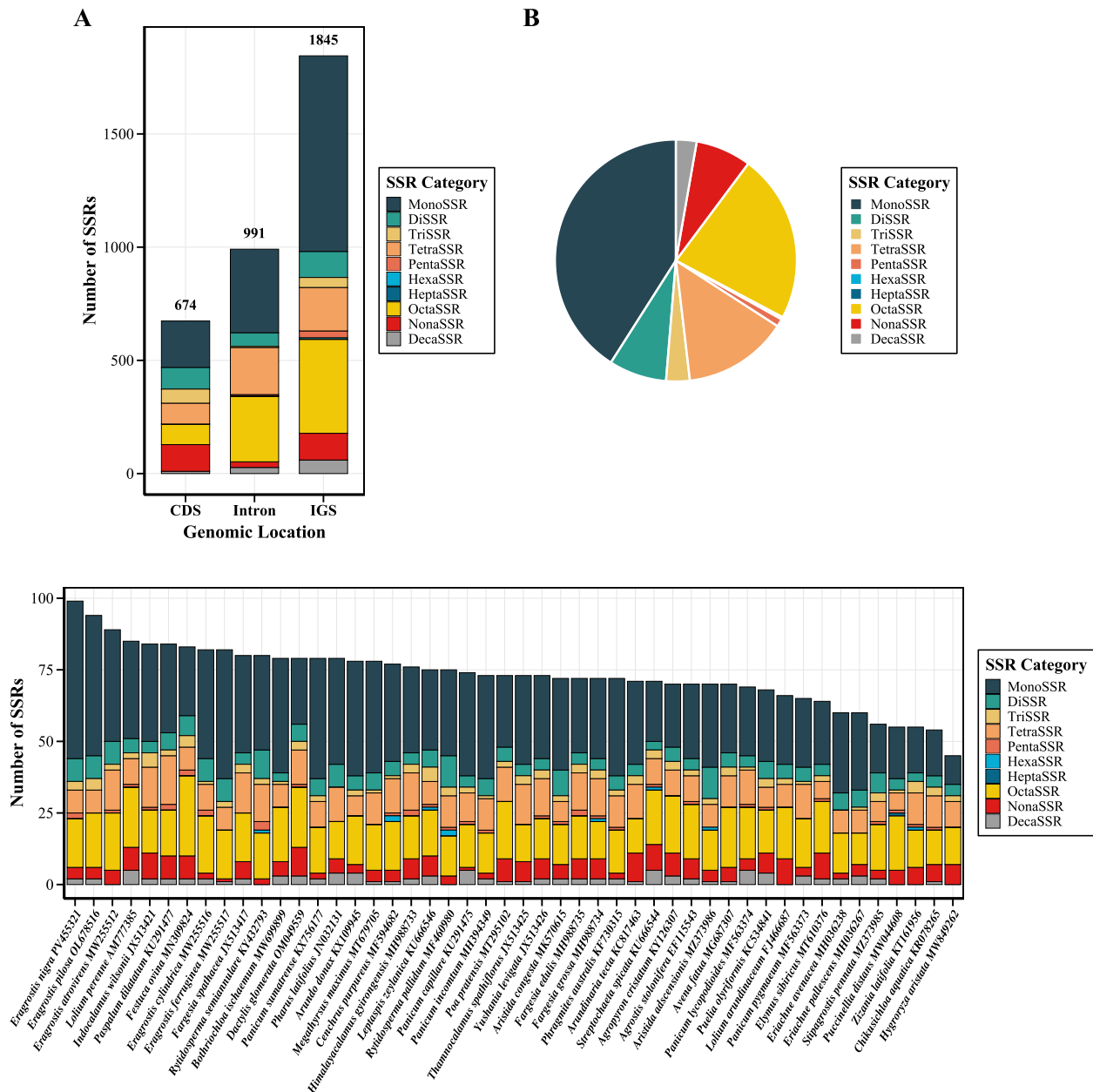


Fig. 3 Simple sequence repeat (SSR) analysis in Poaceae chloroplast genomes. **A** Distribution of SSR categories across different genomic locations in chloroplast genomes. Numbers represent the count of each SSR category found in different genomic regions. **B** Pie chart displaying the percentage distribution of different SSR categories across all analyzed species. Values show the relative abundance of each SSR type. **C** Stacked bar chart showing the distribution and abundance of SSR categories across different pasture and ecological Poaceae species

higher preference for T over A at the third codon position in these 48 forage and ecological species. The horizontal distribution, though more dispersed, showed most data points concentrated in the right half of the plot. Analysis revealed the highest concentration of data points in the fourth quadrant, confirming the preference for T and G at the third codon position. *E. nigra* displayed a broader distribution, with average G_3 proportions of 0.538 and A_3 proportions of 0.471, showing a more concentrated pattern compared to other species. In terms of gene function

classification, Photosystem II-related genes were more concentrated in the third quadrant, indicating a preference for using C and T codons, while other genes showed a bias toward G. Meanwhile, with the exception of genes related to small ribosomal protein and large ribosomal protein, all other genes exhibited a preference for T. Collectively, the scattered distribution of points indicates that natural selection pressure, rather than mutation pressure, is the primary factor affecting third codon position preference.

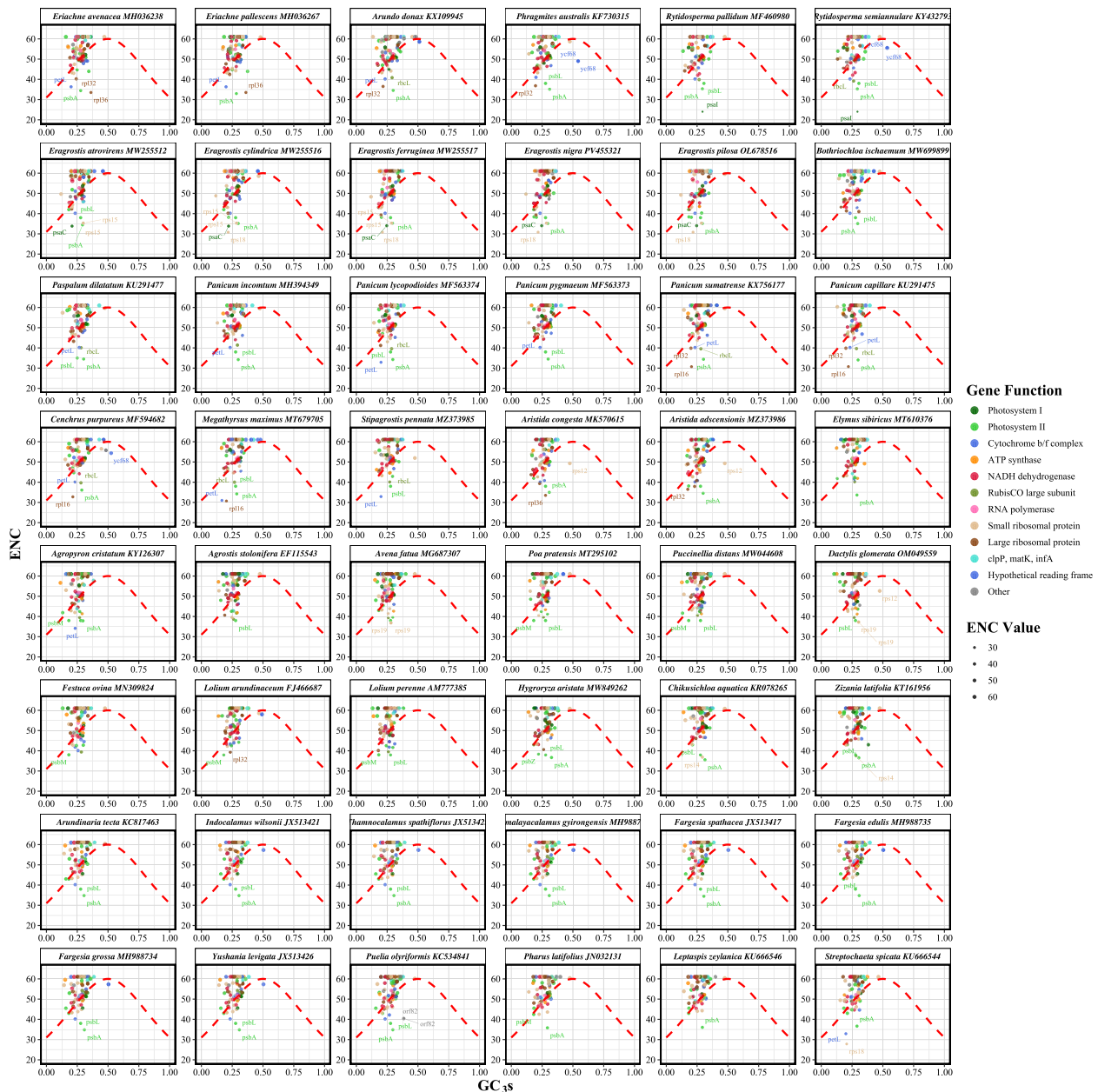


Fig. 4 ENC- GC_{3s} plot analysis of Poaceae chloroplast genomes. Each panel represents a different species with the ENC plotted against the GC content at GC_{3s} . The red dashed curves represent the expected relationship between ENC and GC_{3s} under mutational pressure alone. Data points above the curve indicate the influence of additional factors beyond mutation, particularly natural selection. Points are colored according to gene functional categories. Point size corresponds to Effective Number of Codons (ENC) values. Dashed blue lines indicate the neutral expectation (0.5) where equal frequencies of complementary bases would be expected under neutral evolution. Deviations from the center (0.5, 0.5) indicate the influence of natural selection and/or mutational pressure on codon usage bias in grass chloroplast genomes

Neutrality plot analysis

GC_{12} and GC_3 analysis examined the effects of natural selection and mutation on codon usage bias (Fig. 6). The results indicated no significant correlation between GC_{12} and GC_3 in the chloroplast genomes of the 48 samples. Data points clustered predominantly in one group, with GC_{12} values ranging from 0.3 to 0.6 and GC_3 values from 0.125 to 0.55. This distribution pattern indicates

that CUB is primarily influenced by natural selection. The absolute regression coefficients for GC_{12} versus GC_3 ranged from 0.0037 (*P. pratensis*) to 0.2227 (*Arundo donax*), suggesting mutational pressure contributed between 0.37%–22.27% to CUB formation. Species with regression coefficients exceeding 0.1 included *Rytidosperma semiannulare*, *L. arundinaceum*, *Himalayacalamus gyirongensis*, *Fargesia grossa*, *F. edulis*, *Megathyrsus*

