BMC Plant Biology

RESEARCH

Open Access

Integrative chloroplast genomics of Polygonaceae: evolutionary dynamics, codon optimization, and phylogenetic resolution

Haodi Wang[1], Zishuo Wang[1], Zhuofan Liu[1], Guy Smagghe[2], Xiaoyan Zhao[1], Dong Li[1] and Yunpeng Gai[1*]

Abstract

Background The Polygonaceae family comprises approximately 1,200 species of pharmacological, economic, and ecological importance, including medicinal genera such as *Rheum*, *Polygonum*, and *Fagopyrum*, which produce bioactive anthraquinones, stilbenes, and flavonoids. Despite their therapeutic importance, comprehensive analyses of chloroplast genomes across this family remain limited. Chloroplast genomes provide valuable molecular markers for evolutionary studies, phylogenetic reconstruction, and quality control of herbal medicines.

Results This study comprehensively analyzed chloroplast genomes across 48 Polygonaceae species, focusing on genome architecture, gene content, codon usage patterns, selection pressures, simple sequence repeats (SSRs), and phylogenetic relationships. All chloroplast genomes exhibited conserved quadripartite structures, with genome sizes ranging from 155,838 – 179,064 bp and GC contents ranging from 36.59 to 38.22%, demonstrating AT enrichment and a consistent bias toward A/U-ending codons. RSCU analysis revealed six universally optimal codons (AGA, GAU, GCU, UAU, UCU, and UUA) conserved across the family. ENC values (46.36–48.28) indicated weak overall codon bias; however, ENC-GC3 plots, neutrality, and PR2 analyses confirmed that natural selection, rather than mutational pressure, predominantly shaped synonymous codon usage. Ka/Ks analysis revealed that 89.1% of gene pairs were under strong purifying selection, with photosystem-related genes the most conserved. A total of 4,593 SSR loci were identified, predominantly mononucleotide repeats concentrated in intergenic regions. Collinearity analyses demonstrated high structural conservation within genera, with coding regions more conserved than noncoding sequences. Phylogenetic reconstruction delineated major lineages and suggested that subfamily-specific codon usage patterns correlate with ecological adaptation.

Conclusions This study provides comprehensive insights into chloroplast genome evolution in Polygonaceae, highlighting the predominant role of natural selection in codon usage bias and identifying conserved genomic features, SSR resources, and phylogenetic relationships. These findings establish a foundation for molecular marker development, quality control of medicinal species, and evolutionary investigations of this pharmacologically important plant family.

*Correspondence:
Yunpeng Gai
gaiyunpeng@bjfu.edu.cn

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

Keywords Polygonaceae, Chloroplast genome, Codon usage bias, Simple sequence repeats (SSRs), Purification selection, Phylogenomics, RSCU analysis, Genome evolution

Background

Polygonaceae is a globally distributed plant family of considerable ecological, economic, and pharmacological importance within the order Caryophyllales [1, 2]. This morphologically diverse family comprises approximately 48 genera and 1,200 species, including herbs, shrubs, trees, and lianas, inhabiting temperate, tropical, and subtropical regions worldwide [3]. The characteristic features of Polygonaceae include swollen stem nodes; alternate simple leaves with sheathing ocreae; and inflorescences typically arranged in spikes, racemes, or panicles [4]. Molecular phylogenetic studies have revealed two major subfamilies, Polygonoideae and Eriogonoideae, with Coccoloboideae occasionally treated as a third; however, subfamily boundaries remain under investigation pending comprehensive genomic analyses [5, 6].

Species of Polygonaceae produce a rich diversity of bioactive secondary metabolites with demonstrated pharmacological properties [7, 8]. For example, anthraquinones and stilbenes, which are particularly abundant in the tribe Polygoneae, exhibit anti-inflammatory, antioxidant, antimicrobial, and immunomodulatory activities [9, 10]. *Fagopyrum esculentum* and *F. tataricum* demonstrate metabolism-modulating properties in livestock, whereas *Rheum* species provide compounds with cardiovascular and hepatoprotective potential [11]. In contrast, some members, such as *Reynoutria japonica*, are aggressive invasive species owing to their rapid growth, allelopathy, and tolerance to environmental stress [12]. This duality underscores the necessity of genomic investigations to elucidate the molecular mechanisms underlying both therapeutically beneficial traits and ecological invasiveness [13]. Owing to their conserved structure, uniparental inheritance, and moderate evolutionary rates, chloroplast genomes represent powerful tools for investigating plant evolution, phylogenetic relationships, and adaptive mechanisms [14–16]. In addition to photosynthesis, chloroplasts participate in critical metabolic pathways, including fatty acid synthesis, amino acid biosynthesis, and secondary metabolite production, which are pathways directly relevant to the pharmacological activities of many Polygonaceae species [17, 18]. Typically, angiosperm chloroplast genomes exhibit a quadripartite structure consisting of a large single-copy (LSC) region, a small single-copy (SSC) region, and a pair of inverted repeats (IRs), encoding 120–130 genes, including protein-coding genes (PCGs), tRNAs, and rRNAs [19–21]. These genes are highly conserved, making chloroplast genomes ideal for studies of genetic variation, molecular systematics, and species identification [15, 22, 23].

A key feature of chloroplast genomes is codon usage bias (CUB), which reflects nonrandom patterns of synonymous codon usage [24–27]. Codon usage is influenced by multiple factors, including nucleotide composition, mutational pressure, natural selection, gene expression levels, tRNA availability, and codon position within genes [28]. CUB has functional consequences for protein translation efficiency and accuracy, as well as for adaptive evolution, genetic selection, and functional differentiation [29–31]. In chloroplast genomes, codon usage patterns reveal evolutionary pressures and adaptive strategies, which that differ from those in nuclear genomes [14, 32–34].

With the advent of high-throughput sequencing technologies, it has become possible to rapidly analyze chloroplast genomes, providing a powerful approach for comparative genomic studies among diverse taxa [35–38]. However, integrated investigations encompassing codon usage evolution, SSR polymorphisms, selective pressure variation, and phylogenomic resolution within Polygonaceae remain limited. This gap hinders our understanding of both evolutionary dynamics and the mechanistic basis of bioactive metabolite production, which is critical for phytopharmacological applications.

In the present study, we performed a comprehensive analysis of chloroplast genome features across 48 Polygonaceae species, representing three major subfamilies and diverse ecological adaptations. Our specific objectives were to (i) characterize synonymous codon usage patterns and identify the underlying evolutionary determinants; (ii) assess SSR distributions as potential molecular markers for genetic diversity and breeding applications; (iii) quantify selective pressures across functional gene categories relevant to metabolism and secondary compound biosynthesis; (iv) reconstruct phylogenetic relationships using complete chloroplast genome sequences; and (v) identify genomic signatures of adaptive evolution that may underlie ecological success and bioactive metabolite diversity. By addressing these objectives, this study aims to provide fundamental insights into chloroplast genome evolution in Polygonaceae while generating practical resources for molecular systematics, biodiversity conservation, and biotechnological and pharmacological applications. Understanding the genetic and evolutionary basis of metabolite-rich species may inform the development of standardized herbal preparations, optimize bioactive compound yields, and support their integration into evidence-based phytotherapy.

Materials and methods

Plant materials and chloroplast genome sequencing

Fresh leaf tissues were collected from seven Polygonaceae species, namely, *Oxyria digyna*, *Persicaria chinensis*, *Persicaria maculosa*, *Rumex hastatus*, *Polygonum ajanense*, *Persicaria lapathifolia*, and *Fagopyrum leptopodum*, in Muli Tibetan Autonomous County, Liangshan Yi Autonomous Prefecture, Sichuan Province, China. According to Article 344 of the Criminal Law of the People's Republic of China, those who violate national regulations by illegally logging or destroying precious trees or other plants under state key protection shall be subject to legal sanctions, including fixed-term imprisonment, detention, surveillance, and fines. However, the plant specimens collected in this study are not included in the National Plant Protection Regulations. Voucher specimens were deposited at the School of Grassland Science, Beijing Forestry University. The specimen accession numbers are: *Oxyria digyna* (BJFU-CXY-S02-001R15), *Persicaria chinensis* (BJFU-CXY-S02-001R16), *Persicaria maculosa* (BJFU-CXY-S02-001R17), *Rumex hastatus* (BJFU-CXY-S02-001R18), *Polygonum ajanense* (BJFU-CXY-S02-001R19), *Persicaria lapathifolia* (BJFU-CXY-S02-001R20), and *Fagopyrum leptopodum* (BJFU-CXY-S02-001R21). Species identification was confirmed through morphological examination using standard taxonomic keys (https://www.iplant.cn/frps/vol). Total genomic DNA was extracted from silica gel-dried leaf tissue using a modified cetyltrimethylammonium bromide (CTAB) protocol optimized for high-altitude plant materials [39]. Paired-end libraries (2 × 150 bp) were constructed and sequenced on an Illumina HiSeq platform. Raw sequencing reads were assessed for quality using FastQC v0.11.9 [40] and filtered using Trimmomatic v0.39 [41] with default parameters to remove low-quality reads and adapter sequences. Chloroplast genomes were assembled using the reference-guided seed-and-extend algorithm implemented in NOVOPlasty v4.3.1 [42].

Genome annotation and sequence dataset compilation

Protein-coding genes (PCGs) were extracted from the assembled chloroplast genomes using the web-based organellar genome annotation platforms GeSeq (https://chlorobox.mpimp-golm.mpg.de/geseq.html) [43] and CPGAVAS2 (http://47.96.249.172:16019/analyzer/home) [44]. GeSeq integrates BLAT-based homology searches with profile hidden Markov model (HMM) analyses and utilizes a manually curated chloroplast reference database. Multiple prediction approaches were applied, including BLATn and BLATx searches, profile HMM-based identification of protein-coding and rRNA genes, and de novo tRNA prediction using tRNAscan-SE 2.0 and ARAGORN [45, 46]. IR regions were identified using GeSeq's self-BLATn search algorithm. All annotations were manually verified to ensure accuracy, and sequences shorter than 300 bp were excluded from downstream analyses to maintain statistical reliability [18]. To construct a comprehensive dataset, an additional 41 Polygonaceae chloroplast genomes were retrieved from the NCBI GenBank database (https://www.ncbi.nlm.nih.gov/) [47] (Table 1). The final dataset included 48 species spanning three major subfamilies: Polygonoideae, Eriogonoideae, and Coccoloboideae.

Phylogenetic analysis

For phylogenetic reconstruction, orthologous chloroplast protein-coding genes were extracted from all plastomes, and gene sets were filtered to minimize missing data while retaining broad taxonomic coverage across taxa. Each gene was aligned with MAFFT v7 using the automatic alignment strategy [48]. Poorly aligned and ambiguous regions were removed using trimAl with the -automated1 heuristic [49]. Gene alignments passing quality checks were concatenated into a supermatrix, and maximum-likelihood analysis was performed in IQ-TREE 2 [50]. The most appropriate nucleotide substitution model was selected using ModelFinder based on the Bayesian Information Criterion (BIC) [51, 52], and branch support was assessed using the ultrafast bootstrap approximation UFBoot2 together with the SH-aLRT test [53, 54]. Trees were visualized and annotated in R using the ape and ggtree packages [55, 56]. Ancestral state reconstruction of continuous plastome traits was performed using phytools with maximum-likelihood estimation under a Brownian motion model [57].

Ancestral state reconstruction

Ancestral state reconstruction for continuous chloroplast genome traits was performed using the phytools v2.0 package in R v4.3.0 [58]. The contMap function was employed, which implements maximum likelihood estimation of ancestral states under a Brownian motion model of character evolution. Ancestral values at internal nodes were estimated using the fastAnc function, which employs a rerooting algorithm for efficient computation. The reconstructed values were mapped onto phylogenetic branches using a continuous color gradient, enabling visualization of evolutionary trajectories across the tree.

Ancestral state reconstructions were visualized in three layouts: circular, rectangular, and fan formats. For each of the eight chloroplast genome characteristics, separate contMap visualizations were generated with consistent color scales to facilitate comparative interpretation. The contMap objects were plotted using the type parameter "fan" for circular/fan layouts and "phylogram" for rectangular layouts, with tip labels disabled to improve visual clarity for large datasets.

Table 1 Detailed characteristics of the complete chloroplast genomes of 48 Polygonaceae species

GenBank accession	Species Name	Gene num	Genome size (bp)	GC (%)	GC1%	GC2%	GC3%	GC3s
KX085498	*Fagopyrum tataricum*	131	159272	37.88	45.57	37.74	30.91	27.96
KX774248	*Oxyria sinensis*	128	160404	37.54	45.69	37.77	30.71	27.7
MH359405	*Rumex acetosa*	127	160269	37.2	45.75	37.78	30.45	27.47
MK381448	*Reynoutria japonica*	130	163386	37.52	45.59	37.93	31.07	28.11
MK842154	*Reynoutria sachalinensis*	130	163485	37.53	45.73	37.84	31.09	28.12
MN055629	*Rumex crispus*	129	158851	37.55	45.7	37.79	30.61	27.64
MN202600	*Calligonum caput-medusae*	131	162019	37.51	45.83	37.81	30.64	27.63
MN202602	*Calligonum colubrinum*	130	161321	37.53	45.81	37.8	30.63	27.61
MN202603	*Calligonum cordatum*	131	161276	37.55	45.81	37.81	30.65	27.64
MN720269	*Rumex japonicus*	130	159292	37.5	45.7	37.79	30.6	27.62
MN880789	*Rheum nobile*	125	161250	37.29	45.59	37.63	30.71	27.75
MT341768	*Rheum alexandrae*	128	161634	37.33	45.59	37.71	30.65	27.67
MT572345	*Fagopyrum esculentum*	131	159576	37.97	45.63	37.79	31.02	28.07
MW770449	*Bistorta coriacea*	126	158839	37.99	45.85	37.84	31.06	28.1
MZ573789	*Koenigia divaricata*	131	159126	37.43	45.55	37.63	30.46	27.46
MZ573795	*Koenigia nepalensis*	130	155838	37.37	45.38	37.6	29.99	27.01
MZ579703	*Fallopia multiflora*	130	163425	37.54	45.68	37.74	31.09	28.07
MZ618350	*Pteroxygonum denticulatum*	130	162928	37.38	45.61	37.67	30.62	27.59
MZ618352	*Fallopia aubertii*	130	162393	37.59	45.75	37.69	31.37	28.41
MZ748474	*Polygonum aviculare*	129	179064	37.39	45.97	38.17	31	28.05
MZ748475	*Bistorta officinalis*	130	159476	37.88	45.77	37.79	31.1	28.16
MZ958816	*Rheum tanguticum*	129	161039	37.35	45.57	37.71	30.67	27.71
MZ958817	*Rheum officinale*	129	161093	37.33	45.57	37.68	30.67	27.7
OK054490	*Fagopyrum urophyllum*	139	159350	37.78	45.55	37.8	30.72	27.76
OK661151	*Coccoloba uvifera*	127	169360	36.59	45.69	37.57	30.15	27.15
OK661152	*Triplaris americana*	127	171340	36.9	46.45	38.31	29.71	26.71
OK661157	*Koenigia forrestii*	127	156844	37.29	45.48	37.62	30.12	27.15
ON229543	*Atraphaxis spinosa*	130	164070	37.46	45.73	37.78	31.07	28.09
ON229544	*Bistorta amplexicaulis*	130	159789	37.85	45.75	37.73	31.06	28.12
ON229550	*Polygonum suffultum*	130	159115	37.98	45.79	37.8	31.14	28.19
ON229551	*Bistorta vivipara*	130	158825	37.99	45.73	37.75	31.09	28.14
ON229553	*Calligonum leucocladum*	131	161313	37.53	45.79	37.76	30.61	27.6
ON229569	*Persicaria capitata*	130	158852	37.96	45.72	37.86	30.89	27.91
ON229587	*Persicaria nepalensis*	130	157889	37.93	45.59	37.78	30.76	27.75
ON229593	*Persicaria senticosa*	131	160589	38.02	45.78	37.9	30.97	27.98
ON229608	*Polygonum urumqiense*	130	163376	37.46	45.73	37.79	30.81	27.85
ON229609	*Pteroxygonum giraldii*	130	162009	37.32	45.52	37.64	30.43	27.42
OR570615	*Persicaria hydropiper*	128	159819	38.19	45.86	37.82	31.41	28.43
OR730799	*Persicaria tinctoria*	126	159633	38.18	46.95	38.97	31.59	28.67
OR866441	*Knorringia sibirica*	129	161384	37.63	45.84	37.82	31.01	28.06
PQ523295	*Atraphaxis pungens*	133	164065	37.46	45.63	37.82	31.1	28.12
PX501750	*Persicaria maculosa*	130	159073	38.22	45.9	37.78	31.42	28.45
PX501751	*Rumex hastatus*	130	158985	37.4	45.69	37.92	30.82	27.83
PX501752	*Polygonum ajanense*	125	177536	38.16	45.59	37.78	30.47	27.47
PX501753	*Persicaria lapathifolia*	129	159072	38.22	45.91	37.8	31.42	28.45
PX501754	*Fagopyrum leptopodum*	129	159362	37.82	45.62	37.82	30.72	27.77
PX501763	*Oxyria digyna*	131	160666	37.54	45.71	37.87	30.96	27.96
PX501764	*Persicaria chinensis*	130	159073	37.97	45.83	37.9	31.24	28.27

Statistical analysis

Group comparisons of chloroplast genome characteristics among genera were performed using nonparametric statistical methods appropriate for datasets that did not meet normality assumptions. The Shapiro-Wilk test was applied to assess the normality of trait distributions within each genus. Levene's test was used to evaluate the homogeneity of variances across groups. For

comparisons involving more than two groups (≥3 genera), the Kruskal-Wallis rank–sum test was employed, with Dunn's post hoc test applying Bonferroni correction for multiple pairwise comparisons. For two-group comparisons, the Wilcoxon rank-sum test (Mann–Whitney U test) was utilized.

Effect sizes were calculated to quantify the magnitude of between-group differences. For multigroup comparisons, eta-squared (η^2) was computed as $\eta^2 = H/(n-1)$, where H is the Kruskal–Wallis test statistic and n is the total sample size. The effect size interpretation followed established conventions: small (0.01–0.05), medium (0.06–0.13), and large (≥0.14). For two-group comparisons, Cohen's d was calculated. The significance thresholds were set at $P<0.05$ (*), $P<0.01$ (**), and $P<0.001$ (***), with $P\geq0.05$ considered not significant (NS).

Pearson correlation coefficients were calculated to assess pairwise relationships among the eight chloroplast genome characteristics. Correlation matrices were visualized as heatmaps with hierarchical clustering to identify groups of covarying traits. All the statistical analyses and visualizations were performed using R v4.3.0 with the packages ggplot2 v3.4.2, ggpubr, rstatix, corrplot, patchwork, and ggbeeswarm.

Collinearity analysis

Genome structural conservation among the 48 chloroplast genomes was assessed through comprehensive collinearity analyses. Pairwise genome comparisons were performed using average nucleotide identity (ANI) analysis implemented in pyANI v0.2.11 with MUMmer v3.23 alignments [59]. Synteny detection was conducted using BLASTN searches with stringent parameters (e-value threshold of 1 × 10 − 5, minimum alignment length of 50 bp) to identify homologous genomic regions [60]. Syntenic blocks were visualized in a phylogenetic context using genoPlotR v0.8.11 in R v4.3.0, incorporating hierarchical clustering basis on ANI distance matrices [61, 62]. This approach enables the detection of genome rearrangements, inversions, and conserved structural features, providing insights into evolutionary stability.

Nucleotide composition and codon usage analyses

Base composition analyses were performed to quantify overall adenine-thymine (AT) and guanine-cytosine (GC) contents. Positional GC content was further assessed at the first, second, and third codon positions, with a particular focus on the GC content at synonymous third codon positions (GC3s) to evaluate compositional bias and the relative contributions of mutational pressure versus natural selection [63]. Genome structural features, including the LSC, SSC, and inverted repeat (IR) regions, were compared across all species to identify conserved organizational patterns.

RSCU values were calculated to quantify codon usage bias (CUB), defined as the ratio of the observed codon frequency to the expected frequency under equal usage of synonymous codons for a given amino acid [64]. RSCU values > 1.0 indicate preferentially used codons, whereas values < 1.0 denote underrepresented codons. All calculations were performed using Python v3.12.2 with custom scripts.

SSR analysis

Chloroplast SSRs, consisting of tandemly repeated motifs of 1–6 nucleotides, were identified using MISA-web. The search parameters were configured to detect mononucleotide (≥ 10 repeats), dinucleotide (≥ 5 repeats), trinucleotide (≥ 4 repeats), tetranucleotide (≥ 3 repeats), and penta- and hexanucleotide repeats (≥ 2 repeats each) [65]. SSR distributions were characterized across genomic regions, including coding sequences (PCGs), introns, intergenic spacers (IGSs), tRNA genes, and rRNA genes, to assess regional variation in repeat abundance.

ENC analysis

ENC analysis was conducted to quantify codon usage bias by comparing the observed ENC values to those expected under neutral mutation models [66]. The ENC values range from 20 (extremely biased codon usage) to 61 (no bias). ENC plots were constructed by plotting ENC values against the GC3 content, with the theoretical neutral curve calculated using the formula $ENC = 2 + GC3s + 29/[GC3s^2 + (1 - GC3s)^2]$ [63]. Genes falling significantly below the theoretical curve indicate that codon usage is shaped predominantly by natural selection rather than mutational pressure [67].

PR2 analysis

PR2 analysis was performed to assess base composition symmetry at third codon positions by plotting A3/(A3 + T3) against G3/(G3 + C3), with the central point (0.5, 0.5) representing complete parity and no strand bias [68]. Deviations from this central position indicate compositional asymmetry arising from mutational pressure, natural selection, or their interaction [69]. Gene distributions across the four quadrants of the plot were quantified to characterize directional biases in nucleotide usage patterns [70].

Nucleotide diversity and haplotype analysis

Nucleotide diversity (π) was calculated for each protein-coding gene across the 48 Polygonaceae species following the methods of Nei and Li. For each gene, multiple sequence alignments were generated using MAFFT v7.490 (-autoparameter) [48]. Poorly aligned regions were subsequently removed using trimAl v1.4 with the -automated1 option [49]. Nucleotide diversity (π), which

represents the average number of nucleotide differences per site between all pairwise sequences, was calculated as follows: $\pi = \Sigma_{i<j} d_{ij}/[n(n-1)/2]$, where d_{ij} is the proportion of nucleotide differences between sequences i and j, and n is the number of sequences. Standard errors were estimated using 1,000 bootstrap replicates. Watterson's theta (θ) was calculated on the basis of the number of segregating sites, and Tajima's D statistic was computed to test for departures from neutral evolution. Haplotype diversity (Hd), the number of segregating sites, transitions, and transversions, and the transition-to-transversion ratio (Ti/Tv) were also calculated for each gene. The genes were classified into standard chloroplast functional categories. Nucleotide diversity across categories was compared using Kruskal–Wallis tests, with post hoc pairwise comparisons performed using Dunn's test with Bonferroni correction.

Selection pressure analysis

Selective pressures acting on chloroplast protein-coding genes were evaluated by calculating nonsynonymous (Ka) and synonymous (Ks) substitution rates for pairwise comparisons among the 48 species [71]. PCGs longer than 300 bp were extracted and aligned using MAFFT v7.490 with default parameters optimized for nucleotide sequences [48]. The alignments were mgenomes exhibited the characteristic quadripartite architectureanually inspected and refined in MEGA X v10.2.6 to ensure proper codon alignment [72]. Ka and Ks values were computed using KaKs_Calculator 2.0 using the Nei–Gojobori method, which provides reliable estimates for sequences with mogenomes exhibited the characteristic quadripartite architectureregenomes exhibited the characteristic quadripartite architectureregenomes exhibited the characteristic quadripartite architecture

genomes exhibited the characteristic quadripartite architecturederate divergence typical of intrafamily comparisons [73]. Gene pairs with Ks values > 2.0 were excluded to minimize saturation effects. Selective pressure categories were defined as follows: strong purifying selection: Ka/Ks < 0.5; relaxed purifying selection: 0.5 ≤ Ka/Ks ≤ 1.0; and positive selection: Ka/Ks > 1.0. These cgenomes exhibited the characteristic quadripartite architecture

lassifications follow established standards in plant evolutionary genomics [74].

Results

Chloroplast genome architecture and nucleotide composition of seven newly sequenced Polygonaceae species

The complete chloroplast genomes of seven Polygonaceae species, namely, *O. digyna*, *P. chinensis*, *P. maculosa*, *R. hastatus*, *P. ajanense*, *P. lapathifolia*, and *F. leptopodum*, ranged from 158,985 bp to 177,536 bp in length, with overall GC contents varying between 37.40% and 38.22% (Fig. 1A, Table S1). All the genomes exhibited the characteristic quadripartite architecture typical of angiosperm chloroplast genomes, comprising a large single-copy (LSC) region, a small single-copy (SSC) region, and a pair of inverted repeats (IRs).

Nucleotide composition analyses revealed a consistent adenine-thymine (AT) bias, with adenine comprising 30.76–31.10%, thymine comprising 30.97–31.49%, cytosine comprising 19.02–19.43%, and guanine comprising 18.39–18.82% of the total genomic content. The GC content was consistently greater in coding regions than in noncoding sequences, reflecting evolutionary constraints on protein-coding genes and patterns widely observed across Polygonaceae.

The gene content ranged from 130 to 133 genes per genome, including 85–88 protein-coding genes, 37 tRNA genes, and 8–9 rRNA genes. Protein-coding genes encompassed functional categories critical to chloroplast metabolism and photosynthesis, including components of photosystems I and II, the cytochrome b_6/f complex, ATP synthase, NADH dehydrogenase, ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit (*rbcL*), RNA polymerase, and ribosomal proteins. The coding regions spanned 82,459–85,402 bp, with a GC content of 37.93–38.41%, whereas the noncoding regions extended 61,844–83,161 bp and presented a reduced GC content of 32.48–34.88%.

Codon usage analyses revealed pronounced positional bias, with the GC content decreasing across codon positions in the order GC1 > GC2 > GC3, which is consistent with patterns shaped by both mutational constraints and natural selection. These structural and compositional features provide a foundation for understanding gene expression dynamics, adaptive evolution, and the biosynthetic potential of metabolites with pharmacological relevance in Polygonaceae species.