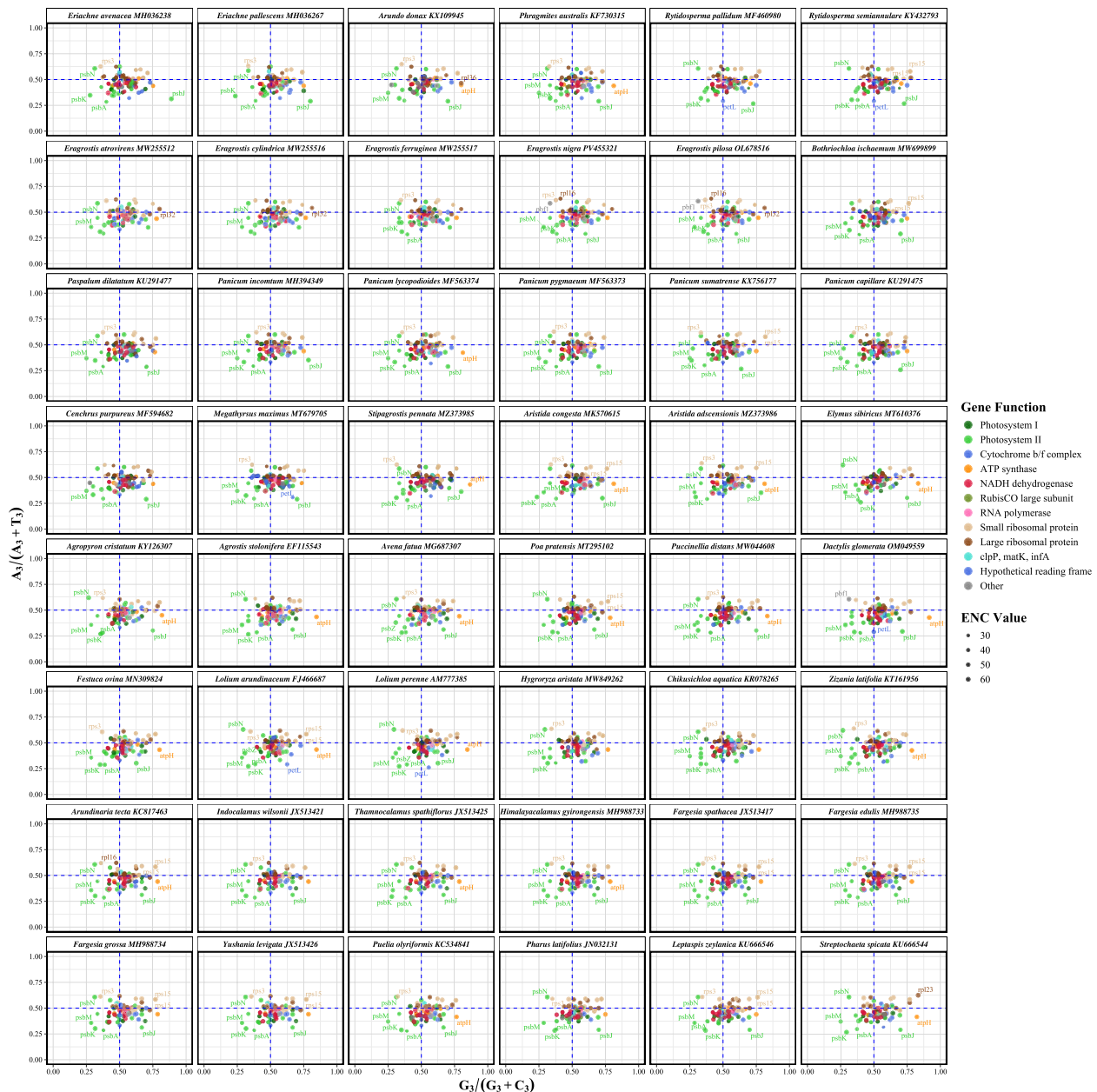


Fig. 5 PR2 plot analysis of codon usage patterns in 48 Poaceae chloroplast genomes. Points are colored according to gene functional categories. The plots display $A_3/(A_3+T_3)$ values on the y-axis against $G_3/(G_3+C_3)$ values on the x-axis, with data points representing individual genes. Blue dashed lines indicate the equilibrium positions (0.5) where no bias exists

maximus, *C. purpureus*, and *A. donax*. The remaining 40 species showed regression coefficients below 0.1, indicating that natural selection and other factors influenced their CUB by more than 90%. These findings emphasize the significant role of natural selection in shaping CUB.

Phylogenetic relationships and evolutionary patterns

To elucidate evolutionary relationships among 65 species (48 Poaceae ingroup taxa representing all 12 subfamilies and 17 outgroup taxa from Bromeliaceae and Typhaceae), we constructed a comprehensive phylogenetic tree based

on 4,887 chloroplast protein-coding sequences. A phylogenetic tree was constructed using the approximately maximum-likelihood method with the WAG+G substitution model (Fig. 7). The phylogenetic analysis successfully resolved relationships across all major Poaceae subfamilies. The phylogeny resolved the outgroups into two major clades, with Bromeliaceae (12 species) and Typhaceae (5 species) forming distinct sister lineages to the Poaceae, consistent with their phylogenetic position as early-diverging lineages within Poales. Within Poaceae, the basal subfamilies Anomochloideae (represented



Fig. 6 Neutrality plot analysis of the chloroplast genomes of 48 Poaceae species. Point size corresponds to Effective Number of Codons (ENC) values. Dashed blue lines indicate the neutral expectation (0.5) where equal frequencies of complementary bases would be expected under neutral evolution. Deviations from the center (0.5, 0.5) indicate the influence of natural selection and/or mutational pressure on codon usage bias in grass chloroplast genomes

by *Streptochaeta spicata*), Pharoideae (*Pharus latifolius*, *Leptaspis zeylanica*), and Puelioideae (*Puelia olyrififormis*) were recovered as successive sister lineages to the core grass clade (BOP + PACMAD), consistent with current grass phylogeny. All outgroup branches showed clear divergence from *Eragrostis*, supporting its placement well within the Poaceae lineage. Within *Eragrostis*, the analysis uncovered a strongly supported clade in which *E. nigra* and *E. pilosa* were most closely related. These two species further clustered with additional *Eragrostis* taxa to form a monophyletic group, which in turn was recovered as sister to the genus *Rytidosperma*, highlighting

their close evolutionary affinity. The phylogenetic tree exhibited high overall support values, with the majority of internal nodes receiving aLRT support values exceeding 0.75, indicating robust resolution of evolutionary relationships across the Poaceae family. The 12 subfamilies were clearly delineated in the tree topology, with Pooideae, Bambusoideae, Oryzoideae forming the BOP clade, and Panicoideae, Arundinoideae, Chloridoideae, Micraioideae, Aristidoideae, and Danthonioideae forming the PACMAD clade.

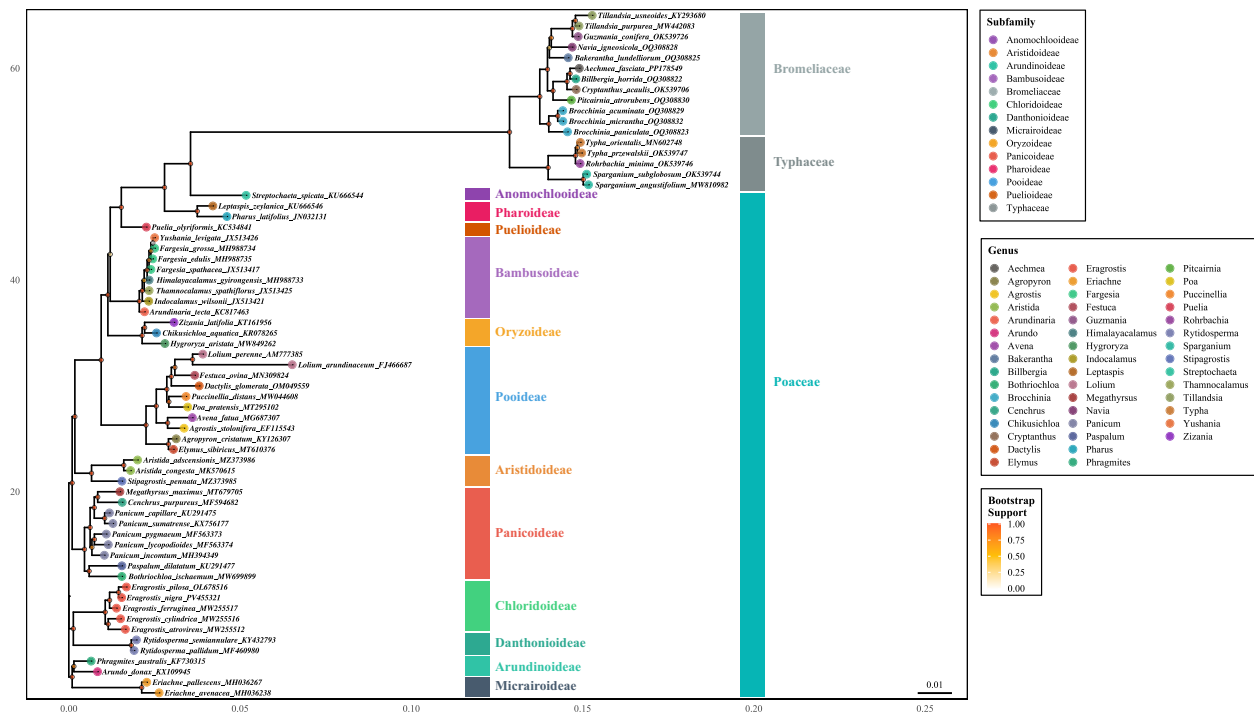


Fig. 7 Phylogenetic analysis of Poaceae and related Poales species based on chloroplast protein sequences. Approximately maximum-likelihood phylogenetic tree of 65 species, including 48 Poaceae species representing all 12 subfamilies (Anomochloideae, Pharoideae, Puelioideae, Bambusoideae, Oryzoideae, Pooideae, Aristidoideae, Panicoideae, Chloridoideae, Danthioideae, Arundinoideae, and Micraioideae) and 17 outgroup taxa from Bromeliaceae and Typhaceae. The tree was constructed using FastTree v2.1.11 with the WAG+G substitution model based on 4,887 concatenated chloroplast protein-coding sequences. Approximate likelihood-ratio test (aLRT) support values with Shimodaira-Hasegawa-like interpretation are indicated by the color gradient shown in the legend (right, top), with values ranging from 0 to 1. Different genera ($n = 46$) are distinguished by colored dots according to the legend (right, bottom). Subfamily affiliations are indicated by colored strip labels adjacent to the tree. The scale bar at the bottom indicates the number of amino acid substitutions per site. Outgroup families (Bromeliaceae and Typhaceae) are shown at the top of the tree, representing early-diverging lineages within Poales

Evolutionary selection pressure on protein-coding genes

Selective pressure analysis across the 48 Poaceae forage and ecological grass species revealed that chloroplast protein-coding genes are predominantly shaped by purifying selection consistent with the essential and conserved nature of organellar genes functions (Fig. 8). A total of 2,401 pairwise comparisons were successfully computed across 11 functional gene categories, representing comprehensive coverage of chloroplast protein-coding sequence evolution. Ka/Ks ratios were heavily skewed toward values below 0.5, with a median of 0.12, indicating that deleterious mutations are efficiently removed and functional stability is maintained across evolutionary time.