Phylogenetic analysis

The concatenated chloroplast genomes protein-coding supermatrix produced a well-resolved maximum-likelihood topology with strong support across most internal nodes, indicating that chloroplast genome coding data contain sufficient signal to resolve both deep and shallow relationships within Polygonaceae (Fig. 2, Table S2). The Polygonaceae family formed a strongly supported monophyletic clade (UFBoot = 100%, SH-aLRT = 100%) clearly separated from the Plumbaginaceae outgroup, confirming the phylogenetic integrity of the family. Within Polygonaceae, several major evolutionary lineages were identified with high support. The subfamily Polygonoideae was resolved as monophyletic, comprising the majority of the sampled taxa, including *Fagopyrum*,

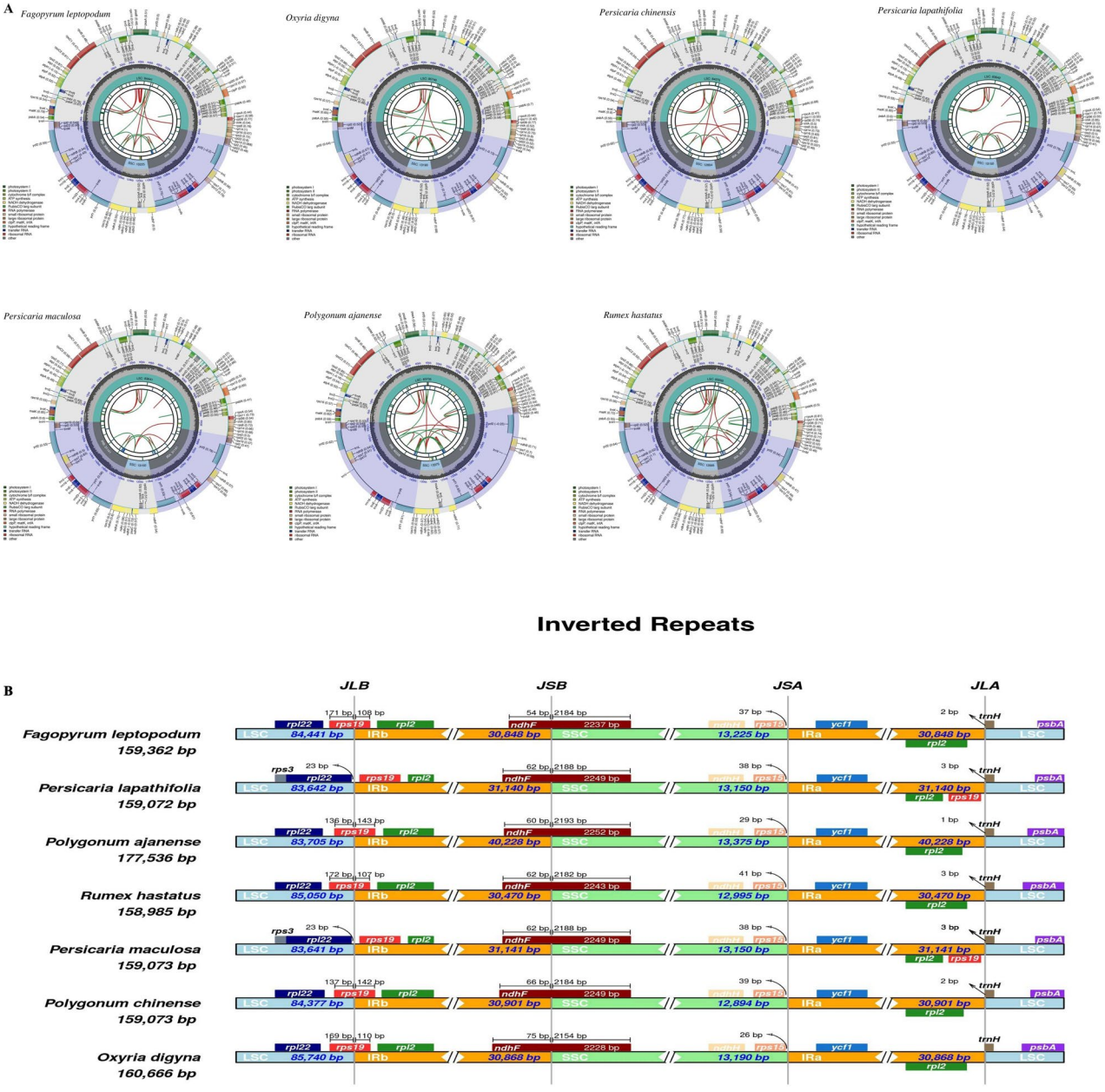

Fig. 1 Chloroplast genome features of seven Polygonaceae species. Species names are indicated in the top left corner. From the center outward: Dispersed repeats: direct (D) and palindromic (P) repeats are connected by red and green arcs, respectively. Long tandem repeats: represented as short blue bars. Short tandem repeats (microsatellites): depicted as short bars in different colors. Genome structure: small single-copy (SSC), inverted repeat (IRa and IRb), and large single-copy (LSC) regions. GC content: plotted along the genome; the base frequency at each site is displayed between tracks 4 and 5. Genes: color-coded by functional classification, with transcription directions indicated clockwise for inner genes and anticlockwise for outer genes. Optional codon usage bias (CUB) is provided in parentheses following gene names. The functional classification legend is shown in the bottom left corner

Polygonum, *Persicaria*, *Bistorta*, *Koenigia*, *Rheum*, *Rumex*, *Oxyria*, *Calligonum*, *Atraphaxis*, *Fallopia*, *Reynoutria*, *Pteroxygonum*, *Knorringia*, and *Muehlenbeckia*. The tropical genera *Coccoloba* and *Triplaris* (subfamily Coccoloboideae) formed a distinct clade sister to the remaining Polygonoideae species.

The genus-level relationships were generally well supported. The buckwheat genus *Fagopyrum* (11 species) formed a monophyletic clade with strong support (UFBoot ≥ 99%), which is consistent with previous molecular phylogenetic studies. *Calligonum* (28 species), the largest sampled genus, was recovered as monophyletic, with species clustering according to previously recognized sections. The medicinal genera *Rheum* (18 species) and *Rumex* (7 species) were each resolved as monophyletic with high support, although their

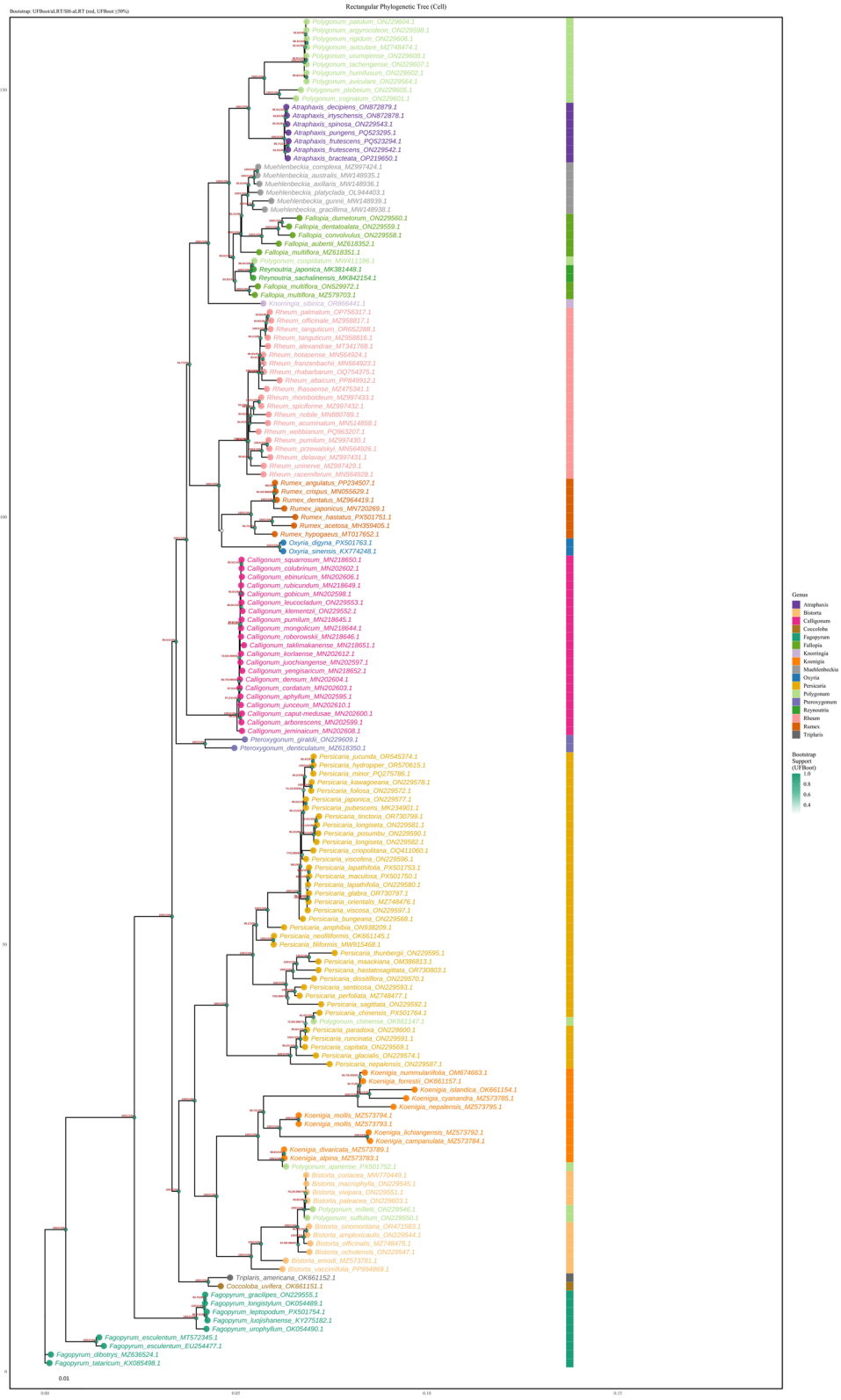

Fig. 2 Maximum likelihood phylogenetic reconstruction of 48 Polygonaceae species on the basis of chloroplast protein-coding genes (PCGs). The rectangular tree displays genus-level classification, with colored circles at branch tips indicating genus membership. Node circles represent bootstrap support values, with color intensity reflecting statistical confidence. Branch lengths correspond to evolutionary distances. The tree topology resolves major evolutionary lineages, including Polygonoideae, Eriogonoideae, and Coccoloboideae, and largely conforms to morphology-based taxonomy, providing a robust framework for investigating diversification, ecological adaptation, and biogeographic history within the Polygonaceae family

intergeneric relationships presented moderate bootstrap values (70–85%). The genus *Persicaria* (22 species) was recovered as paraphyletic with respect to *Bistorta* (9 species), with *Bistorta* nested within a larger *Persicaria* + *Koenigia* clade. This finding aligns with recent taxonomic revisions that have questioned the monophyly of traditional *Persicaria* delimitation. *Koenigia* (13 species) formed a distinct clade sister to a subset of *Persicaria* species, supporting its recognition as a separate genus. The knotweed clade comprising *Fallopia*, *Reynoutria*, and *Pteroxygonum* presented complex relationships, with *Reynoutria japonica* and *R. sachalinensis* forming a well-supported clade distinct from *Fallopia* species. *Polygonum* sensu stricto was resolved as monophyletic, although several species traditionally placed in this genus were dispersed across other clades, reflecting ongoing taxonomic realignment within the family.

Ancestral state reconstruction of chloroplast genome characteristics

Ancestral state reconstruction using maximum likelihood under a Brownian motion model revealed distinct evolutionary trajectories for each of the eight chloroplast genome characteristics (Fig. 3, Table S3). Mapping of reconstructed ancestral values onto the phylogeny enabled visualization of trait evolution across the diversification history of Polygonaceae. Evolutionary rate: Branch-specific evolutionary rates showed considerable variation across lineages, with the ancestral Polygonaceae state reconstructed as moderate. Accelerated evolution was observed in the *Calligonum* clade, potentially reflecting adaptation to arid environments. In contrast, the *Fagopyrum* and *Rheum* lineages presented relatively conserved evolutionary rates. GC content: overall GC content across the family ranged from 35% to 45%, with the ancestral state reconstructed near the family median. The *Calligonum* lineage presented notably higher GC contents, whereas *Rheum* and *Oxyria* presented lower values. The GC content appeared relatively stable across most lineages, suggesting conserved selective constraints on nucleotide composition. Genome size: chloroplast genome sizes ranged from approximately 140 kb to 200 kb, with substantial variation observed across genera. The ancestral genome size was reconstructed as approximately 160 kb. Notable genome expansion has been observed in *Calligonum* and some *Persicaria* species, whereas genome contraction has occurred in Rheum. These size variations correlated with changes in the length of the IR region. Gene density and the coding ratio: Gene density and the coding ratio showed inverse relationships with genome size, as expected, given the relatively conserved gene content across the family. The ancestral gene density was reconstructed as approximately 0.8 genes/kb. Lineages with expanded genomes (*Calligonum*) presented lower gene density, whereas compact genomes (*Rheum*) presented higher density. The coding ratios varied from approximately 40% to 80%, with most species clustering at approximately 55–65%. tRNA genes and IR length: the number of tRNA genes whose ancestral state was reconstructed near 37 tRNA genes showed limited variation. IR region length exhibited substantial variation, representing a major contributor to genome size polymorphism. The *Calligonum* lineage presented notably expanded IR regions, whereas *Rheum* presented contracted IRs, which is consistent with previous reports of IR boundary dynamics in these genera. AT skew: Strand compositional asymmetry ranged from approximately − 0.02–0.08, with most species showing slight positive skew. The ancestral state was reconstructed near neutrality. *Polygonum* and *Rheum* presented elevated positive AT skew, whereas Persicaria species presented values closer to zero, potentially reflecting differences in replication-associated mutational biases.

Statistical analysis of chloroplast genome variations: intergenerational differences and trait correlations

Comparative statistical analysis of chloroplast genome characteristics across 16 Polygonaceae genera revealed significant intergeneric variation for most traits examined (Fig. 4, Table S4). Kruskal–Wallis tests detected significant differences among genera for all eight characteristics, with varying effect sizes. The evolutionary rate showed highly significant variation among genera, representing the largest effect size observed. Post hoc Dunn's tests with Bonferroni correction revealed that *Calligonum* presented significantly elevated evolutionary rates compared with *Fagopyrum* ($P < 0.001$), *Rheum* ($P < 0.001$), and *Persicaria* ($P < 0.001$). Atraphaxis also presented elevated rates compared with those of the family median, which is consistent with the ability of Calligonum to adapt to arid environments. The GC content differed significantly among genera. Compared with most other genera, Calligonum presented significantly greater GC contents, whereas Rheum and Rumex presented lower values characteristic of temperate-adapted species. Genome size variation was highly significant, with Calligonum showing significantly larger genomes than *Rheum* ($P < 0.001$), *Fagopyrum* ($P < 0.001$), and *Rumex* ($P < 0.001$). The genomes of the tropical genus Triplaris were also larger than those of the temperate genera. Gene density showed moderate but significant variation. *Rheum* presented the highest gene density, whereas *Calligonum* presented the lowest density. Atraphaxis presented intermediate values despite the close phylogenetic relationship with *Calligonum*. The coding ratio varied significantly among genera. *Koenigia* and *Oxyria* presented the highest coding ratios, whereas *Calligonum*

Fig. 3 Ancestral state reconstruction displayed in a rectangular phylogram layout. The same ancestral state reconstruction analysis as in Fig. 3 is displayed in rectangular phylogram format to facilitate interpretation of evolutionary trajectories along individual branches. The color gradients (blue to red) indicate trait values as described in Fig. 3. This layout enables visual comparison of branch lengths (evolutionary distance) with trait values, revealing that lineages with elevated evolutionary rates (longer branches) tend to show correlated changes in other chloroplast genome characteristics. The panel arrangement and trait designations (**A**–**H**) are identical to those in Fig. 3. Species names and tip trait values are shown at branch termini. The internal node values represent the maximum likelihood ancestral state estimates under the Brownian motion model