A strong positive correlation was observed between Ka and Ks, with most genes positioned below the neutrality threshold ($K_a = K_s$) (Fig. 8A), reinforcing purifying selection as the dominant evolutionary constraint. Genes were colored by functional category, with photosystem genes (PSA, PSB) clustering in the lower-left region indicating strong purifying selection, while genes such as *atpF* and *rps15* exhibited higher Ka/Ks ratios approaching or exceeding the neutrality

line. Among the 20 major gene types analyzed, significant variation in selection pressure was observed (Fig. 8B). Both large and small subunit ribosomal protein genes generally showed strong purifying selection, with median Ka/Ks ratios ranging from 0.0623 to 0.6603, reflecting stringent conservation due to their critical involvement in chloroplast translation. Likewise, RNA polymerase genes, including *rpoA*, *rpoB*, *rpoC1*, and *rpoC2*, exhibited Ka/Ks ratios consistently below 0.5, indicating purifying selection pressure typical of essential transcriptional machinery. Distinct selection signatures were detected across additional gene groups (Fig. 8C). Photosystem I genes exhibited the lowest mean Ka/Ks ratio among all functional categories, followed by cytochrome b/f complex genes and ATP synthase genes. Photosystem II genes showed a mean Ka/Ks of 0.2108, with core genes *psbA* and *psbD* displaying the strongest purifying selection among all analyzed genes, reflecting their essential role in the photosynthetic reaction center. NADH dehydrogenase genes exhibited consistent purifying selection, with 99.26% of comparisons showing Ka/Ks < 0.5. Analysis of selection category distribution revealed that

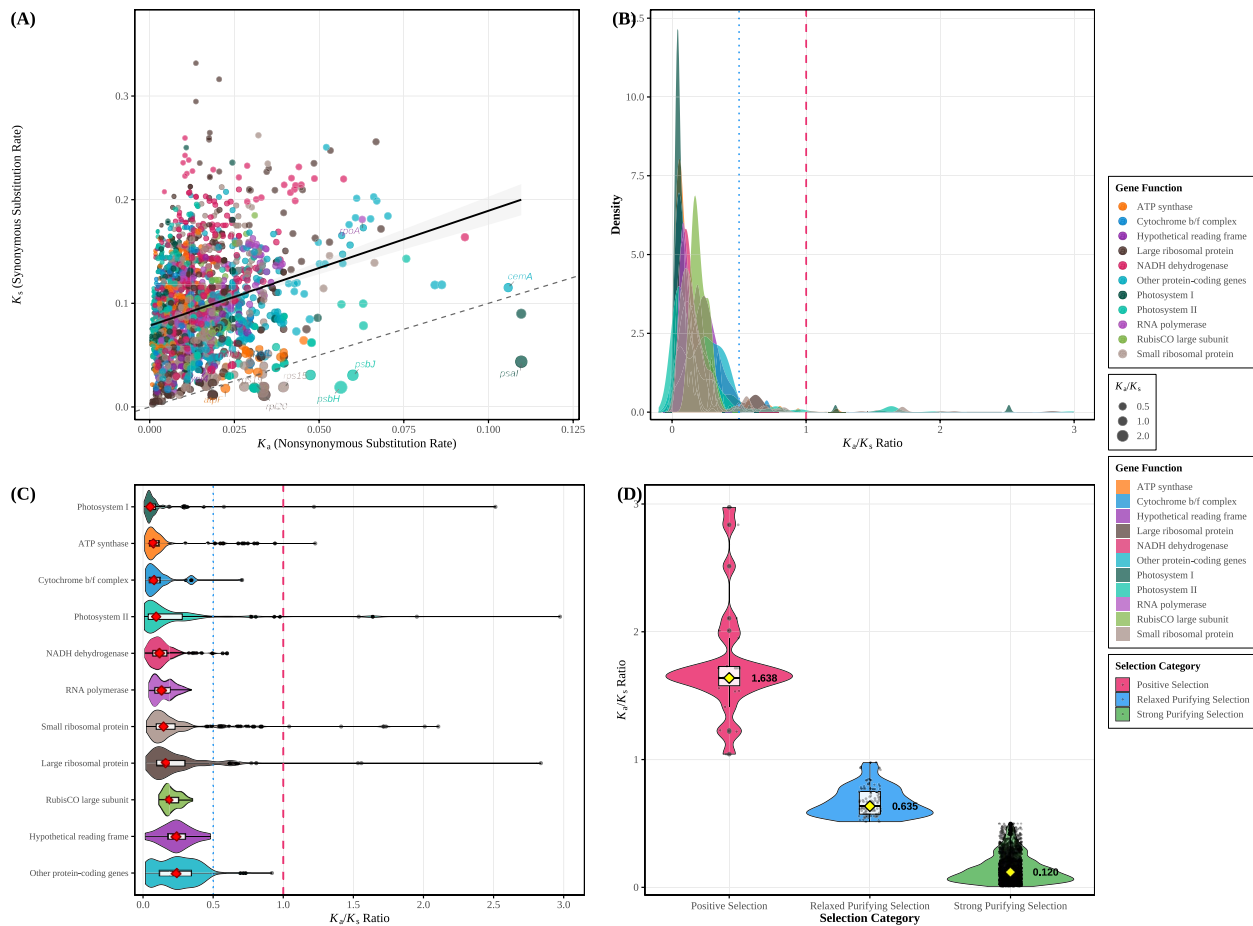


Fig. 8 Evolutionary selection analysis of chloroplast genes in Poaceae species. **A** Scatter plot illustrating the relationship between K_a and K_s substitution rates across chloroplast genes. Points are colored by gene family and sized according to K_a/K_s ratio. Key genes are labeled. **B** Density distribution of K_a/K_s ratios across different gene families. **C** Violin plot showing the distribution of K_a/K_s ratios for the 20 most important chloroplast gene types, ordered by median K_a/K_s values. **D** Violin plot with density curve showing the overall distribution of K_a/K_s ratios, with background shading and text annotations indicating selection categories (strong purifying: $K_a/K_s < 0.5$; relaxed purifying: $0.5 \leq K_a/K_s < 1$; positive selection: $K_a/K_s \geq 1$). Color gradient represents increasing K_a/K_s values

2,272 comparisons (94.67%) exhibited strong purifying selection ($K_a/K_s < 0.5$), 102 comparisons (4.25%) showed relaxed purifying selection ($0.5 \leq K_a/K_s < 1.0$), and only 26 comparisons (1.08%) displayed evidence of positive selection ($K_a/K_s \geq 1.0$) (Fig. 8D).

Among individual genes, *atpF* showed the highest mean K_a/K_s ratio (0.6928) within the ATP synthase category, with one comparison exhibiting positive selection ($K_a/K_s = 1.2281$). The ribosomal protein genes *rps15* (mean = 0.6045, 15.56% positive selection) and *rps12* (mean = 0.6603) displayed elevated K_a/K_s ratios, suggesting relaxed functional constraints. The photosystem II gene *psbH* exhibited the highest variance in selection pressure, with 19.57% of comparisons showing evidence of positive selection. The hypothetical reading frame gene *ycf4* showed consistent moderate purifying selection across all comparisons. RuBisCO large subunit (*rbcL*) exhibited uniform purifying selection, consistent with its essential role in carbon fixation.

Genome synteny and structural conservation

To further resolve evolutionary affinities, we performed chloroplast genome synteny analysis across 48 Poaceae forage and ecological grasses (Fig. 9). The results revealed a high degree of structural conservation, particularly among closely related taxa. Gene order remained extensively preserved across species, with core plastid components, including photosynthetic system genes, ribosomal protein genes, tRNAs, and rRNAs, exhibiting stable and consistent genomic organization throughout the family. Quantitative assessment of the 47 pairwise genome comparisons revealed that syntenic conservation was strongly correlated with phylogenetic distance. Within-subfamily comparisons exhibited consistently high sequence identity, while cross-subfamily comparisons showed greater variability. The BOP clade (Bambusoideae, Oryzoideae, Pooideae) demonstrated remarkably conserved synteny, with all pairwise comparisons

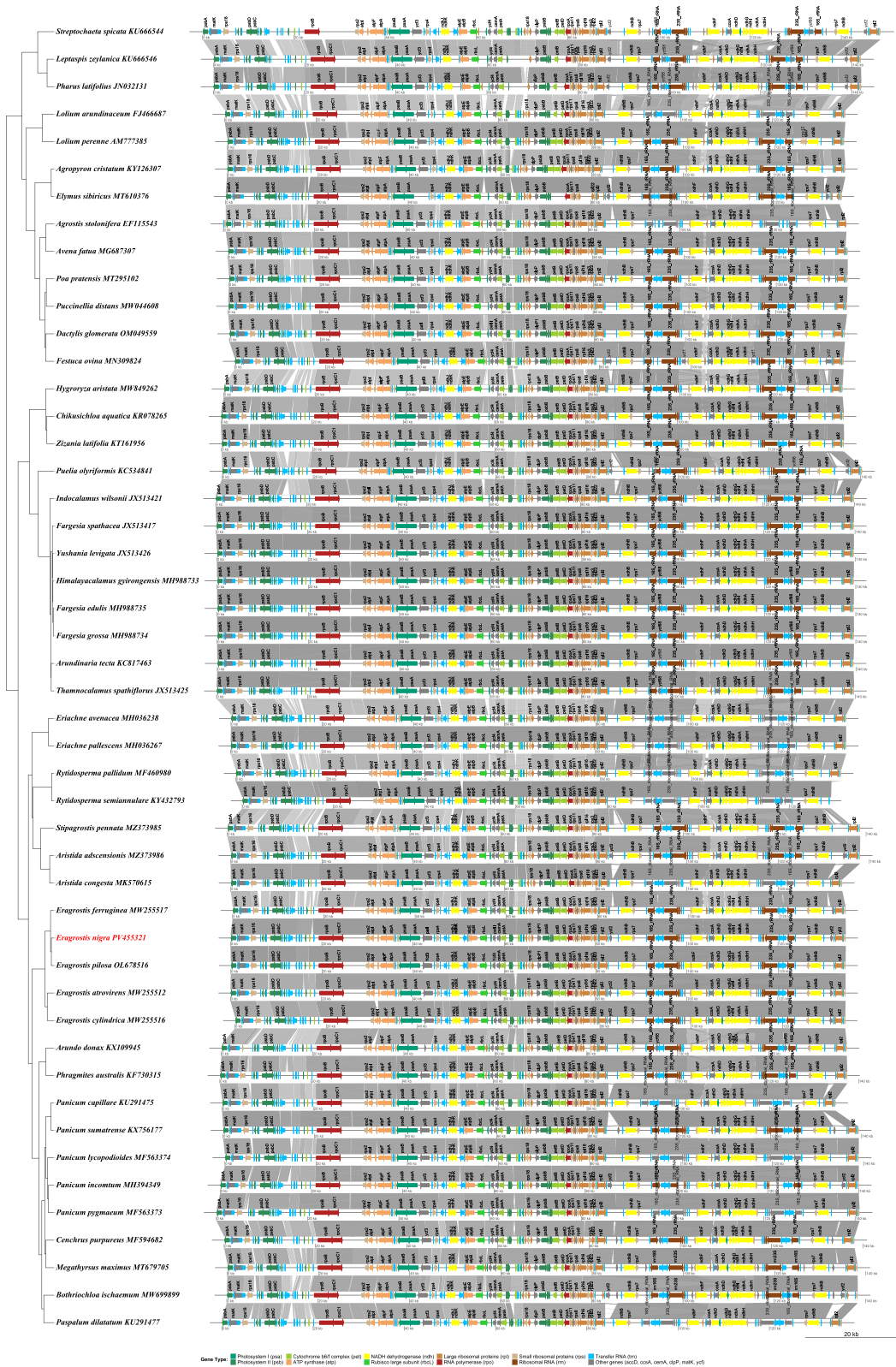


Fig. 9 (See legend on next page.)

(See figure on previous page.)

Fig. 9 Synteny analysis and phylogenetic relationships among chloroplast genomes of 48 Poaceae species. The phylogenetic tree on the left illustrates evolutionary relationships, while the genome maps show gene arrangement and content across species. Each horizontal track represents a linearized chloroplast genome, with genes depicted as directional arrows indicating transcriptional orientation. Colored blocks represent different functional gene categories as indicated in the legend. Gray connecting lines between genomes highlight conserved syntenic regions, with line intensity corresponding to sequence identity levels (darker shading indicates higher conservation). The 47 pairwise comparisons reveal high structural conservation within sub-families and lineage-specific variations at IR boundaries and intergenic regions

exceeding 92% sequence identity in shared regions. In contrast, the PACMAD clade exhibited more variable syntenic patterns, particularly in the SSC region where IR boundary shifts were detected in Chloridoideae and Panicoideae lineages.

Despite this overarching structural stability, the synteny analysis also detected localized variations in gene composition and genome arrangement among different phylogenetic lineages. Specifically, lineage-specific variations were identified in three main categories: (i) IR boundary dynamics, where the *rps19-rpl22* region showed variable positioning at the LSC-IR junction across Pooideae species; (ii) gene content variations, with the presence/absence polymorphism of *ycf68* among Oryzoideae species; and (iii) intergenic spacer length polymorphisms, particularly evident in the *trnS-trnG* and *petN-psbM* regions. These structural divergences were largely congruent with major clade separations, suggesting that genome architecture evolution is closely associated with phylogenetic branching.

Analysis of functional gene categories revealed differential conservation patterns across the chloroplast genome. Photosystem I and II genes (*psa* and *psb* clusters) exhibited the highest syntenic conservation, consistent with their essential role in photosynthetic electron transport. ATP synthase genes showed similarly high conservation, while NADH dehydrogenase genes displayed moderate variability, particularly in the *ndhF* region near the IR-SSC boundary. Ribosomal protein genes demonstrated subfamily-specific patterns, with *rps* genes showing greater positional variability than *rpl* genes. The *ycf1* and *ycf2* genes, despite their hypothetical function annotation, exhibited strong positional conservation across all 48 genomes, supporting their functional significance in chloroplast maintenance. Furthermore, patterns of structural variation corresponded with species' biogeographic distribution, implying that chloroplast genome evolution within Poaceae has been shaped jointly by ancestral diversification and ecological adaptation across varied environments. Notably, C4 photosynthetic lineages (Chloridoideae, Panicoideae, and Aristidoideae) displayed distinct syntenic signatures compared to C3 lineages, particularly in regions surrounding photosystem-related genes, suggesting potential co-evolutionary relationships between genome architecture and photosynthetic pathway diversification. Collectively, these findings indicate that while Poaceae

chloroplast genomes retain a highly conserved core structure, lineage-specific syntenic modifications reflect evolutionary divergence and potential adaptations to distinct ecological niches.

Discussion

Chloroplast genome structural conservation and SSR distribution patterns

Our comparative analysis of 48 Poaceae chloroplast genomes reveals remarkable structural conservation closely linked to the central role of chloroplasts in essential plant functions. The complete chloroplast genome lengths range from 129,095 bp to 148,617 bp, which is highly consistent with the typical size range of angiosperm chloroplast genomes [2], indicating that Poaceae chloroplast genomes follow the universal structural pattern of angiosperms. The newly sequenced *E. nigra* possesses the typical quadripartite structure configuration, including two inverted repeat regions (IRA and IRB), a small single-copy region (SSC), and a large single-copy region (LSC), and this structure remains highly consistent across all 48 species [119–121]. The conserved gene content and arrangement further underscore the strong functional constraints on chloroplast genome evolution, particularly for genes encoding core photosynthetic components, which exhibit extremely high sequence and structural conservation [122, 123]. The evolutionary significance of this conservation resides in the critical role of the chloroplast as the plant's energy conversion center, where any structural variations affecting photosynthetic efficiency could compromise photosynthetic efficiency and reduce plant fitness [124, 125]. The presence and conservation of IR regions not only provide structural stability to the chloroplast genome but also enhance the expression of ribosomal RNA genes through gene dosage effects, which is crucial for chloroplast protein synthesis [126, 127].

SSRs are prone to segmental duplications, deletions, and rearrangements, which influence species evolution and intraspecific genetic diversity [128–130]. Due to their high polymorphism, SSR markers are frequently utilized in molecular breeding, germplasm resource evaluation, and evolutionary research [131–133]. These genetic elements are of great significance as molecular markers in chloroplast genome studies due to their applications in species identification, population genetics, and phylogenetic studies [134, 135]. SSRs in Poaceae chloroplast

genomes exhibit a clear non-random distribution pattern: intergenic regions > introns > coding sequence, aligning with observations in species of Asteraceae [136], *Primula vulgaris* [137], and *Poa annua* [138]. One possible explanation for SSRs presenting highest density in intergenic regions is that these regions experience relatively less selective pressure, allowing accumulation and expansion of repetitive sequences [129]. Additionally, these SSRs might also serve as regulators of gene expression, such as through formation of transcription factor binding sites or organization of transcriptional regulatory elements [139]. From an evolutionary biology perspective, the distribution pattern of SSRs provides important clues for understanding the evolutionary dynamics of chloroplast genomes.

Due to SSR marker ability to show genetic differences and identify background variations, they are invaluable in molecular marker development [140, 141]. The high density distribution of SSRs in intergenic regions reflects their key functions in plant genomes [142]. With further in-depth study of SSR functions and distribution mechanisms, SSRs will play an increasingly important role in the research and breeding of forage grasses, ecological grasses and other plant species. In this study, comparative analysis of chloroplast genomes from 48 Poaceae species revealed significant interspecific variations in the quantity and composition of repetitive sequences. These variations align with previous findings that link repetitive sequence abundance to chloroplast genome stability, suggesting their potential role in genomic structural maintenance or adaptive evolution. For example, inverted repeats can enhance stability by suppressing rearrangements, with instability increasing when they are absent [143], and repetitive sequence expansion or contraction in *Jasminum* may respond to evolutionary pressures to optimize genomic function [144]. The differences in the number, type, and length of repetitive sequences between species provide a potential basis for the development of new molecular genetic markers.

Evolutionary forces shaping codon usage bias in Poaceae chloroplasts

The marked preference for A/U-ending codons is consistent with the observed low GC content, which on average is 38.55%. This indicates that the Poaceae chloroplast genomes have been under directional evolutionary pressure toward A and T enrichment, which supports the “A/T bias in higher plant codons” hypothesis [145]. The GC content at the third codon position is significantly lower than that at the first and second positions. This gradient reflects weaker functional constraints on synonymous sites, thereby allowing greater evolutionary flexibility. CUB affects gene expression efficiency and can also be associated with species’ environmental adaptations,

genetic stability, and genome evolution [146, 147]. From a functional perspective, CUB directly influences translation efficiency and protein expression levels, which has important implications for plant adaptation to different environmental conditions [148, 149].

The favored use of A/U-ending codons may reflect the compositional characteristics of the chloroplast tRNA pool, optimizing the efficiency of codon-anticodon pairing during translation [150, 151]. This optimization is especially important for rapid synthesis of photosynthetic proteins, as photosynthesis requires large quantities of protein complexes to sustain efficient light energy conversion [152, 153]. ENC value analysis showed that most species had an effective number of codons exceeding 35, indicating moderate CUB intensity. This “moderate bias” represents an evolutionary balance between optimizing translation efficiency and maintaining genetic code flexibility. The interspecies consistency of codon usage patterns suggests that Poaceae plants have adopted similar evolutionary strategies for chloroplast protein synthesis, potentially explaining the successful global adaptation of the family [120]. The observed CUB may result from multiple concurrent selective pressures, including translational selection for efficient protein synthesis, mRNA structural stability, and optimal protein folding kinetics [57]. This multi-level selection has fine-tuned codon usage patterns across Poaceae, representing a sophisticated evolutionary solution to maintaining high photosynthetic efficiency under diverse environmental challenges.

CUB in Poaceae chloroplast genomes is influenced by multiple biological factors including gene expression, gene length [154, 155], mutation pressure [156], GC content [157], and natural selection [158], with the latter two being considered the primary evolutionary forces. In the chloroplast genomes of 48 grasses, we observed significant CUB, which may be closely associated with the efficiency and stability of gene expression. Notably, the observed codon usage pattern suggests that natural selection and mutation pressure play significant roles in shaping CUB in Poaceae forage and ecological grasses. ENC-GC_{3s} plot analysis shows that most data points are distributed above or deviate from the expected curve, and this deviation pattern strongly suggests that selective forces rather than pure mutational pressure influence codon choice [159]. PR2 plot analysis further confirms this conclusion, with the concentrated distribution of data points in the fourth quadrant indicating selective preference for G and T at the third codon position. This non-random distribution pattern cannot be explained solely by mutational pressure [160]. Neutrality plot analysis then provides the most direct quantitative evidence, with extremely weak correlation between GC₁₂ and GC₃ showing regression coefficients ranging from 0.0037 to

0.1531, suggesting that mutational pressure contributes ranging from 0.37% to 15.31% to codon bias, while natural selection and other evolutionary forces contribute from 84.69% to 99.96%. This selection-dominated evolutionary pattern has profound biological significance, indicating that codon usage in chloroplast genomes is not a random process but rather a product of precise regulatory optimization for gene expression efficiency [161].

Natural selection can influence different aspects of the genome: first, translational selection, which favors codons that match the most abundant tRNAs, and thus speeds up and makes more accurate translation [162]; second, mRNA structural selection, where the use of specific codons may stabilize secondary structures of mRNA [163]; and third, protein folding selection where specific codon usage may optimize protein co-translational folding processes by influencing translation speed [164]. The RSCU analysis also validates that ecological factors exert a notable influence on the chloroplast codon usage patterns of Poaceae species. The Pooideae subfamily, with greater ecological diversity spanning arctic to temperate zones, exhibits higher variability in codon bias. However, the Chloridoideae subfamily, adapted to arid and semi-arid environments with relatively constrained ecological niches, shows more uniform codon bias. This variation in codon usage patterns, combined with the widespread presence of highly preferred codons across Poaceae, further confirms the strong role of natural selection. By optimizing translational efficiency and tRNA abundance, natural selection drives the evolution of chloroplast codons to meet the adaptive demands of different ecological niches, providing key insights into C_3/C_4 photosynthetic differences and habitat-specific functional adaptations. This optimization underscores the importance of efficient gene expression in chloroplasts, where rapid and accurate protein synthesis is crucial for maintaining photosynthetic competence under varying environmental conditions [165].