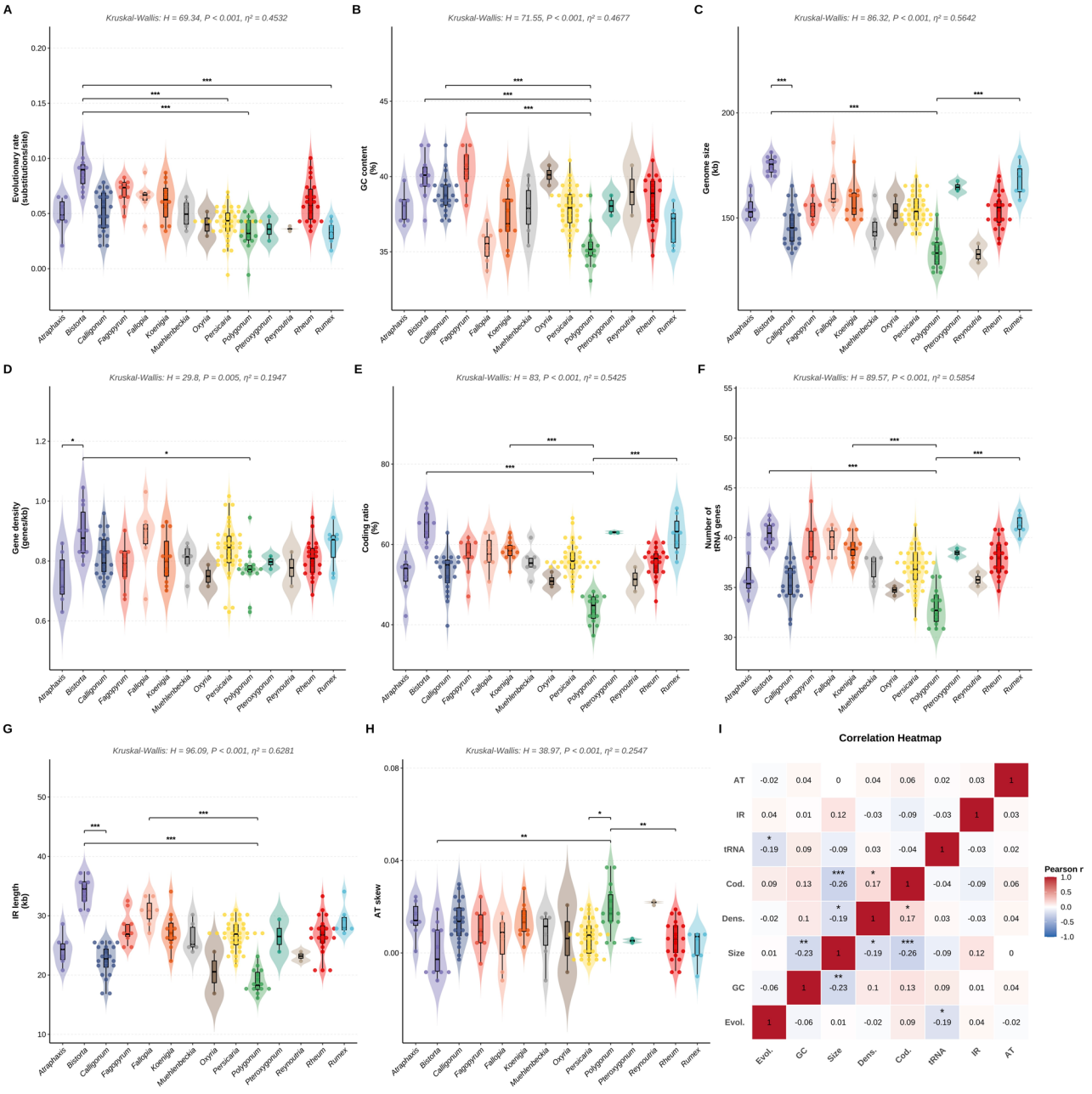

Fig. 4 Statistical comparison of chloroplast genome characteristics among Polygonaceae genera and correlation analysis. (A–H) Violin plots with overlaid box plots and individual data points (beeswarm) showing the distribution of eight chloroplast genome characteristics across 16 Polygonaceae genera (genera with ≥ 2 species). The violin width represents the kernel density estimation of the data distribution. Box plot elements: box boundaries = interquartile range (IQR), horizontal line = median, whiskers = 1.5 × IQR. Kruskal–Wallis test statistics are displayed for each trait, including the *H* statistic, *P* value, and effect size (η^2, eta-square). Effect size interpretation: small (0.01–0.06), medium (0.06–0.14), and large (> 0.14). Letters above violins indicate Dunn's post hoc test groupings with Bonferroni correction; genera sharing the same letter are not significantly different ($P > 0.05$). Trait panels: (**A**) Evolutionary rate ($H = 69.34$, $P < 0.001$, $\eta^2 = 0.4532$); (**B**) GC content ($H = 71.55$, $P < 0.001$, $\eta^2 = 0.4677$); (**C**) genome size ($H = 86.32$, $P < 0.001$, $\eta^2 = 0.5642$); (**D**) gene density ($H = 29.80$, $P = 0.005$, $\eta^2 = 0.1947$); (**E**) coding ratio ($H = 83.00$, $P < 0.001$, $\eta^2 = 0.5425$); (**F**) tRNA number ($H = 89.57$, $P < 0.001$, $\eta^2 = 0.5854$); (**G**) IR length ($H = 96.09$, $P < 0.001$, $\eta^2 = 0.6281$); (**H**) AT skew ($H = 38.97$, $P < 0.001$, $\eta^2 = 0.2547$). **I** Pearson correlation matrix heatmap of eight chloroplast genome characteristics with hierarchical clustering (complete linkage method). Color scale: red = positive correlations, blue = negative correlations; color intensity reflects correlation strength. Significance levels: *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$. Strong positive correlations included genome size–IR length ($r = 0.89$) and genome size–tRNA number ($r = 0.82$). Strong negative correlations included the following: genome size–gene density ($r = -0.76$) and the genome size–coding ratio ($r = -0.71$). Sample size: $n = 155$ Polygonaceae species. Visualization was performed using ggplot2 v3.4 and corrplot in R

and *Muehlenbeckia* presented lower values, reflecting differences in noncoding sequence accumulation. tRNA gene number had the largest effect size among all the traits. Compared with all the other genera, *Calligonum* contained significantly more tRNA genes, likely reflecting the fact that IR expansion has resulted in the duplication of tRNA gene copies. *Persicaria*, *Koenigia*, and *Rheum* presented intermediate values. IR length exhibited highly significant variation, which is consistent with its role as a major driver of genome size polymorphism. *The* IRs of Calligonum were significantly greater than those of *Rheum*, *Fagopyrum*, and *Persicaria*. The AT skew differed significantly among genera, with *Rheum* showing elevated positive skew values compared with *Persicaria* ($P < 0.01$) and Koenigia ($P < 0.05$). This pattern suggests lineage-specific differences in replication or transcription-coupled mutational biases.

Pearson correlation analysis revealed significant relationships among several chloroplast genome characteristics. Genome size was strongly positively correlated with IR length and tRNA gene number, confirming that IR expansion is the primary driver of genome size variation in Polygonaceae. Conversely, genome size was negatively correlated with gene density and the coding ratio, reflecting the dilution of coding sequences within expanded genomes. The evolutionary rate showed a weak positive correlation with the GC content, suggesting that lineages with elevated nucleotide substitution rates tend toward higher GC compositions. This relationship may reflect codon usage bias optimization or environmental adaptation. The AT skew showed a weak negative correlation with the coding ratio, potentially indicating that genomes with higher noncoding content experience different replication-associated mutational pressures. Hierarchical clustering of the correlation matrix identified two primary groups of covarying traits: (1) a size-associated cluster comprising genome size, IR length, and tRNA number versus gene density and coding ratio and (2) a composition-associated cluster linking GC content, evolutionary rate, and AT skew. These groupings reflect the mechanistic relationships among chloroplast genome characteristics, with structural features forming one functional module and compositional features forming another.

Collinearity analysis

Collinearity analysis revealed high sequence similarity among closely related genera, with clearly defined genus-level clustering patterns in comparative genomic alignments. Within *Calligonum*, chloroplast genome structures were nearly identical among *C. caput-medusae*, *C. cordatum*, *C. colubrinum*, and *C. leucocladum*, with sequence identities exceeding 98% across entire genome lengths, which is consistent with highly conserved chloroplast genome organization despite their broad distribution across arid environments (Fig. 5, Table S5). Similarly, *Fagopyrum* species (*F. leptopodum*, *F. urophyllum*, *F. esculentum*, and *F. tataricum*) presented marked structural homogeneity, with minimal variation in gene arrangement and intergenic spacing, reflecting recent evolutionary divergence within this economically and nutritionally significant genus. In the *Rheum* cluster (*R. nobile*, *R. alexandrae*, *R. officinale*, and *R. tanguticum*), the chloroplast genomes presented exceptional collinearity with minimal structural divergence despite distinct morphological and ecological adaptations, suggesting that strong purifying selection maintained the chloroplast genome architecture. Notably, *R. officinale* and *R. tanguticum* share the most similar genome structure, which is consistent with their close phylogenetic affinity.

Interspecific comparisons within *Oxyria* (*O. digyna*, *O. sinensis*) demonstrated near-perfect collinearity despite geographic isolation and ecological differentiation, with identical gene orders and similar IR configurations, indicating recent divergence and limited chloroplast genome evolution in this alpine-adapted lineage. Across all comparisons, coding regions were significantly more conserved than noncoding intergenic spacers were, highlighting functional constraints on protein-coding genes and relaxed selective pressures on nonfunctional regions. These findings demonstrate that the chloroplast genome structure in Polygonaceae is highly conserved at both the genus and subfamily levels, reflecting evolutionary constraints that likely preserve essential photosynthetic, metabolic, and regulatory functions, with implications for genetic engineering, molecular breeding, and the production of pharmacologically relevant secondary metabolites.

Comparative genome characteristics across 48 Polygonaceae species

The chloroplast genomes of the 48 Polygonaceae species analyzed presented a highly conserved quadripartite structure, with total lengths ranging from 155,838 bp to 179,064 bp. The overall GC content was relatively uniform (36.59–38.22%), which is consistent with previously reported Polygonaceae chloroplast genome compositions. Nucleotide composition analyses confirmed a pronounced AT bias, with average adenine (31.05%) and thymine (31.32%) contents exceeding cytosine (19.13%) and guanine (18.50%) contents, reflecting patterns characteristic of dicotyledonous chloroplast genomes (Table S1).

Positional GC content analysis revealed a consistent hierarchy of GC1 > GC2 > GC3, indicating differential selective pressures across codon positions. The average

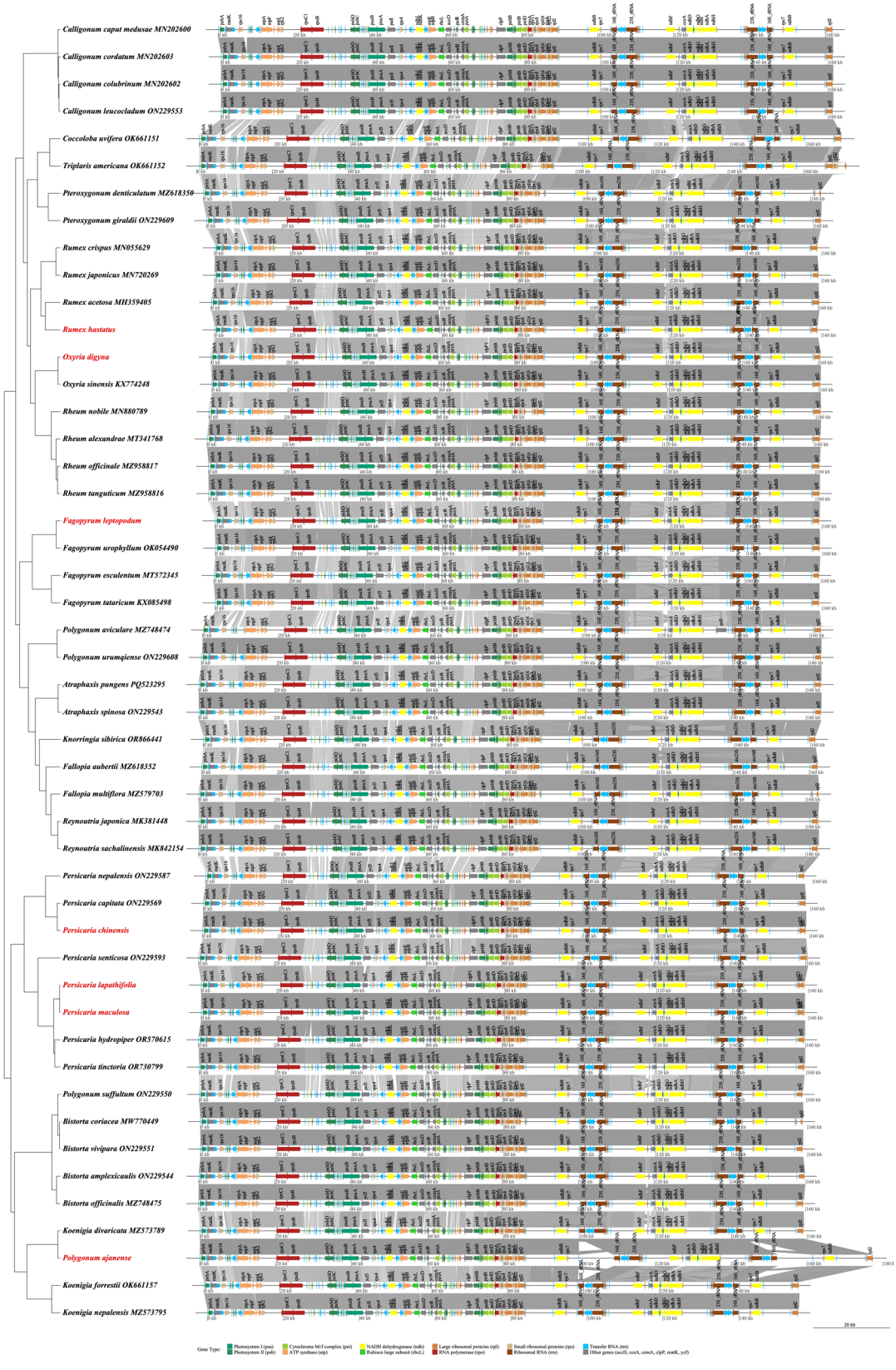

Fig. 5 Collinearity analysis of the chloroplast genomes of 48 Polygonaceae species. Each point represents a chloroplast gene, color-coded by gene type. The arrows indicate gene orientation and are colored according to functional categories, including photosystem components, ATP synthase, NADH dehydrogenase, ribosomal proteins, and other genes. Comparative alignment highlights conserved gene order and structural collinearity across genera, with minor variations in intergenic regions providing phylogenetically informative markers for resolving evolutionary relationships

GC3 content was 27.82%, suggesting substantial synonymous codon usage bias (CUB). Gene content was highly conserved among species, with coding regions averaging 82,332 bp and exhibiting GC content relative to noncoding regions (average 67,419 bp), which is consistent with functional constraints on protein-coding genes (PCGs). The gene repertoire included 81–88 protein-coding genes, 8–9 rRNA genes, and 35–45 tRNA genes per genome, yielding total gene counts of 125–139.

Codon usage analysis indicated relatively weak codon bias, with ENC values ranging from 46.36 to 48.28. Elevated ENC values, together with average RSCU values near 1.0, suggest that synonymous codon usage in Polygonaceae chloroplast genomes reflects a balance between mutational pressures and natural selection. These conserved structural and compositional features provide a foundation for understanding the evolutionary constraints on chloroplast genes and their potential influence on the biosynthesis of bioactive metabolites relevant to pharmacological applications.

Relative synonymous codon usage (RSCU) analysis

RSCU analysis across 48 Polygonaceae species representing the three subfamilies, namely, Polygonoideae, Eriogonoideae, and Coccoloboideae, revealed clear phylogenetic clustering corresponding to taxonomic classification (Fig. 6A, Table S6A). Hierarchical clustering based on RSCU values revealed that species within the same subfamily presented highly similar codon usage patterns, with Polygonoideae forming the most cohesive cluster, followed by distinct Eriogonoideae and Coccoloboideae groupings. Within subfamilies, further clustering reflected habitat adaptation: alpine-adapted species (e.g., *Bistorta* spp., *Koenigia* spp.) and arid-adapted taxa (e.g., *Calligonum* spp., *Atraphaxis* spp.) segregated into separate clusters, indicating a strong phylogenetic and ecological signal in CUB (Fig. 6B). The ENC values ranged from 45.8 to 48.3, which is consistent with the relatively weak CUB typical of plant chloroplast genomes. Nonetheless, habitat-specific differences were observed: alpine species generally presented slightly higher ENC values than arid-adapted species did, suggesting subtle variation in selective pressures linked to environmental adaptation.Among the subfamilies, Coccoloboideae species (*Coccoloba uvifera*,*Triplaris americana*) presented the highest ENC values, whereas several*Calligonum*species presented the lowest ENC values, reflecting subfamily specific differences in codon bias intensity. Overall, the RSCU patterns demonstrated a pronounced preference for A/U-ending codons, with a bimodal distribution peaking near 0.6 and 1.4, clearly distinguishing preferred from underrepresented codons (Fig. 6C). A total of 29 codons displayed positive bias, 28 of which terminated in A or U, underscoring a strong pyrimidine-rich codon preference. The ten most preferred codons were UUA, AGA, GCU, UAU, UCU, GAU, UUU, CAU, AAU, and CCU, whereas the ten least preferred codons predominantly terminated in G or C, with AGC and CGC exhibiting the strongest negative bias (Fig. 6D, Table S6D).

These results highlight that codon usage in Polygonaceae chloroplast genomes is shaped by a combination of phylogenetic inheritance and ecological adaptation, with potential implications for translational efficiency, gene expression optimization, and the biosynthesis of secondary metabolites relevant to pharmacological applications.

SSR analysis

Comprehensive SSR analysis across 48 Polygonaceae chloroplast genomes revealed ten distinct repeat categories ranging from mononucleotide to decanucleotide repeats (Fig. 7, Table S7A, Table S7D). Mononucleotide SSRs predominated, with 1,810 occurrences, which is consistent with patterns observed in angiosperm chloroplast genomes and is likely attributable to slipped-strand mispairing during DNA replication (Table S7C). Octanucleotide SSRs were the second most frequent category (846 occurrences), followed by dinucleotide (612 occurrences) and nonanucleotide (502 occurrences) SSRs. The remaining SSR types presented progressively lower abundances, with heptanucleotide SSRs representing the rarest category (< 0.1% of total repeats), suggesting potential selective constraints on repeat stability or functional integrity.

SSR distribution across genomic regions revealed distinct localization patterns: intergenic spacer (IGS) regions presented the highest SSR density (2,241 occurrences), followed by introns (1,278 occurrences) and coding sequences (1,074 occurrences). This distribution reflects relaxed selective pressures in noncoding regions, permitting greater accumulation of repetitive elements (Table S7A).

Motif composition analysis revealed a pronounced AT bias, mirroring the overall chloroplast genome nucleotide composition. Among the mononucleotide repeats, A/T motifs overwhelmingly dominated. For dinucleotide repeats, AT/TA motifs occurred most frequently, followed by AG/CT motifs, whereas GC-rich motifs were rare. This AT enrichment extended to longer repeat units, where AT-rich motifs consistently predominated (Table S7B).

The observed SSR patterns provide valuable molecular markers for population genetics, phylogeography, and breeding programs in Polygonaceae. Additionally, the preferential localization of SSRs in noncoding regions may influence chloroplast genome stability and gene regulation, with potential implications for the biosynthesis of pharmacologically active secondary metabolites.

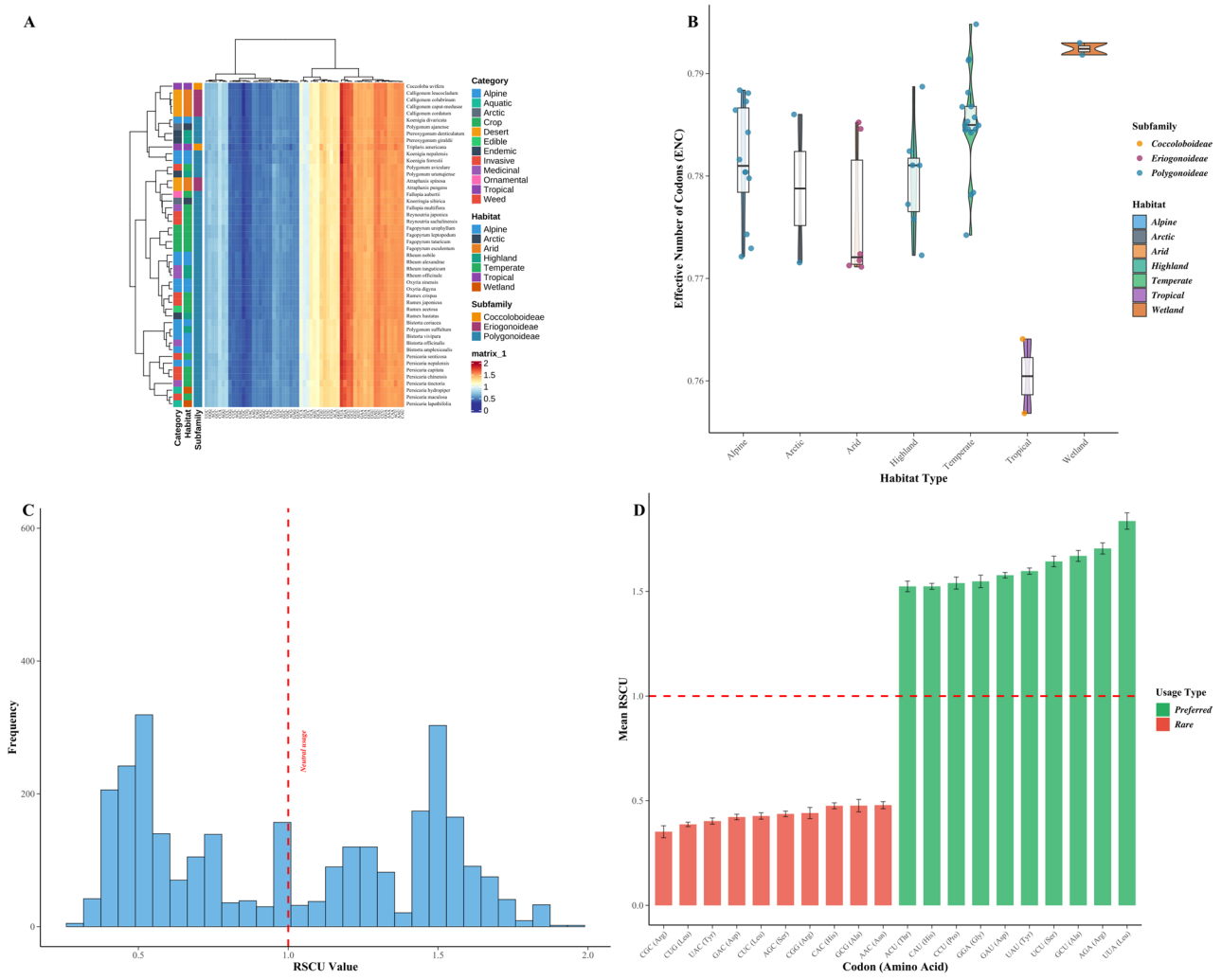

Fig. 6 CUB analysis across 48 Polygonaceae chloroplast genomes. Comprehensive evaluation reveals subfamily specific evolutionary patterns and habitat-dependent optimization strategies. **A**. Hierarchical clustering heatmap of RSCU values for 48 species representing three subfamilies (Polygonoideae, Eriogonoideae, and Coccoloboideae). Clear phylogenetic and ecological groupings are evident. Species are annotated by habitat type (alpine, temperate, arid, tropical, highland, Arctic, wetland) and ecological usage (crop, medicinal, weed, endemic, etc.). Color scale: blue = low usage (RSCU < 1), red = high usage (RSCU > 2) (see Table S6A). **B**. Distribution of the effective number of codons (ENC) across species, illustrating habitat and subfamily specific variation. Violin plots indicate density distributions; box plots represent quartiles; points represent individual species colored by subfamily. Lower ENC values reflect stronger codon bias, indicating adaptation to environmental constraints (see Table S6B). **C**. Frequency distribution of RSCU values across all 48 species, revealing overall CUB patterns. The dashed red line indicates neutral usage (RSCU = 1.0) (see Table S6C). **D**. Mean RSCU values for the top 10 preferred (green) and 10 rare (red) codons across all species. The error bars indicate the standard deviation. The dashed red line denotes neutral usage (RSCU = 1.0) (see Table S6D). This integrated analysis highlights evolutionary constraints and adaptive strategies in chloroplast codon optimization across the phylogenetic and ecological diversity of Polygonaceae, encompassing crops (e.g., buckwheat), medicinal plants (e.g., rhubarb), and wild taxa adapted to extreme environments

ENC analysis

ENC analysis across the 48 Polygonaceae species revealed codon usage bias patterns ranging from 27.2 to 61 across all protein-coding genes, with species-level means clustering between 51.95 and 53.72, indicative of the relatively weak CUB characteristic of plant chloroplast genomes (Fig. 8, Table S8). The GC3 values showed a narrow distribution, confirming the AT-rich composition and supporting the family-wide preference for A/U-ending codons observed in the RSCU analysis.

In the ENC-GC3 plots, the majority of protein-coding genes from all the species were positioned markedly below the theoretical neutrality curve, with only approximately 5–10% of the genes lying on or near the expected curve. Most genes clustered in the lower region, exhibiting ENC values 10–15 units below theoretical predictions, indicating that natural selection exerts a predominant influence over codon usage rather than neutral mutational pressure.