Ka/Ks ratio analysis revealed significant heterogeneity in selection pressure within Poaceae chloroplast genomes, reflecting differences in functional importance of different genes in plant essential metabolic processes. 94.67% of genes experience strong purifying selection ($Ka/Ks < 0.5$), and this proportion is much higher than most nuclear genomes, indicating that chloroplast genomes are subject to extremely strong functional constraints. This study and profound previous studies on the chloroplast genomes of Poaceae plants have all shown that core photosynthetic genes such as *psbA*, *rbcL*, and *atpB* exhibited the lowest Ka/Ks ratios [166], as these genes encode proteins directly involved in key processes such as light energy capture, electron transport, and ATP synthesis, where any amino acid changes could significantly impact photosynthetic efficiency. In contrast,

some genes such as *rpl12*, *ndhI*, and *matK* show relatively higher Ka/Ks ratios [167], suggesting that these genes may either be subject to weaker selective constraints or experience some form of adaptive evolution. This variation in the pattern of selection pressure has important evolutionary biological significance.

Phylogenomic relationships and genome evolution patterns

Phylogenetic trees based on chloroplast genomes provide detailed records of the evolution of Poaceae and serve as a primary reference for linking genomic structural variation to phylogenetic differentiation [168]. The selection of Bromeliaceae and Typhaceae as outgroups is well-supported by recent phylogenomic studies demonstrating their position as early-diverging lineages within Poales. This outgroup selection ensures appropriate rooting of the Poaceae phylogeny while maintaining sufficient phylogenetic distance to avoid long-branch attraction artifacts. Our comprehensive sampling of 17 outgroup species from these two families provides robust resolution of the Poaceae root. *Eragrostis* forms a monophyletic group in the tree, supporting its taxonomic monophyly. The close relationship between *E. nigra* and *E. pilosa* aligns with morphological classification, validating the use of chloroplast genomes for species identification and phylogenetic inference. The relatively close relationship between *Eragrostis* and *Rytidosperma* may reflect shared ancestry or convergent ecological adaptation strategies. The approximately maximum-likelihood approach implemented in FastTree with the WAG+G model provides a computationally efficient method for constructing large-scale phylogenies from protein sequences while maintaining accuracy comparable to traditional maximum-likelihood methods. The WAG model is particularly appropriate for chloroplast protein evolution, as it was derived from a large database of globular proteins and performs well for diverse protein families. The aLRT support values provide a rapid assessment of branch reliability, with values exceeding 0.90 generally indicating well-supported clades.

The clear resolution of all 12 Poaceae subfamilies in our phylogenetic analysis demonstrates the utility of chloroplast protein sequences for resolving grass evolutionary relationships. The topology recovered in this study is largely consistent with the most recent Poaceae classification, with the BOP clade (Bambusoideae, Oryzoideae, Pooideae) and PACMAD clade (Panicoideae, Arundinoideae, Chloridoideae, Micrairoideae, Aristidoideae, Danthonioideae) forming the two major lineages of core grasses. The basal position of Anomochlooideae, Pharoideae, and Puelioideae further supports the deep divergence of these early-diverging grass lineages. Phylogenetic trees clarify evolutionary branching within

Poaceae, providing a framework to study the relatedness of species and track evolutionary trajectories, with implications for plant taxonomy, evolutionary ecology, and population genetics [169].

Synteny analysis confirms a strong correlation between genomic structure and phylogenetic relationships. Closely related species exhibit high similarity in gene arrangement and genome structure, while distantly related species show greater genomic differences. This supports the hypothesis of chloroplast genome conservation, consistent with studies in other plant lineages [4, 170, 171]. The quantitative framework employed in our synteny analysis, utilizing 47 pairwise comparisons with graduated sequence identity thresholds, provides robust statistical support for these observations. The observed genomic stability across species reflects the essential role of chloroplasts in photosynthesis and energy production, which requires genome integrity to maintain vital functions. Despite interspecific variations in morphology and ecology, the core chloroplast genome structure remains conserved, highlighting shared evolutionary requirements for basic biological processes [146]. The conservation of synteny in Poaceae chloroplast genomes also has practical implications for molecular marker development. The highly conserved regions identified in our analysis, particularly the *psaA-psaB-rps14* and *atpB-rbcL* gene clusters, provide excellent targets for universal primers in phylogenetic studies spanning deep evolutionary timescales. Conversely, the variable regions at IR boundaries and in specific intergenic spacers offer potential for population-level genetic markers and cultivar identification in agronomically important species.

Conclusions

We present a comprehensive comparative analysis of chloroplast genomes from 48 Poaceae forage and ecological grass species, revealing key patterns in genome structure, codon usage, SSR distribution, and selection pressure. Our findings indicate: (i) Genome composition: Low GC content (38.16%–39.04%, mean 38.55%) with a clear A/U bias at the third codon position; cool-season grasses exhibit stronger codon bias than warm-season species. (ii) Codon usage: Weak to moderate codon bias driven primarily by natural selection, with mutational pressure contributing only 0.37%–22.27% to overall bias; selection accounts for >90% of codon preference in most species. (iii) SSR distribution: Non-random, with the highest abundance in intergenic regions (52.56%), followed by introns (28.23%) and CDS (19.21%). Mononucleotide and octanucleotide repeats dominate, and 11 SSR types are nearly universal across species, underscoring their potential as phylogenetic markers. (iv) Selection pressure: Nearly 95% of chloroplast genes are under strong purifying selection,

especially photosystem and ribosomal protein genes, reflecting functional constraints; genes like *ndhF* and *rpl22* exhibit adaptive variations. (v) Phylogenomics: Structural conservation correlates strongly with phylogenetic relationships; *Eragrostis* forms a monophyletic clade closely related to Rytidosperma, confirming chloroplast genomes' utility for taxonomic and evolutionary studies. We believe that these insights clarify the adaptive evolution of Poaceae chloroplasts, providing a molecular framework for understanding functional constraints and ecological adaptation. They also generate valuable resources for future evolutionary, ecological, and breeding research, particularly for forage and C₄ grasses under changing environmental conditions.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12864-025-12498-6>.

Supplementary Material 1.

Acknowledgements

We gratefully acknowledge the High-Performance Computing Platform of Beijing Forestry University for providing computational resources. We also thank the anonymous reviewers for their constructive comments that significantly improved the manuscript. Special thanks are extended to Rashid Khan (Department of Agronomy, Faculty of Agriculture, Gomal University, KPK, Pakistan) and Qaisar Hayyat (Department of Plant Breeding and Genetics, Faculty of Agricultural Sciences, University of Punjab, Lahore, Pakistan) for their invaluable contributions to language polishing, meticulous revision, and insightful suggestions during the preparation of this paper.

Authors' contributions

Conceptualization, Y.G.; methodology, X.G. and Y.G.; software, Y.G.; validation, X.G. and Y.G.; formal analysis, X.G. and Y.G.; sample collection, H.W.; data curation, X.G., H.W. and Y.G.; writing—original draft preparation, X.G., Y.X., H.W., H.T., R.J., Z.L., and Y.G.; writing—review and editing, X.G., Y.X., H.W., H.T., R.J., Z.L., and Y.G.; project administration, Y.G.; funding acquisition, Y.G. All authors have read and agreed to the published version of the manuscript.

Funding

This research was financially supported by the Fundamental Research Funds for the Central Universities (No: ZZK202503).

Data availability

The data sets generated in this study are available in GenBank, National Center for Biotechnology Information (<http://www.ncbi.nlm.nih.gov/genbank/>) under the accession number of plastome: PV455321.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹School of Grassland Science, Beijing Forestry University, Beijing 100083, China