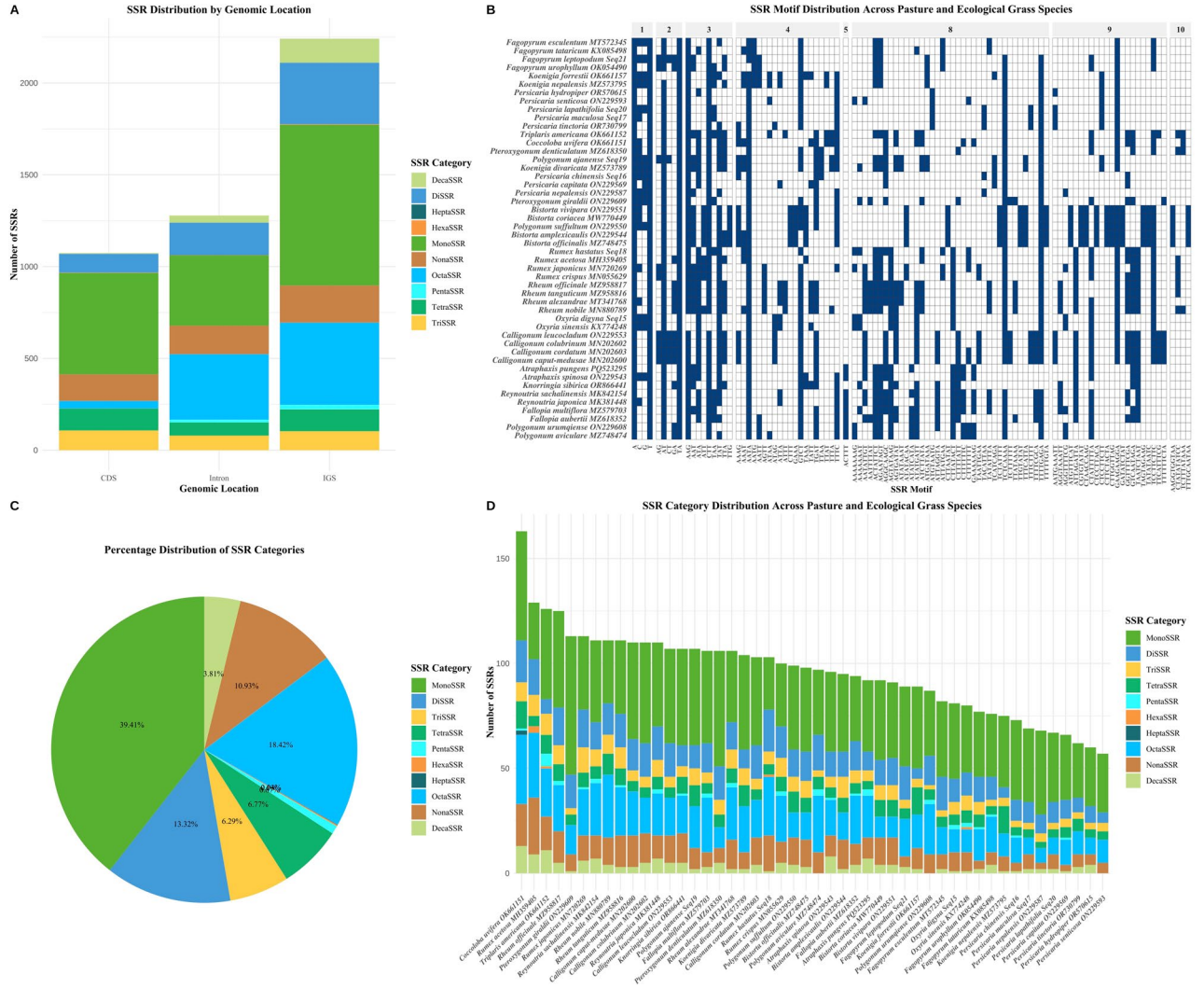

Fig. 7 Simple sequence repeat (SSR) analysis of the chloroplast genomes of 48 Polygonaceae species. **A**. Distribution of SSR categories across different genomic regions (PCG, intergenic spacers, introns, tRNA, and rRNA). The numbers indicate the number of each SSR type within specific regions. **B**. Heatmap showing the presence/absence patterns of SSR motifs across 48 species. Blue indicates presence, and white indicates absence. SSR motifs are grouped by repeat unit length (1–10 bp) and arranged using hierarchical clustering. **C**. Pie chart depicting the relative abundance of different SSR categories across all analyzed species. **D**. Stacked bar chart showing the distribution and abundance of SSR categories across individual species, ordered by total SSR count. SSR categories are defined by repeat unit length: monoSSR (1 bp), DiSSR (2 bp), TriSSR (3 bp), TetraSSR (4 bp), PentaSSR (5 bp), HexaSSR (6 bp), HeptaSSR (7 bp), OctaSSR (8 bp), NonaSSR (9 bp), DecaSSR (10 bp), and ExtendedSSR (> 10 bp). This analysis reveals the predominance of A/T-rich mononucleotide repeats and highlights variation in SSR abundance across coding and noncoding regions, providing valuable markers for population genetics, phylogeography, and molecular breeding in Polygonaceae

Gene functional classification revealed differential codon usage patterns across chloroplast gene categories. The *rbcL* (ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit) genes presented the strongest codon bias, which was significantly lower than that of the other categories, reflecting strong selective pressure for translational optimization of this key photosynthesis-related enzyme. Small ribosomal protein genes displayed the weakest bias, suggesting relaxed constraints on codon usage for these structural components. NADH dehydrogenase genes showed relatively strong bias, whereas ATP synthase genes exhibited moderate bias, highlighting functional differences in selective pressures among chloroplast gene categories.

These results indicate that codon usage in Polygonaceae chloroplast genomes is shaped by both evolutionary and functional constraints, with translational selection likely optimizing the expression of photosynthesis-related and metabolically critical genes, which may, in turn, influence secondary metabolite production relevant to phytotherapeutic applications.

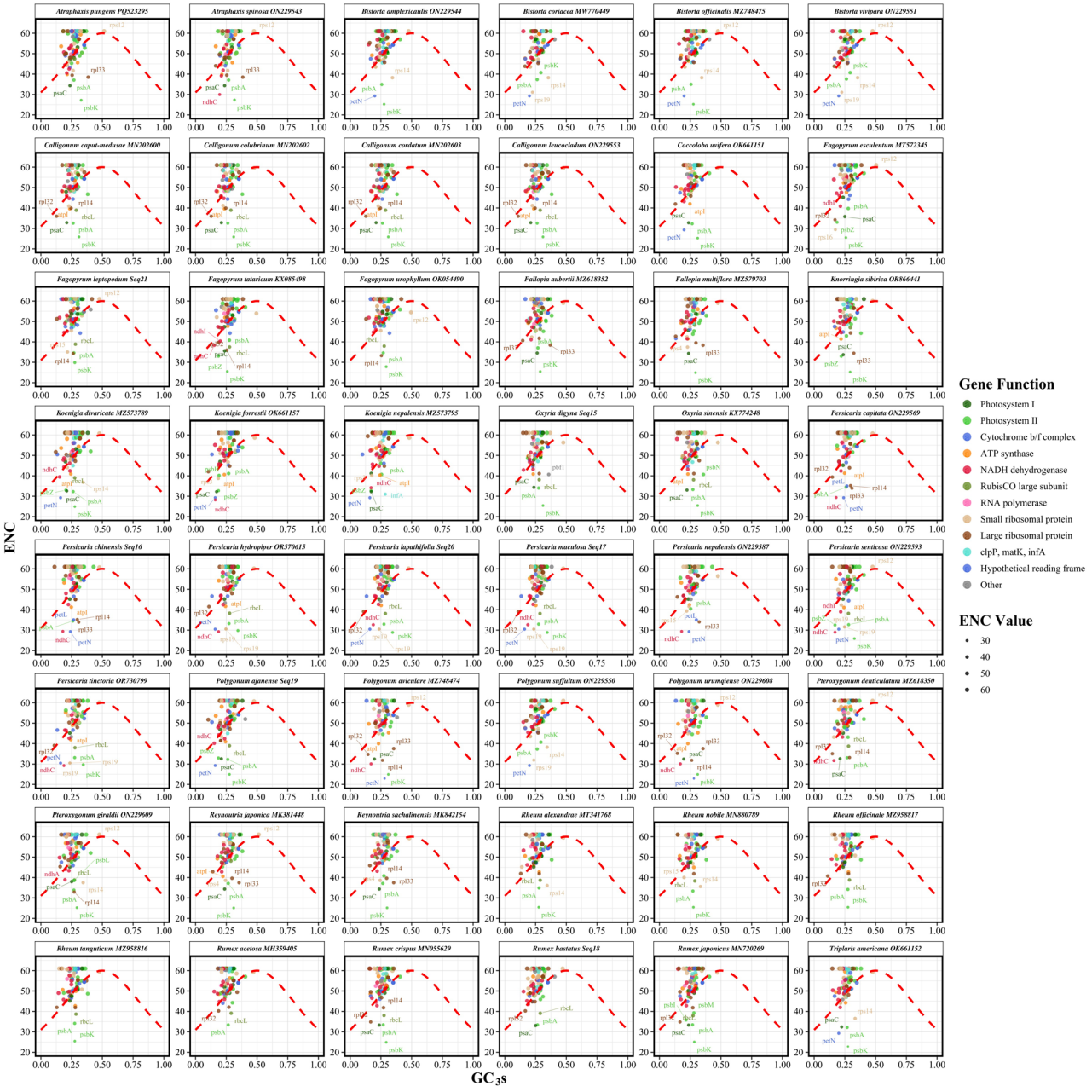

Fig. 8 ENC plotted against the GC content at synonymous third codon positions (GC3s) for 48 Polygonaceae species. Each point represents a protein-coding gene, colored according to functional category: photosystem I (dark green), photosystem II (light green), cytochrome b_6/f complex (royal blue), ATP synthase (dark orange), NADH dehydrogenase (crimson), RubisCO large subunit (*rbcL*, olive), RNA polymerase (hot pink), small ribosomal proteins (burlywood), large ribosomal proteins (saddle brown), *clpP/matK/infA* (turquoise), hypothetical reading frames (royal blue), tRNA (black), rRNA (medium purple), and other genes (gray). The point size corresponds to the ENC value. The red dashed line indicates the theoretical neutrality curve. Genes clustering below this line exhibit CUB that is stronger than expected under neutral mutation, reflecting selective pressures. Functional grouping revealed that photosynthesis-related genes (e.g., *rbcL*, photosystem subunits) and NADH dehydrogenase genes display distinct codon usage patterns, suggesting differential evolutionary constraints within Polygonaceae chloroplast genomes

PR2 analysis

PR2 analysis revealed pronounced asymmetry in third-position nucleotide composition across all investigated Polygonaceae species. The gene distributions deviated substantially from the theoretical center (0.5, 0.5), which represents neutral base usage. The A3/(A3 + T3) ratios ranged from 0.32 to 0.58, indicating a systematic bias toward thymine over adenine at synonymous third codon positions (Fig. 9, Table S9). The G3/(G3 + C3) ratios exhibited a broader range, reflecting a consistent preference for guanine over cytosine at synonymous sites.

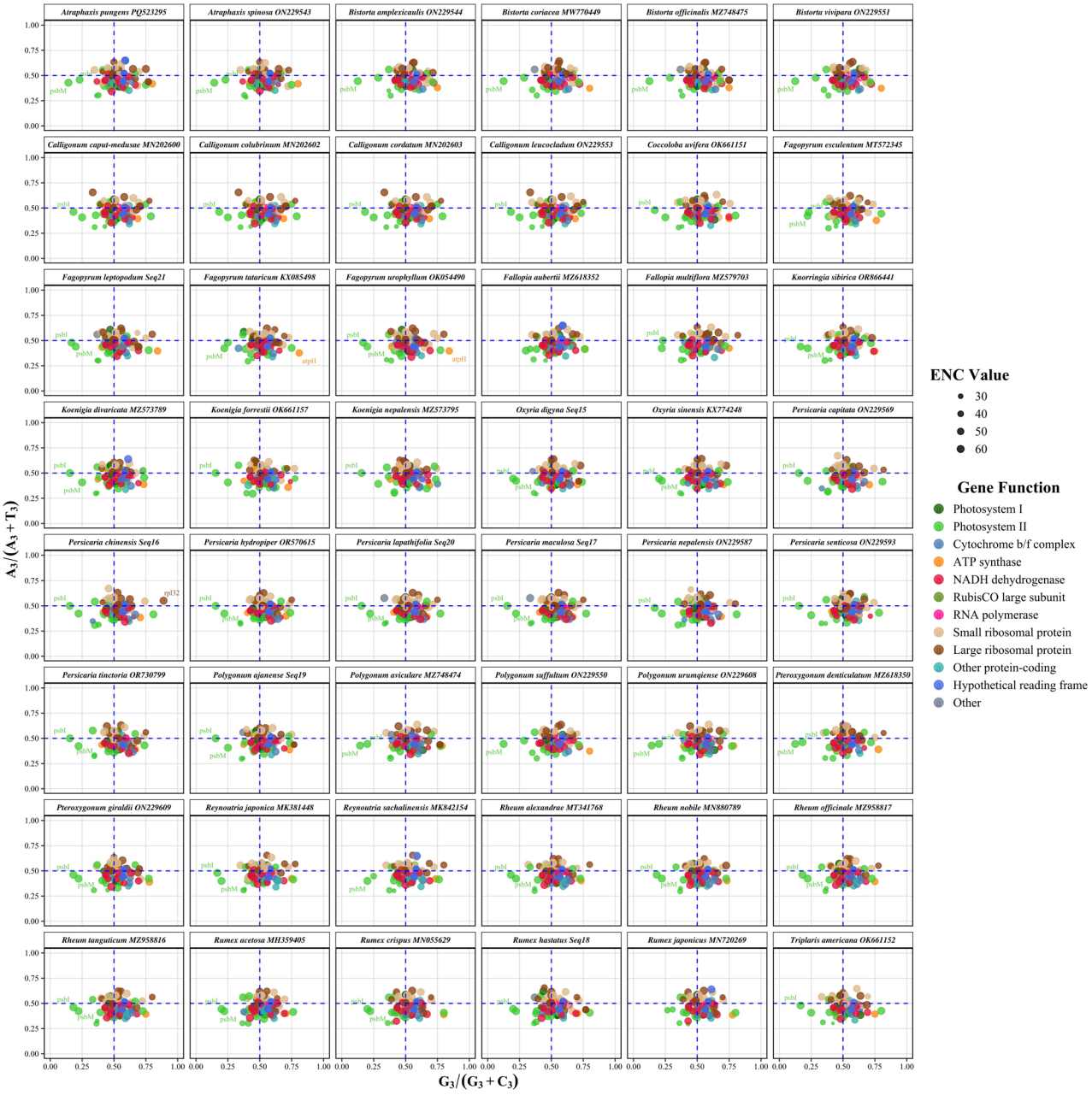

Fig. 9 Parameter rule 2 (PR2) analysis of codon usage patterns in Polygonaceae chloroplast genomes. The plot depicts the relationship between A3/(A3 + T3) and G3/(G3 + C3) ratios at synonymous third codon positions for all protein-coding genes across 48 Polygonaceae species. Points are colored according to gene functional categories. Individual gene names are annotated using the ggrepel algorithm to prevent overlapping labels, providing clear identification of genes contributing to codon usage patterns. Optimized label positioning allows detailed visualization of genes with unusual codon usage characteristics. A deviation from the central point (0.5, 0.5) indicates asymmetry in base composition, reflecting the relative influences of natural selection and mutational pressure. The quadrant distribution highlights trends such as preferential T and G usage at third codon positions, with functional distinctions evident among photosynthesis-related genes, ribosomal proteins, and NADH dehydrogenase genes. This annotated version enables comprehensive assessment of individual gene contributions to overall CUB within the Polygonaceae family

This compositional asymmetry was consistently observed across the three major Polygonaceae subfamilies, namely, Polygonoideae, Eriogonoideae, and Coccoloboideae, suggesting family-wide evolutionary constraints on chloroplast codon usage. The gene distribution within the PR2 plot quadrants was highly skewed: 68.4% of the genes clustered in the fourth quadrant (A3/(A3 + T3) < 0.5, G3/(G3 + C3) > 0.5), reflecting preferential thymine and guanine usage at the third codon position. The second quadrant contained 21.7% of the genes, whereas the first and third quadrants included only 6.2% and 3.7% of the genes, respectively (Fig. 9, Table S9). This