Received: 5 May 2025 / Accepted: 26 December 2025

Published online: 27 January 2026

References

- Soreng RJ, Peterson PM, Romaschenko K, Davidse G, Teisher JK, Clark LG, et al. A worldwide phylogenetic classification of the Poaceae (Gramineae) II: an update and a comparison of two 2015 classifications. *J Syst Evol.* 2017;55(4):259–90.
- Hu Y, Sun Y, Zhu QH, Fan L, Li J. Poaceae chloroplast genome sequencing: great leap forward in recent ten years. *Curr Genomics.* 2023;23(6):369–84.
- Kirschner JA, Hoorn C. The onset of grasses in the Amazon drainage basin, evidence from the fossil record. *Front Biogeography.* 2020;12(2):e44827.
- Burke SV, Wysocki WP, Zuloaga FO, Craine JM, Pires JC, Edgar PP, et al. Evolutionary relationships in Panicoid grasses based on plastome phylogenomics (Panicoidae; Poaceae). *BMC Plant Biol.* 2016;16(1):140.
- Azizbek MV, Bekzod MJ, Akmal AL, Azizbek AS, Odilbek MT, Ozodbek AS. Distribution area and yield indicators of *Poa bulbosa* L. in Uzbekistan. *Am J Plant Sci.* 2023;14(08):869–81.
- Bouchenak-Khelladi Y, Verboom GA, Savolainen V, Hodkinson TR. Biogeography of the grasses (Poaceae): a phylogenetic approach to reveal evolutionary history in geographical space and geological time. *Bot J Linn Soc.* 2010;162(4):543–57.
- Kellogg EA. Relationships of cereal crops and other grasses. *Proc Natl Acad Sci U S A.* 1998;95(5):2005–10.
- Kellogg EA. Flowering plants. *Monocots*: Springer; 2016.
- Wills B, Douglas G, McKenzie J, Trainor K, Foote A. *Thinopyrum intermedium* (Host) Barkw. & Dewey—a review, and evaluation of intermediate and pubescent wheatgrass for dryland agriculture in New Zealand. 1998.
- Mårtensson L-MD, Barreiro A, Li S, Jensen ES. Agronomic performance, nitrogen acquisition and water-use efficiency of the perennial grain crop *Thinopyrum intermedium* in a monoculture and intercropped with alfalfa in Scandinavia. *Agron Sustain Dev.* 2022;42(2):21.
- Pardo J, VanBuren R. Evolutionary innovations driving abiotic stress tolerance in C4 grasses and cereals. *Plant Cell.* 2021;33(11):3391–401.
- Ingram AL, Doyle JJ. The origin and evolution of *Eragrostis tef* (Poaceae) and related polyploids: evidence from nuclear waxy and plastid rps16. *Am J Bot.* 2003;90(1):116–22.
- Wang R. The occurrence of C4 plants and their morphological functional types in the vegetation of Xinjiang, China. *Photosynthetica.* 2006;44(2):293–8.
- Stewart RR. The grasses of northwest India. *Brittonia.* 1945;5(4):404–68.
- Peterson PM, Romaschenko K, Johnson G. A classification of the Chloridoideae (Poaceae) based on multi-gene phylogenetic trees. *Mol Phylogenet Evol.* 2010;55(2):580–98.
- Faye N, Diallo A, Sagna MB, Peiry JL, Sarr PS, Guisse A. Influence of anthropic and eco-hydrological factors on the floristic diversity of the herbaceous vegetation around the temporary ponds in Ferlo, northern Senegal. *J Plant Ecol.* 2022;15(1):26–38.
- Davidson JM. Evaluating teff grass as a summer forage. 2018.
- Tóth SZ, Lőrincz T, Szarka A. Concentration does matter: the beneficial and potentially harmful effects of ascorbate in humans and plants. *Antioxid Redox Signal.* 2018;29(15):1516–33.
- Daniell H, Lin C-S, Yu M, Chang W-J. Chloroplast genomes: diversity, evolution, and applications in genetic engineering. *Genome Biol.* 2016;17(1):134.
- Palmer JD. Comparative organization of chloroplast genomes. 1985.
- McFadden GI. Primary and secondary endosymbiosis and the origin of plastids. *J Phycol.* 2001;37(6):951–9.
- Martin W, Herrmann RG. Gene transfer from organelles to the nucleus: how much, what happens, and why? *Plant Physiol.* 1998;118(1):9–17.
- Gray MW, Doolittle WF. Has the endosymbiont hypothesis been proven? *Microbiol Rev.* 1982;46(1):1–42.
- Martin WF, Garg S, Zimorski V. Endosymbiotic theories for eukaryote origin. *Philos Trans R Soc Lond B Biol Sci.* 2015;370(1678):20140330.
- Yi X, Dong J, Chen J, Zhou H, Wu T, Gao S, et al. Molecular and morphological evidence support a new species of Rosaceae *Prunus* subg. *Cerasus* from Wuyishan National Park, southeast China. *PhytoKeys.* 2024;237:269.
- Wicke S, Schneeweiss GM, Depamphilis CW, Müller KF, Quandt D. The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. *Plant Mol Biol.* 2011;76(3):273–97.
- Park I, Kim WJ, Yeo S-M, Choi G, Kang Y-M, Piao R, et al. The complete chloroplast genome sequences of *Fritillaria ussuriensis* Maxim. and *Fritillaria cirrhosa* D. Don, and comparative analysis with other *Fritillaria* species. *Molecules.* 2017;22(6):982.
- Huang Y, Wang J, Yang Y, Fan C, Chen J. Phylogenomic analysis and dynamic evolution of chloroplast genomes in Salicaceae. *Front Plant Sci.* 2017;8:1050.
- Liu D, Cui Y, Li S, Bai G, Li Q, Zhao Z, et al. A new chloroplast DNA extraction protocol significantly improves the chloroplast genome sequence quality of foxtail millet (*Setaria italica* (L.) P. Beauv.). *Sci Rep.* 2019;9(1):16227.
- Parks M, Cronn R, Liston A. Increasing phylogenetic resolution at low taxonomic levels using massively parallel sequencing of chloroplast genomes. *BMC Biol.* 2009;7(1):84.
- Shaw J, Lickey EB, Schilling EE, Small RL. 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(3):275–88.
- Pei J, Wang Y, Zhuo J, Gao H, Vasupalli N, Hou D, et al. Complete chloroplast genome features of *Dendrocalamus farinosus* and its comparison and evolutionary analysis with other Bambusoideae species. *Genes.* 2022;13(9):1519.
- Shi Y, Chen Z, Jiang J, Li X, Zeng W. Comparative analysis of chloroplast genomes of “Tiantai wu-yao” (*Lindera aggregata*) and taxa of the same genus and different genera. *Genes.* 2024;15(3):263.
- Takamatsu T, Baslam M, Inomata T, Oikawa K, Itoh K, Ohnishi T, et al. Optimized method of extracting rice chloroplast DNA for high-quality plastome resequencing and de novo assembly. *Front Plant Sci.* 2018;9:266.
- Duvall MR, Fisher AE, Columbus JT, Ingram AL, Wysocki WP, Burke SV, et al. Phylogenomics and plastome evolution of the chloridoideae (Chloridoideae; Poaceae). *Int J Plant Sci.* 2016;177(3):235–46.
- Saarela JM, Burke SV, Wysocki WP, Barrett MD, Clark LG, Craine JM, et al. A 250 plastome phylogeny of the grass family (Poaceae): topological support under different data partitions. *PeerJ.* 2018;6:e4299.
- Wysocki WP, Burke SV, Swingley WD, Duvall MR. The first complete plastid genome from Joinvilleaceae (*J. ascendens*; Poales) shows unique and unpredicted rearrangements. *PLoS One.* 2016;11(9):e0163218.
- Sharp PM, Li W-H. The codon adaptation index—a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res.* 1987;15(3):1281–95.
- Bulmer M. The selection-mutation-drift theory of synonymous codon usage. *Genetics.* 1991;129(3):897–907.
- Quax TE, Claessens NJ, Söhl D, van der Oost J. Codon bias as a means to fine-tune gene expression. *Mol Cell.* 2015;59(2):149–61.
- Négadi T. Revealing the genetic code symmetries through computations involving Fibonacci-like sequences and their properties. *Computation.* 2023;11(8):154.
- Zhou M, Long W, Li X. Patterns of synonymous codon usage bias in chloroplast genomes of seed plants. *Forest Stud China.* 2008;10(4):235–42.
- Li X, Liu L, Ren Q, Zhang T, Hu N, Sun J, et al. Analysis of synonymous codon usage bias in the chloroplast genome of five *Caragana*. *BMC Plant Biol.* 2025;25(1):322.
- Sharp PM, Li W-H. An evolutionary perspective on synonymous codon usage in unicellular organisms. *J Mol Evol.* 1986;24(1):28–38.
- Duan H, Zhang Q, Wang C, Li F, Tian F, Lu Y, et al. Analysis of codon usage patterns of the chloroplast genome in *Delphinium grandiflorum* L. reveals a preference for AT-ending codons as a result of major selection constraints. *PeerJ.* 2021;9:e10787.
- Hershberg R, Petrov DA. Selection on codon bias. *Annu Rev Genet.* 2008;42(1):287–99.
- Zhou T, Weems M, Wilke CO. Translationally optimal codons associate with structurally sensitive sites in proteins. *Mol Biol Evol.* 2009;26(7):1571–80.
- Parvathy ST, Udayasuriyan V, Bhadana V. Codon usage bias. *Mol Biol Rep.* 2022;49(1):539–65.
- Hanson G, Collier J. Codon optimality, bias and usage in translation and mRNA decay. *Nat Rev Mol Cell Biol.* 2018;19(1):20–30.
- Morton BR. Selection on the codon bias of chloroplast and cyanelle genes in different plant and algal lineages. *J Mol Evol.* 1998;46(4):449–59.
- Zhang R, Zhang L, Wang W, Zhang Z, Du H, Qu Z, et al. Differences in codon usage bias between photosynthesis-related genes and genetic system-related genes of chloroplast genomes in cultivated and wild solanum species. *Int J Mol Sci.* 2018;19(10):3142.
- Tuller T, Carmi A, Vestsigian K, Navon S, Dorfan Y, Zaborse J, et al. An evolutionarily conserved mechanism for controlling the efficiency of protein translation. *Cell.* 2010;141(2):344–54.
- Drummond DA, Wilke CO. Mistranslation-induced protein misfolding as a dominant constraint on coding-sequence evolution. *Cell.* 2008;134(2):341–52.