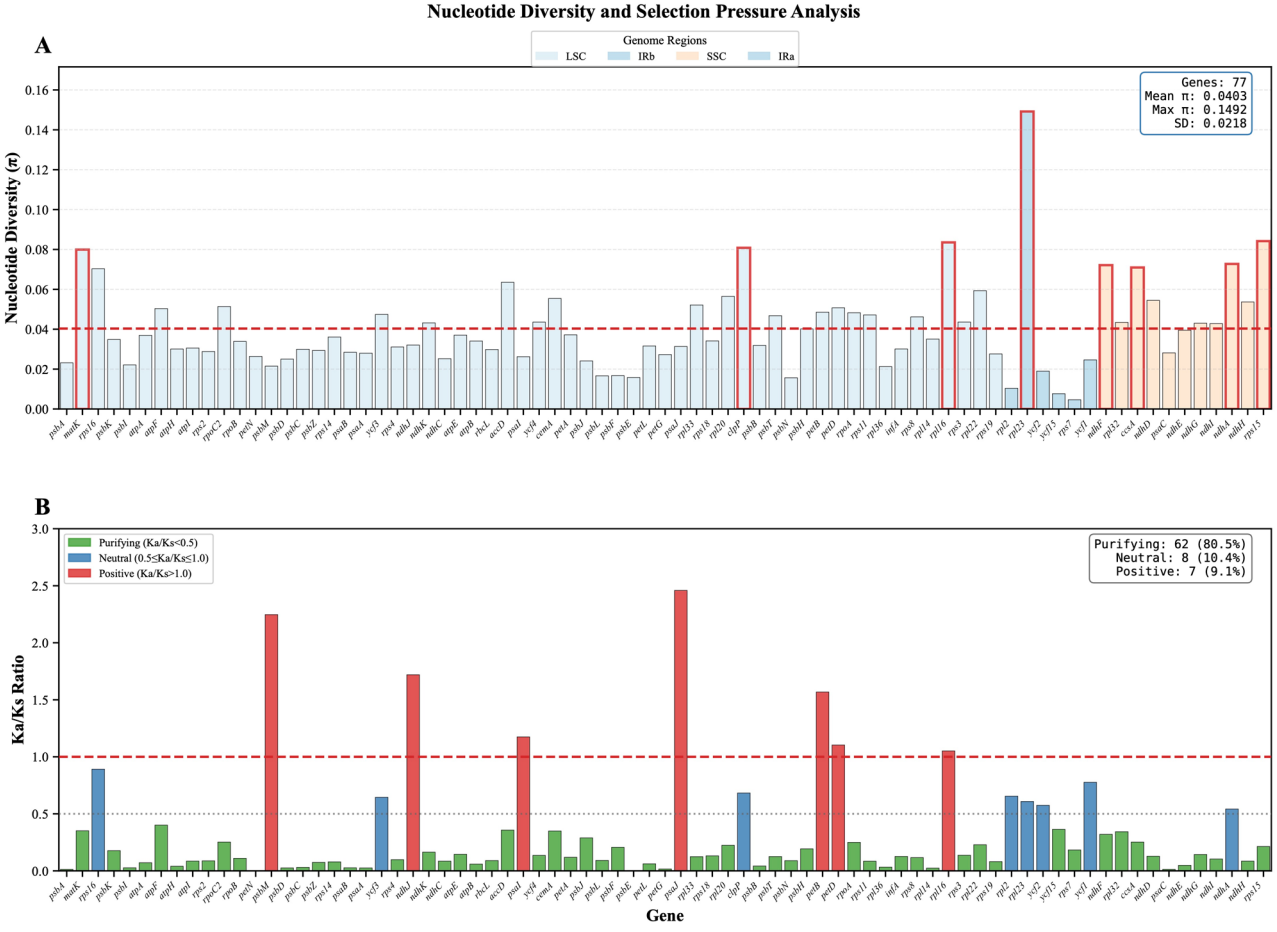

Fig. 10 This figure illustrates the nucleotide diversity (Panel **A**) and selection pressure distribution, as measured by the Ka/Ks ratio (Panel **B**), across different genomic regions. Panel A displays the nucleotide diversity (π) values for the ISC, Rb, SSC, and IRa regions, ranging from approximately 0.00-0.16. Panel B categorizes genes into three types on the basis of their Ka/Ks ratios: purifying selection (Ka/Ks < 0.5), neutral selection (0.5 ≤ Ka/Ks < 1.0), and positive selection (Ka/Ks ≥ 1.0), with the ratio scale extending up to approximately 7.5

pronounced quadrant asymmetry indicates the action of selective pressures on codon composition, which contrasts with the equal quadrant distribution expected under neutral evolution.

Functional gene categorization revealed distinct PR2 patterns associated with cellular roles. The photosystem I-related genes presented the most pronounced thymine and guanine bias, the photosystem II genes presented similar but slightly weaker biases, and the NADH dehydrogenase genes presented intermediate levels of compositional asymmetry. These results suggest that codon usage bias in Polygonaceae chloroplast genomes is shaped by both evolutionary constraints and functional optimization, potentially influencing translational efficiency and the biosynthesis of secondary metabolites relevant to pharmacological applications.

Nucleotide diversity and selection pressure analysis

Analysis of 77 protein-coding genes across 48 Polygonaceae chloroplast genomes revealed considerable heterogeneity in nucleotide diversity (π), with values ranging from 0.0076 (*ycf15*) to 0.1492 (*rpl23*) and a mean of 0.0403 (Fig. 10A). Diversity was unevenly distributed among functional categories. The genes encoding ribosomal proteins, particularly the large subunit, presented the highest mean diversity (π = 0.0654). In contrast, genes involved in the core photosynthetic machinery (Photosystems I and II) were highly conserved, with the lowest mean diversity (π = 0.0312 and 0.0298, respectively). At the genomic level, average diversity was highest in the small single-copy region (π = 0.0498), intermediate in the large single-copy region (π = 0.0391), and lowest in the inverted repeat region (π = 0.0285), reflecting the homogenizing effect of gene conversion in the IRs.

The ratio of nonsynonymous to synonymous substitutions (Ka/Ks) indicated that strong purifying selection was the dominant evolutionary force, constraining 80.5% of the genes (Ka/Ks < 0.5) (Fig. 10B). This was especially pronounced for the photosystem and ATP synthase genes, whose mean Ka/Ks ratios were less than

0.1. Evidence for positive selection (Ka/Ks > 1.0) was detected in 9.1% of the genes. Notably, *psaJ* (Photosystem I) presented the highest Ka/Ks ratio (2.89), suggesting potential adaptive evolution. The genes encoding the NADH dehydrogenase complex (*ndh*) and the maturase *matK* displayed more heterogeneous and relaxed selection patterns, which is consistent with greater functional flexibility.

Discussion

This study, incorporating seven newly sequenced and 41 publicly available chloroplast genomes, provides comprehensive insights into codon usage evolution, molecular marker development, and phylogenomic relationships across this economically and medicinally important plant family [75, 76]. By integrating codon usage bias (CUB) analysis, SSR characterization, selective pressure quantification (Ka/Ks ratios), and complete chloroplast genome phylogenetic reconstruction, we demonstrate that natural selection predominantly governs synonymous codon evolution in Polygonaceae, with distinct subfamily-specific clustering patterns reflecting adaptive optimization strategies across diverse ecological niches.

Compositional architecture and evolutionary determinants of chloroplast codon usage patterns

Chloroplast genomes in Polygonaceae exhibit pronounced adenine-thymine (AT) enrichment, with overall GC contents ranging from 36.59% to 38.22%, reflecting compositional biases characteristic of plant organellar genomes [77]. Positional GC content analysis revealed the conserved pattern GC1 > GC2 > GC3, with GC3s averaging 27.82%, indicative of substantial codon usage bias at synonymous third-codon positions [78]. This hierarchical GC distribution reflects differential selective constraints, wherein first and second codon positions experience stronger purifying selection owing to their direct impact on amino acid identity, whereas third positions permit greater latitude for synonymous substitutions.

RSCU analysis identified 29 codons exhibiting positive bias (RSCU > 1.0), 28 of which terminated in adenine or uracil, demonstrating a family-wide preference for A/U-ending codons [66]. Six optimal codons (AGA, GAU, GCU, UAU, UCU, and UUA) displayed consistent overrepresentation across all the examined taxa, suggesting functional optimization for translational efficiency [79, 80]. The ENC values ranged from 46.36 to 48.28, reflecting weak overall codon bias consistent with organellar genome characteristics. However, hierarchical clustering on the basis of RSCU values revealed clear phylogenetic structuring, with species grouped according to subfamily affiliation and ecological adaptation, implying that evolutionary history and environmental pressures jointly shape synonymous codon bias.

Multiple analytical frameworks, including ENC-GC3 analysis, neutrality analysis, and PR2 analysis, converge to demonstrating that natural selection, rather than mutational pressure, is the dominant evolutionary force shaping codon usage bias in Polygonaceae chloroplast genomes. In the ENC-GC3 plots, the majority of protein-coding genes clustered substantially below the theoretical neutrality curve, with deviation magnitudes of 10–15 ENC units, indicating selection-driven optimization rather than mutation-driven compositional equilibrium. Neutrality analysis revealed weak correlations between GC_{12} and GC_3 (regression slopes 0.04–1.26), with most species exhibiting slopes approaching zero, further supporting the predominance of natural selection over mutational pressure. PR2 analysis demonstrated pronounced base composition asymmetry, with 68.4% of genes clustering in the fourth quadrant ($A3/(A3+T3) < 0.5$ and $G3/(G3+C3) > 0.5$), reflecting nonrandom, selection-driven codon bias rather than neutral compositional drift.

RSCU values are widely used to quantify codon bias across genomic sequences [81]. Mutation-dominated codon usage typically results in RSCU values converging toward unity, reflecting uniform utilization of synonymous codons. In Polygonaceae, our RSCU analysis revealed extensive variation among taxa, with thirteen highly conserved preferred codons ending in A/U. Species-specific codon bias is shaped by both evolutionary history and ecological adaptation, as exemplified by the distinct RSCU patterns of *T. americana* and *O. digyna*, which occupy markedly divergent habitats [82].

Functional gene categorization further highlighted that differential codon usage patterns were correlated with physiological importance. Photosynthesis-related genes, particularly those encoding photosystem subunits and the RubisCO large subunit (*rbcL*), exhibited significantly stronger codon bias than ribosomal protein genes did, reflecting intensified selective pressure for translational optimization in photosynthetic apparatus components [83]. This observation aligns with the codon adaptation hypothesis, which proposes that highly expressed genes evolve bias toward codons matching abundant tRNA isoacceptors, thereby increasing translation efficiency and minimizing ribosomal pausing [84, 85]. Similar patterns have been reported across diverse angiosperm lineages, where natural selection favors AT-ending codons correlated with gene expression levels [86].

Simple sequence repeat polymorphisms and molecular marker resources

Chloroplast SSRs constitute highly polymorphic tandemly repeated motifs that are extensively employed in population genetics, phylogeography, and molecular

breeding owing to their codominant inheritance, high reproducibility, and genome-wide distribution [87, 88]. We identified 4,593 SSR loci across 48 Polygonaceae chloroplast genomes, with mononucleotide repeats predominating (39.41%), followed by octanucleotide (18.42%), dinucleotide (13.33%), and nonanucleotide (10.93%) repeats. The prevalence of A/T mononucleotide repeats over G/C motifs reflects the overall AT-rich chloroplast composition, with CCG/CGG dinucleotide repeats being rare or absent, which is consistent with angiosperm patterns [89, 90]. Regional distribution revealed substantial heterogeneity: intergenic spacer (IGS) regions harbored 2,241 SSRs (48.8%), introns 1,278 SSRs (27.8%), and coding sequences 1,074 SSRs (23.4%). This pattern reflects relaxed selective constraints in noncoding regions, permitting greater accumulation of repetitive elements without impacting protein function. These SSRs provide valuable molecular markers for genetic diversity studies, cultivar fingerprinting, parentage analysis, and tracing of maternal lineages in phylogeographic reconstructions [79, 91, 92]. ENC values across species ranged from 40 to 61, quantitatively assessing deviations from random codon usage. Neutrality analysis indicated greater constraints at the first and second codon positions (GC12) than at the third codon position (GC3), suggesting that CUB is shaped predominantly by natural selection, optimizing translational efficiency and energy utilization. Codon bias diversity was visualized using correspondence analysis, revealing clustering patterns consistent with phylogenetic and functional constraints [93].

Selective pressures, structural collinearity, and phylogenomic resolution

Ka/Ks ratios provide quantitative measures of selection constraints, with values < 1 indicating purifying selection, ≈ 1 indicating neutral evolution, and > 1 indicating positive selection [94]. In this study, analysis of 79 protein-coding genes across 2,622 pairwise comparisons revealed photosystem genes presented the greatest degree of conservation, followed by the ATP synthase and ribosomal protein genes. The NADH dehydrogenase (*ndh*) genes presented relatively relaxed constraints (85.4% under strong purifying selection), with *ndhD*, *ndhF*, *ndhH*, and *matK* occasionally showing Ka/Ks ≥ 1, which is consistent with documented *ndh* evolutionary dynamism and occasional pseudogenization in angiosperms [95–98]. The *ndh* complex contributes to cyclic electron flow around photosystem I, with functional redundancy potentially relaxing selective constraints relative to the core photosynthetic machinery [97].

Collinearity analysis revealed exceptional structural conservation within genera (*Calligonum*, *Fagopyrum*, *Rheum*), with sequence identities > 98% [99, 100]. Coding regions are more conserved than intergenic spacers are, and IRs display minimal variation relative to single-copy regions, reflecting stable gene conversion and purifying selection [101]. Minor structural variations, including IR boundary shifts and small indels, were phylogenetically informative, facilitating the resolution of subfamily and genus-level relationships.

Chloroplast genome-scale phylogenies have repeatedly demonstrated their ability to stabilize backbone relationships and provide strong support for genus-level hypotheses, especially when compared with earlier studies that relied on a small number of markers such as *rbcL* [6]. The topology recovered here is broadly congruent with the major splits identified in Polygonaceae-wide molecular syntheses, while offering clearer internal resolution and more consistent node support across subfamilies and tribes [1, 3, 6]. This improvement is expected given the larger number of informative sites in concatenated plastid coding matrices and the fit of modern likelihood models and support procedures implemented in contemporary tree-search frameworks [50, 53].