54. Xiao M, Hu X, Li Y, Liu Q, Shen S, Jiang T, et al. Comparative analysis of codon usage patterns in the chloroplast genomes of nine forage legumes. *Physiol Mol Biol Plants*. 2024;30(2):153–66.
55. Reis Md, Savva R, Wernisch L. Solving the riddle of codon usage preferences: a test for translational selection. *Nucleic Acids Res*. 2004;32(17):5036–44.
56. Rak R, Dahan O, Pilpel Y. Repertoires of tRNAs: the couplers of genomics and proteomics. *Annu Rev Cell Dev Biol*. 2018;34(1):239–64.
57. Sakamoto W, Miyagishima S-y, Jarvis P. Chloroplast biogenesis: control of plastid development, protein import, division and inheritance. *Arabidopsis Book*. 2008;6:e0110.
58. Spencer PS, Siller E, Anderson JF, Barral JM. Silent substitutions predictably alter translation elongation rates and protein folding efficiencies. *J Mol Biol*. 2012;422(3):328–35.
59. Mehmood F, Li M, Bertolli A, Prosser F, Varotto C. Comparative plastomics of plantains (*Plantago*, Plantaginaceae) as a tool for the development of species-specific DNA barcodes. *Plants*. 2024;13(19):2691.
60. Wu B, Zhu J, Ma X, Jia J, Luo D, Ding Q, et al. Comparative analysis of switchgrass chloroplast genomes provides insights into identification, phylogenetic relationships and evolution of different ecotypes. *Ind Crops Prod*. 2023;205:117570.
61. Young HA, Lanzatella CL, Sarath G, Tobias CM. Chloroplast genome variation in upland and lowland switchgrass. *PLoS One*. 2011;6(8):e23980.
62. Diekmann K, Hodgkinson TR, Wolfe KH, van den Bekerom R, Dix PJ, Barth S. Complete chloroplast genome sequence of a major allogamous forage species, perennial ryegrass (*Lolium perenne* L.). *DNA Res*. 2009;16(3):165–76.
63. Zhang Y, Nie X, Jia X, Zhao C, Biradar SS, Wang L, et al. Analysis of codon usage patterns of the chloroplast genomes in the Poaceae family. *Aust J Bot*. 2012;60(5):461–70.
64. Jung-Hoon L, Ki-Joong K, Bo-Yun K, Young-Dong K. Molecular systematics of Poaceae based on eight chloroplast markers, emphasizing the phylogenetic positions of Korean taxa. *Korean J Plant Taxonomy*. 2022;52(3):127–43.
65. Pandey DK, Chaudhary B. Evolutionary expansion and structural functionalism of the ancient family of profilin proteins. *Gene*. 2017;626:70–86.
66. Yang Z, Nielsen R. Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. *Mol Biol Evol*. 2000;17(1):32–43.
67. Yang L, Deng S, Zhu Y, Da Q. Comparative chloroplast genomics of 34 species in subtribe Swertiinae (Gentianaceae) with implications for its phylogeny. *BMC Plant Biol*. 2023;23(1):164.
68. Rono PC, Dong X, Yang J-X, Mutie FM, Oulo MA, Malombe I, et al. Initial complete chloroplast genomes of *Alchemilla* (Rosaceae): comparative analysis and phylogenetic relationships. *Front Genet*. 2020;11:560368.
69. Liu H, Ye H, Zhang N, Ma J, Wang J, Hu G, et al. Comparative analyses of chloroplast genomes provide comprehensive insights into the adaptive evolution of *Paphiopedilum* (Orchidaceae). *Horticulturae*. 2022;8(5):391.
70. Wolfe KH, Li W-H, Sharp PM. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *Proc Natl Acad Sci U S A*. 1987;84(24):9054–8.
71. Muse SV, Gaut BS. A likelihood approach for comparing synonymous and nonsynonymous nucleotide substitution rates, with application to the chloroplast genome. *Mol Biol Evol*. 1994;11(5):715–24.
72. Huang R, Xiao D, Wang X, Zhan J, Wang A, He L. Genome-wide identification, evolutionary and expression analyses of LEA gene family in peanut (*Arachis hypogaea* L.). *BMC Plant Biol*. 2022;22(1):155.
73. Hermida-Carrera C, Fares MA, Fernandez A, Gil-Pelegrin E, Kapralov MV, Mir A, et al. Positively selected amino acid replacements within the RuBisCO enzyme of oak trees are associated with ecological adaptations. *PLoS One*. 2017;12(8):e0183970.
74. Gao C, Deng Y, Wang J. The complete chloroplast genomes of *Echinacanthus* species (Acanthaceae): phylogenetic relationships, adaptive evolution, and screening of molecular markers. *Front Plant Sci*. 2019;9:1989.
75. Ahmed M, Min DS, Kim DR. Integrating binding and expression data to predict transcription factors combined function. *BMC Genomics*. 2020;21(1):610.
76. Iida S, Miyagi A, Aoki S, Ito M, Kadono Y, Kosuge K. Molecular adaptation of *rbcl* in the heterophyllous aquatic plant *Potamogeton*. *PLoS One*. 2009;4(2):e4633.
77. Kapralov MV, Smith JAC, Filatov DA. Rubisco evolution in C4 eudicots: an analysis of *Amaranthaceae* sensu lato. *PLoS One*. 2012;7(12):e52974.
78. Amiteye S. Basic concepts and methodologies of DNA marker systems in plant molecular breeding. *Heliyon*. 2021;7(10):e08093.
79. 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 U S A*. 1995;92(17):7759–63.
80. Jin J-J, Yu W-B, Yang J-B, Song Y, DePamphilis CW, Yi T-S, et al. GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. *Genome Biol*. 2020;21(1):241.
81. Dierckxens N, Mardulyn P, Smits G. NOVOPlasty: de novo assembly of organelle genomes from whole genome data. *Nucleic Acids Res*. 2017;45(4):e18–e18.
82. Turudić A, Liber Z, Grdiša M, Jakše J, Varga F, Šatović Z. Chloroplast genome annotation tools: prolegomena to the identification of inverted repeats. *Int J Mol Sci*. 2022;23(18):10804.
83. Tillich M, Lehwark P, Pellizzer T, Ulbricht-Jones ES, Fischer A, Bock R, et al. GeSeq—versatile and accurate annotation of organelle genomes. *Nucleic Acids Res*. 2017;45(W1):W6–11.
84. Xia X. Bioinformatics and translation elongation. In: *Bioinformatics and the cell: Modern computational approaches in genomics, proteomics and transcriptomics*. Cham: Springer; 2018. p. 197–238.
85. Wang Z, Cai Q, Wang Y, Li M, Wang C, Wang Z, et al. Comparative analysis of codon bias in the chloroplast genomes of Theaceae species. *Front Genet*. 2022;13:824610.
86. Park H, Sa KJ, Hyun DY, Lee S, Lee JK. Identifying SSR markers related to seed fatty acid content in *Perilla frutescens* L. *Plants*. 2021;10(7):1404.
87. Beier S, Thiel T, Münch T, Scholz U, Mascher M. MISA-web: a web server for microsatellite prediction. *Bioinformatics*. 2017;33(16):2583–5.
88. Fuglsang A. The 'effective number of codons' revisited. *Biochem Biophys Res Commun*. 2004;317(3):957–64.
89. Wright F. The 'effective number of codons' used in a gene. *Gene*. 1990;87(1):23–9.
90. Tian T, Zhang Y, Du W, Wang Z. Comparative genomics of *Triticum*, *Secale*, and *Triticale*: codon usage bias in chloroplast genomes and its implications for evolution and genetic engineering. *Int J Mol Sci*. 2025;26(21):10266.
91. Jia X, Wei J, Chen Y, Zeng C, Deng C, Zeng P, et al. Codon usage patterns and genomic variation analysis of chloroplast genomes provides new insights into the evolution of Aroideae. *Sci Rep*. 2025;15(1):4333.
92. Shi N, Yuan Y, Huang R, Wen G. Analysis of codon usage patterns in complete plastomes of four medicinal *Polygonatum* species (Asparagaceae). *Front Genet*. 2024;15:1401013.
93. He B, Dong H, Jiang C, Cao F, Tao S. Xu L-a: analysis of codon usage patterns in *Ginkgo biloba* reveals codon usage tendency from A/U-ending to G/C-ending. *Sci Rep*. 2016;6(1):35927.
94. Sueoka N. Directional mutation pressure and neutral molecular evolution. *Proc Natl Acad Sci U S A*. 1988;85(8):2653–7.
95. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*. 2013;30(4):772–80.
96. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. TrimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*. 2009;25(15):1972–3.
97. Wang D, Zhang Y, Zhang Z, Zhu J, Yu J. KaKs_calculator 2.0: a toolkit incorporating gamma-series methods and sliding window strategies. *Genomics Proteomics Bioinform*. 2010;8(1):77–80.
98. Hasegawa M, Zhong B, Zhong Y. Adaptive evolution of chloroplast genomes in ancestral grasses. *Plant Signal Behav*. 2009;4(7):623–4.
99. Chen Y, Liu N, Wang H, Luo H, Xie Z, Yu H, et al. Comparative chloroplast genomics of Rutaceae: structural divergence, adaptive evolution, and phylogenomic implications. *Front Plant Sci*. 2025;16:1675536.
100. Liu S, Ni Y, Li J, Zhang X, Yang H, Chen H, et al. CPGview: a package for visualizing detailed chloroplast genome structures. *Mol Ecol Resour*. 2023;23(3):694–704.
101. Chen N, Sha LN, Dong ZZ, Tang C, Wang Y, Kang HY, et al. Complete structure and variation of the chloroplast genome of *Agropyron cristatum* (L.) Gaertn. *Gene*. 2018;640:86–96.
102. Guo X-X, Qu X-J, Zhang X-J, Fan S-J. Comparative and phylogenetic analysis of complete plastomes among aristidoideae species (Poaceae). *Biology*. 2022;11(1):63.
103. Wang Y, Liu P, Li H, Shao C. The complete mitochondrial genome of *Opisthotropis latouchii* (Squamata: Colubridae). *Mitochondr DNA B Resources*. 2019;4(1):1437–8.
104. Burke SV, Clark LG, Triplett JK, Grennan CP, Duvall MR. Biogeography and phylogenomics of new world Bambusoideae (Poaceae), revisited. *Am J Bot*. 2014;101(5):886–91.
105. Yu Q-q, Liu Z-c, Xiong Y, Xiong Y-I, Nie C, Gao H-d, Liu W-h, Ma X. Comparative and phylogenetic analysis of complete chloroplast genome sequences of two wildrye grasses *Elymus sibiricus* and *E. nutans* (Triticeae, Poaceae). 2021.

106. Liu K, Wang R, Guo X-X, Zhang X-J, Qu X-J, Fan S-J. Comparative and phylogenetic analysis of complete chloroplast genomes in Eragrostideae (Chloridoideae, Poaceae). *Plants*. 2021;10(1):109.
107. Wang J, Fu GF, Tembrock LR, Liao XZ, Ge S, Wu ZQ. Mutational meltdown or controlled chain reaction: the dynamics of rapid plastome evolution in the hyperdiversity of Poaceae. *J Syst Evol*. 2023;61(2):328–44.
108. Teisher J, McKain M, Schaal B, Kellogg E. Plastome phylogenetics of tribe Eriachneae and evolution of C₄ photosynthesis in subfamily Micrairoideae (Poaceae). *Syst Bot*. 2019;44(1):32–40.
109. Zhou Y, Zhang Y-Q, Xing X-C, Zhang J-Q, Ren Y. Straight from the plastome: molecular phylogeny and morphological evolution of *Fargesia* (Bambusoideae: Poaceae). *Front Plant Sci*. 2019;10:981.
110. Ma P-F, Zhang Y-X, Zeng C-X, Guo Z-H, Li D-Z. Chloroplast phylogenomic analyses resolve deep-level relationships of an intractable bamboo tribe Arundinarieae (Poaceae). *Syst Biol*. 2014;63(6):933–50.
111. Qiu Y, Hirsch CD, Yang Y, Watkins E. Towards improved molecular identification tools in fine fescue (*Festuca* L., Poaceae) turfgrasses: nuclear genome size, ploidy, and chloroplast genome sequencing. *Front Genet*. 2019;10:1223.
112. Cahoon AB, Sharpe RM, Mysayphonh C, Thompson EJ, Ward AD, Lin A. The complete chloroplast genome of tall fescue (*Lolium arundinaceum*; Poaceae) and comparison of whole plastomes from the family Poaceae. *Am J Bot*. 2010;97(1):49–58.
113. Diekmann K, Hodkinson TR, Fricke E, Barth S. An optimized chloroplast DNA extraction protocol for grasses (Poaceae) proves suitable for whole plastid genome sequencing and SNP detection. *PLoS One*. 2008;3(7):e2813.
114. Zeng C-X, Hollingsworth PM, Yang J, He Z-S, Zhang Z-R, Li D-Z, et al. Genome skimming herbarium specimens for DNA barcoding and phylogenomics. *Plant Methods*. 2018;14(1):43.
115. Piot A, Hackel J, Christin P-A, Besnard G. One-third of the plastid genes evolved under positive selection in PACMAD grasses. *Planta*. 2018;247(1):255–66.
116. Jones SS, Burke SV, Duvall MR. Phylogenomics, molecular evolution, and estimated ages of lineages from the deep phylogeny of Poaceae. *Plant Syst Evol*. 2014;300(6):1421–36.
117. Burke SV, Lin C-S, Wysocki WP, Clark LG, Duvall MR. Phylogenomics and plastome evolution of tropical forest grasses (*Leptaspis*, *Streptochaeta*: Poaceae). *Front Plant Sci*. 2016;7:1993.
118. Wu Z, Gu C, Tembrock LR, Ge S. Limited polymorphisms between two whole plastid genomes in the genus *Zizania* (Zizaniinae). *J Proteomics Bioinformatics*. 2015;8(11):253.
119. Kousar M, Park J. Comparative analysis of the chloroplast genome of *Sicyos angulatus* with other seven species of Cucurbitaceae family. *Genes*. 2023;14(9):1776.
120. Wang R, Liu K, Zhang X-J, Chen W-L, Qu X-J, Fan S-J. Comparative plastomes and phylogenetic analysis of *Cleistogenes* and closely related genera (Poaceae). *Front Plant Sci*. 2021;12:638597.
121. Kumar S, Mohapatra T. Epigenetic modifications in genome help remembering the stress tolerance strategy adopted by the plant. *Front Biosci Landmark*. 2024;29(3):126.
122. Li B, Zheng Y. Dynamic evolution and phylogenomic analysis of the chloroplast genome in Schisandraceae. *Sci Rep*. 2018;8(1):9285.
123. Tsuji S, Ueda K, Nishiyama T, Hasebe M, Yoshikawa S, Konagaya A, et al. The chloroplast genome from a lycophyte (microphyllophyte), *Selaginella uncinata*, has a unique inversion, transpositions and many gene losses. *J Plant Res*. 2007;120(2):281–90.
124. Li D-M, Zhu G-F, Yu B, Huang D. Comparative chloroplast genomes and phylogenetic relationships of *Aglaonema modestum* and five variegated cultivars of *Aglaonema*. *PLoS One*. 2022;17(9):e0274067.
125. Long L, Li Y, Wang S, Liu Z, Wang J, Yang M. Complete chloroplast genomes and comparative analysis of *Ligustrum* species. *Sci Rep*. 2023;13(1):212.
126. Li H, Guo Q, Xu L, Gao H, Liu L, Zhou X. CPJSDraw: analysis and visualization of junction sites of chloroplast genomes. *PeerJ*. 2023;11:e15326.
127. Pombert J-F, Lemieux C, Turmel M. The complete chloroplast DNA sequence of the green alga *Oltmannsiellopsis viridis* reveals a distinctive quadripartite architecture in the chloroplast genome of early diverging ulvophytes. *BMC Biol*. 2006;4(1):3.
128. Biswas MK, Biswas D, Bagchi M, Yi G, Deng G. A comprehensive plant microRNA simple sequence repeat marker database to accelerate genetic improvements in crops. *Agronomy*. 2021;11(1):2298.
129. Zhao M, Shu G, Hu Y, Cao G, Wang Y. Pattern and variation in simple sequence repeat (SSR) at different genomic regions and its implications to maize evolution and breeding. *BMC Genomics*. 2023;24(1):136.
130. Mokhtar MM, Alsamman AM, El Allali A. Megassr: a web server for large scale microsatellite identification, classification, and marker development. *Front Plant Sci*. 2023;14:1219055.
131. Dong M, Yu N, Li R, He D, Yuan Z, Yang J, et al. Fingerprinting Chinese sweetgum (*Liquidambar formosana* Hance) accessions and constructing a core collection using newly developed SSR markers. *Forests*. 2025;16(2):281.
132. Gonzaga ZJ, Aslam K, Septiningsih EM, Collard BC. Evaluation of SSR and SNP markers for molecular breeding in rice. 2015.
133. 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(11):260.
134. Yang J-B, Tang M, Li H-T, Zhang Z-R, Li D-Z. Complete chloroplast genome of the genus *Cymbidium*: lights into the species identification, phylogenetic implications and population genetic analyses. *BMC Evol Biol*. 2013;13(1):84.
135. Pugh T, Fouet O, Risterucci A-M, Brottier P, Abouladze M, Deletrez C, et al. A new cacao linkage map based on codominant markers: development and integration of 201 new microsatellite markers. *Theor Appl Genet*. 2004;108(6):1151–61.
136. Meng D, Xiaomei Z, Wenzhen K, Xu Z. Detecting useful genetic markers and reconstructing the phylogeny of an important medicinal resource plant, *Artemisia selengensis*, based on chloroplast genomics. *PLoS One*. 2019;14(2):e0211340.
137. Ren T, Yang Y, Zhou T, Liu Z-L. Comparative plastid genomes of *Primula* species: sequence divergence and phylogenetic relationships. *Int J Mol Sci*. 2018;19(4):1050.
138. Ren J, Tian J, Jiang H, Zhu X-X, Mutie FM, Wanga VO, et al. Comparative and phylogenetic analysis based on the chloroplast genome of *Coleanthus subtilis* (Tratt.) Seidel, a protected rare species of monotypic genus. *Front Plant Sci*. 2022;13:828467.
139. Neumann A, Meinke S, Goldammer G, Strauch M, Schubert D, Timmermann B, et al. Alternative splicing coupled mRNA decay shapes the temperature-dependent transcriptome. *EMBO Rep*. 2020;21(12):e51369.
140. Cheng Y, Yang Y, Wang Z, Qi B, Yin Y, Li H. Development and characterization of EST-SSR markers in *Taxodium zhongshansa*. *Plant Mol Biol Rep*. 2015;33(6):1804–14.
141. Srdić J, Nikolić A, Pajić Z. SSR markers in characterization of sweet corn inbred lines. *Genetika*. 2008;40(2):169–77.
142. Wang Y, Yang C, Jin Q, Zhou D, Wang S, Yu Y, et al. Genome-wide distribution comparative and composition analysis of the SSRs in Poaceae. *BMC Genet*. 2015;16(1):18.
143. Li J, Cullis C. Comparative analysis of 84 chloroplast genomes of *Tylosema esculentum* reveals two distinct cytotypes. *Front Plant Sci*. 2023;13:1025408.
144. Xu X, Huang H, Lin S, Zhou L, Yi Y, Lin E, et al. Twelve newly assembled jasmine chloroplast genomes: unveiling genomic diversity, phylogenetic relationships and evolutionary patterns among Oleaceae and *Jasminum* species. *BMC Plant Biol*. 2024;24(1):331.
145. Campbell WH, Gowri G. Codon usage in higher plants, green algae, and cyanobacteria. *Plant Physiol*. 1990;92(1):1–11.
146. Hsieh Y-C, Chang C-T, Chung J-D, Hwang S-Y. Demographic history and adaptive synonymous and nonsynonymous variants of nuclear genes in *Rhododendron oldhamii* (Ericaceae). *Sci Rep*. 2020;10(1):16658.
147. Yang Q, Xin C, Xiao Q-S, Lin Y-T, Li L, Zhao J-L. Codon usage bias in chloroplast genes implicate adaptive evolution of four ginger species. *Front Plant Sci*. 2023;14:1304264.
148. Sun M, Li S, Gong Q, Xiao Y, Peng F. Leucine contributes to copper stress tolerance in peach (*Prunus persica*) seedlings by enhancing photosynthesis and the antioxidant defense system. *Antioxidants*. 2022;11(12):2455.
149. Ingrisano R, Tosato E, Trost P, Gurrieri L, Sparla F. Proline, cysteine and branched-chain amino acids in abiotic stress response of land plants and microalgae. *Plants*. 2023;12(19):3410.
150. Urban C, Beier H. Cysteine tRNAs of plant origin as novel UGA suppressors. *Nucleic Acids Res*. 1995;23(22):4591–7.
151. Rocha EP. Codon usage bias from tRNA's point of view: redundancy, specialization, and efficient decoding for translation optimization. *Genome Res*. 2004;14(11):2279–86.
152. Zalai Y, Margalot M, Tuller T. On the ribosomal density that maximizes protein translation rate. *PLoS One*. 2016;11(11):e0166481.
153. Fages-Lartaud M, Hundvin K, Hohmann-Marriott MF. Mechanisms governing codon usage bias and the implications for protein expression in the chloroplast of *Chlamydomonas reinhardtii*. *Plant J*. 2022;112(4):919–45.

154. Ingvarsson PK. Gene expression and protein length influence codon usage and rates of sequence evolution in *Populus tremula*. *Mol Biol Evol*. 2007;24(3):836–44.
155. Duret L, Mouchiroud D. Expression pattern and surprisingly, gene length shape codon usage in *Caenorhabditis*, *Drosophila*, and *Arabidopsis*. *Proc Natl Acad Sci U S A*. 1999;96(8):4482–7.
156. Sheng J, She X, Liu X, Wang J, Hu Z. Comparative analysis of codon usage patterns in chloroplast genomes of five *Miscanthus* species and related species. *PeerJ*. 2021;9:e12173.
157. Liu Q, Xue Q. Comparative studies on codon usage pattern of chloroplasts and their host nuclear genes in four plant species. *J Genet*. 2005;84(1):55–62.
158. Wang Z-K, Liu Y, Zheng H-Y, Tang M-Q, Xie S-Q. Comparative analysis of codon usage patterns in nuclear and chloroplast genome of *dalbergia* (Fabaceae). *Genes*. 2023;14(5):1110.
159. Lee J, Kang Y, Shin SC, Park H, Lee H. Combined analysis of the chloroplast genome and transcriptome of the Antarctic vascular plant *Deschampsia antarctica* Desv. *PLoS One*. 2014;9(3):e92501.
160. Zhou J, Zhang S, Wang J, Shen H, Ai B, Gao W, et al. Chloroplast genomes in *Populus* (Salicaceae): comparisons from an intensively sampled genus reveal dynamic patterns of evolution. *Sci Rep*. 2021;11(1):9471.
161. Tyagi S, Kabade PG, Gnanapragasam N, Singh UM, Gurjar AKS, Rai A, et al. Codon usage provide insights into the adaptation of rice genes under stress condition. *Int J Mol Sci*. 2023;24(2):1098.
162. Kalkus A, Barrett J, Ashok T, Morton BR. Evidence from simulation studies for selective constraints on the codon usage of the Angiosperm *psbA* gene. *PLoS Comput Biol*. 2021;17(10):e1009535.
163. Faure G, Ogurtsov AY, Shabalina SA, Koonin EV. Role of mRNA structure in the control of protein folding. *Nucleic Acids Res*. 2016;44(22):10898–911.
164. Pechmann S, Willmund F, Frydman J. The ribosome as a hub for protein quality control. *Mol Cell*. 2013;49(3):411–21.
165. Yang R, Xue Y, He X, Zhang T. Comprehensive analysis of chloroplast genomes in leguminous forage species: codon usage, phylogenetic relationships, and evolutionary insights. *Agronomy*. 2025;15(4):765.
166. Liu Q, Li X, Li M, Xu W, Schwarzacher T, Heslop-Harrison JS. Comparative chloroplast genome analyses of *Avena*: insights into evolutionary dynamics and phylogeny. *BMC Plant Biol*. 2020;20(1):406.
167. Feng L, Zhao G, An M, Wang C, Yin Y. Complete chloroplast genome sequences of the ornamental plant *Prunus cistena* and comparative and phylogenetic analyses with its closely related species. *BMC Genomics*. 2023;24(1):739.
168. Xu X, Wang D. Comparative chloroplast genomics of *Corydalis* species (Papaveraceae): evolutionary perspectives on their unusual large scale rearrangements. *Front Plant Sci*. 2021;11:600354.
169. Wei CX, Jardine PE, Mao LM, Mander L, Li M, Gosling WD, et al. Grass pollen surface ornamentation is diverse across the phylogeny: evidence from northern South America and the global literature. *J Syst Evol*. 2024;62(4):687–701.
170. Li C, Cai C, Tao Y, Sun Z, Jiang M, Chen L, et al. Variation and evolution of the whole chloroplast genomes of *Fragaria* spp. (Rosaceae). *Front Plant Sci*. 2021;12:754209.
171. Wang Y, Wen F, Hong X, Li Z, Mi Y, Zhao B. Comparative chloroplast genome analyses of *Paraboea* (Gesneriaceae): insights into adaptive evolution and phylogenetic analysis. *Front Plant Sci*. 2022;13:1019831.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.