The repeated recovery of monophyly for *Fagopyrum*, *Rheum*, and *Rumex* strengthens confidence in chloroplast genome-based identification and phylogenetic placement for economically important and medicinal taxa, and it aligns with recent genus-focused chloroplast genome phylogenomic work in *Fagopyrum* and *Rheum* [2, 102]. At the same time, the signal of *Persicaria* nonmonophyly and the embedding of Bistorta-like lineages within a broader *Persicaria–Koenigia* complex reinforces the view that traditional circumscriptions in parts of Polygonoideae remain difficult, even with chloroplast genome data. One plausible explanation is that short internal branches reflect rapid radiation, while another is that the chloroplast genome represents a single inherited locus that can be displaced by introgression and chloroplast capture, yielding topologies that are locally discordant with the species history. This is not a hypothetical edge case, because gene-tree conflict within plastomes and chloroplast genome–nuclear discordance are now documented across broad angiosperm radiations [103], and chloroplast genomes can move across species boundaries through grafting, hybridization, or other transfer processes [104]. In Polygonaceae, where hybridization and complex species boundaries are suspected in several genera, plastome-based inferences should therefore be interpreted as a robust maternal-lineage hypothesis that benefits from validation against nuclear datasets. The trait reconstructions mapped onto the tree provide a complementary evolutionary layer that is useful for connecting plastome structural dynamics to lineage diversification. The prominent role of IR boundary change in explaining genome size differences is consistent with general chloroplast genome evolutionary principles and underscores why size-associated traits often show strong

phylogenetic signal [15, 18]. From an applied perspective, a stable, well-supported chloroplast genome phylogeny provides a practical scaffold for marker selection, taxon authentication, and comparative analyses of codon bias or selection patterns along lineages, but downstream ecological or pharmacological interpretation still benefits from integrating nuclear markers, population-level sampling, and explicit tests for introgression.

Taken together, the phylogenetic results support a chloroplast genome-informed framework that largely matches established Polygonaceae systematics while clarifying several previously ambiguous relationships. The remaining areas of uncertainty, especially those associated with short internodes and potential reticulation, point to a clear next step that combines chloroplast genome data with nuclear phylogenomics to disentangle lineage sorting from chloroplast capture and to refine genus boundaries where plastid signal alone is insufficient.

Conclusions

This integrative analysis of 48 Polygonaceae chloroplast genomes provides novel insights into chloroplast genome architecture, codon usage evolution, selective pressures, and phylogenetic relationships. The chloroplast genome exhibits four characteristics: significant size variation, conserved tetragonal organization, AT rich composition, and variable gene content. Codon usage was biased toward A/U-ending codons. Multiple analyses confirmed that natural selection is the primary force shaping synonymous codon usage. Ka/Ks analyses highlighted strong purifying selection in photosystem genes and relatively relaxed constraints in *ndh* genes, which is consistent with functional dispensability. A total of 4,593 SSR loci, predominantly mononucleotide repeats in intergenic regions (48.8%) and introns (27.8%), provide valuable molecular markers for population genetics, cultivar identification, and phylogeography. Chloroplast genome structural collinearity was exceptionally conserved within genera, with coding regions more constrained than noncoding regions. Phylogenomic reconstruction robustly resolved subfamily and genus-level relationships, with codon usage clustering reflecting both evolutionary history and ecological adaptation.

These findings advance our understanding of chloroplast genome evolution in angiosperms and establish resources for practical applications, including heterologous protein expression optimization, molecular marker development, and investigations of adaptive evolution, biogeographic history, and secondary metabolite biosynthesis in this ecologically and economically important plant family. Future work should focus on functional validation of codon bias effects, integration with nuclear genome data, and empirical assessment of SSR markers for breeding and conservation purposes.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12870-026-08560-4.

Supplementary Material 1: Table S1. Detailed characteristics of the complete chloroplast genomes of 48 Polygonaceae species, including large single-copy (LSC), small single-copy (SSC), and inverted repeat (IR) regions; GC content; and gene content.

Supplementary Material 2: Table S2. Information on the Polygonaceae species used in the phylogenetic analysis.

Supplementary Material 3: Table S3. Ancestral State Reconstruction of 48 Polygonaceae species.

Supplementary Material 4: Table S4. rRNA gene statistics, PCG gene statistics, tRNA gene statistics, and correlation matrix of 48 Polygonaceae species.

Supplementary Material 5: Table S5. Collinearity analysis of 48 Polygonaceae species.

Supplementary Material 6: Table S6. Relative synonymous codon usage (RSCU) values for 48 Polygonaceae species used to generate hierarchical clustering heatmaps, showing codon preference patterns across subfamilies and ecological adaptations.

Supplementary Material 7: Table S7. Distribution of simple sequence repeats (SSRs) across different genomic regions [protein-coding genes (PCG), introns, intergenic spacers (IGS), tRNA, rRNA] in 48 Polygonaceae chloroplast genomes, including counts and motif compositions.

Supplementary Material 8: Table S8. Summary statistics of codon usage bias (CUB) analysis for 48 Polygonaceae chloroplast genomes, including the effective number of codons (ENC), GC content at codon positions, and GC3 values.

Supplementary Material 9: Table S9. Statistical summary of parity rule 2 (PR2) analysis for CUB in 48 Polygonaceae chloroplast genomes, including quadrant distributions and base composition asymmetry measures.

Supplementary Material 10: Table S10. Nucleotide diversity and selection statistics for 77 chloroplast genes. π, nucleotide diversity; SE, standard error; θ, Watterson's estimator; Hd, haplotype diversity; Ka/Ks, ratio of non-synonymous to synonymous substitutions; S, number of synonymous sites; N, number of non-synonymous sites; Sig., significance of the Fisher exact test (*P < 0.05, **P < 0.01, ***P < 0.001, ns not significant).

Acknowledgements

We gratefully thank the anonymous reviewers for their constructive comments which significantly improved the manuscript.

Authors' contributions

Conceptualization, Y.G.; methodology, H.W., and Y.G.; software, Y.G.; validation, Y.G.; formal analysis, H.W. and Y.G.; data curation, H.W. and Y.G.; writing—original draft preparation, H.W., Z.W., Z.L., G.S., X.Z., D.L., and Y.G.; writing—review and editing, H.W., Z.W., Z.L., G.S., X.Z., D.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 work was supported by the National Undergraduate Innovation and Entrepreneurship Training Program (Project No. 202510022484) and the Fundamental Research Funds for the Central Universities (No. ZZK202503).

Data availability

All data generated or analysed during this study are included in this published article and its supplementary information files.

Declarations

Ethics approval and consent to participate
Not applicable.
The data submitted to the system is identical to the data in the main manuscript and is available.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]School of Grassland Science, Beijing Forestry University, Beijing 100083, China
[2]Department of Biology, Vrije Universiteit Brussel, Brussels 1050, Belgium

Received: 5 January 2026 / Accepted: 9 March 2026
Published online: 17 March 2026

References

1. Burke JM, Sanchez A. Revised subfamily classification for Polygonaceae, with a tribal classification for Eriogonoideae. Brittonia. 2011;63(4):510–20.
2. Zhang HJ, Zhang X, Landis JB, Sun YX, Sun J, Kuang TH, Li LJ, Tiamiyu BB, Deng T, Sun H. Phylogenomic and comparative analyses of *Rheum* (Polygonaceae, Polygonoideae). J Syst Evol. 2022;60(6):1229–40.
3. Sanchez A, Schuster TM, Kron KA. A large-scale phylogeny of Polygonaceae based on molecular data. Int J Plant Sci. 2009;170(8):1044–55.
4. DECRAENE L-PR, Akeroyd J. Generic limits in Polygonum and related genera (Polygonaceae) on the basis of floral characters. Bot J Linn Soc. 1988;98(4):321–71.
5. Kubitzki K, Rohwer JG, Bittrich V: Flowering plants· dicotyledons: Magnoliid, hamamelid and caryophyllid families: Springer Science & Business Media; 2013.
6. Frye ASL, Kron KA. *rbcL* phylogeny and character evolution in Polygonaceae. Syst Bot. 2003;28(2):326–32.
7. Demirpolat A. Chemical composition of essential oils of seven *Polygonum* species from Turkey: A chemotaxonomic approach. Molecules. 2022;27(24):9053.
8. Inoue M, Nishimura H, Li H, Mizutani J. Allelochemicals from *Polygonum sachalinense* Fr. Schm.(Polygonaceae). J Chem Ecol. 1992;18(10):1833–40.
9. Zhang Z-L, Li Y-Z, Wu G-Q, Zhang D-D, Deng C, Wang Z-M, Song X-M, Wang W. A comprehensive review of traditional uses, phytochemistry and pharmacology of *Reynoutria* genus. J Pharm Pharmacol. 2022;74(12):1718–42.
10. Bounda G-A, Feng Y. Review of clinical studies of *Polygonum multiflorum* Thunb. and its isolated bioactive compounds. Pharmacognosy Res. 2015;7(3):225.
11. Türk Z, Leiber F, Schlittenlacher T, Hamburger M, Walkenhorst M. Multiple benefits of herbs: Polygonaceae species in veterinary pharmacology and livestock nutrition. Veterinary Anim Sci. 2025;27:100416.
12. Forman J, Kesseli RV. Sexual reproduction in the invasive species *Fallopia japonica* (Polygonaceae). Am J Bot. 2003;90(4):586–92.
13. Weston LA, Barney JN, DiTommaso A. A review of the biology and ecology of three invasive perennials in New York State: Japanese knotweed (*Polygonum cuspidatum*), mugwort (*Artemisia vulgaris*) and pale swallow-wort (*Vincetoxicum rossicum*). Plant Soil. 2005;277(1):53–69.
14. Geng X, Huang N, Zhu Y, Qin L, Hui L. Codon usage bias analysis of the chloroplast genome of cassava. South Afr J Bot. 2022;151:970–5.
15. 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.
16. Li B, Zheng Y. Dynamic evolution and phylogenomic analysis of the chloroplast genome in Schisandraceae. Sci Rep. 2018;8(1):9285.
17. Cao Z, Qu Y, Song Y, Xin P. Comparative genomics and phylogenetic analysis of chloroplast genomes of Asian Caryodaphnopsis taxa (Lauraceae). Gene. 2024;907:148259.
18. 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.
19. Sun M, Zhang J, Huang T, Yang M, Ma L, Duan L. Genome structure and variation of *Reynoutria japonica* Houtt. chloroplast genome. Sheng wu Gong Cheng xue bao= Chin J Biotechnol. 2022;38(5):1953–64.
20. Palmer JD. Chloroplast DNA exists in two orientations. Nature. 1983;301(5895):92–3.
21. Sato S, Nakamura Y, Kaneko T, Asamizu E, Tabata S. Complete structure of the chloroplast genome of *Arabidopsis thaliana*. DNA Res. 1999;6(5):283–90.
22. Sheng W. The complete chloroplast genome of *Gerbera piloselloides* (L.) Cass., 1820 (Carduoideae, Asteraceae) and its phylogenetic analysis. Open Life Sci. 2025;20(1):20251070.
23. Ruhlman T, Lee S-B, Jansen RK, Hostetler JB, Tallon LJ, Town CD, Daniell H. Complete plastid genome sequence of *Daucus carota*: implications for biotechnology and phylogeny of angiosperms. BMC Genomics. 2006;7(1):222.
24. Matsushita T, Kano-Sueoka T. Non-random codon usage of synonymous and non-synonymous mutations in the human HLA-A gene. J Mol Evol. 2023;91(2):169–91.
25. 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.
26. Parvathy ST, Udayasuriyan V, Bhadana V. Codon usage bias. Mol Biol Rep. 2022;49(1):539–65.
27. Quax TE, Claassens NJ, Söll D, van der Oost J. Codon bias as a means to fine-tune gene expression. Mol Cell. 2015;59(2):149–61.
28. Sharp PM, Emery LR, Zeng K. Forces that influence the evolution of codon bias. Philosophical Trans Royal Soc B: Biol Sci. 2010;365(1544):1203–12.
29. Li X, Liu L, Ren Q, Zhang T, Hu N, Sun J, Zhou W. Analysis of synonymous codon usage bias in the chloroplast genome of five *Caragana*. BMC Plant Biol. 2025;25(1):322.
30. Chakraborty S, Yengkhom S, Uddin A. Analysis of codon usage bias of chloroplast genes in *Oryza* species: Codon usage of chloroplast genes in *Oryza* species. Planta. 2020;252(4):67.
31. Zhang P, Xu W, Lu X, Wang L. Analysis of codon usage bias of chloroplast genomes in *Gynostemma* species. Physiol Mol Biology Plants. 2021;27(12):2727–37.
32. 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.
33. Chen H, Zhang J. Chloroplast genome evolution and codon usage in the medicinal plant *Pothos chinensis* (Araceae). Genes. 2025;16(9):1017.
34. Bímová K, Mandák B, Kašparová I. How does Reynoutria invasion fit the various theories of invasibility? J Veg Sci. 2004;15(4):495–504.
35. Komarova N, Barkova D, Kuznetsov A. Implementation of high-throughput sequencing (HTS) in aptamer selection technology. Int J Mol Sci. 2020;21(22):8774.
36. Atherton RA, McComish BJ, Shepherd LD, Berry LA, Albert NW, Lockhart PJ. Whole genome sequencing of enriched chloroplast DNA using the Illumina GAII platform. Plant methods. 2010;6(1):22.
37. Jin J-J, Yu W-B, Yang J-B, Song Y, DePamphilis CW, Yi T-S, Li D-Z. GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. Genome Biol. 2020;21(1):241.
38. Gallaher SD, Fitz-Gibbon ST, Strenkert D, Purvine SO, Pellegrini M, Merchant SS. High-throughput sequencing of the chloroplast and mitochondrion of Chlamydomonas reinhardtii to generate improved de novo assemblies, analyze expression patterns and transcript speciation, and evaluate diversity among laboratory strains and wild isolates. Plant J. 2018;93(3):545–65.
39. Jj D. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem Bull. 1987;19:11–15.
40. de Sena Brandine G, Smith AD. Falco: high-speed FastQC emulation for quality control of sequencing data. F1000Research. 2021;8:1874.
41. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30(15):2114–20.
42. Dierckxsens N, Mardulyn P, Smits G. NOVOPlasty: de novo assembly of organelle genomes from whole genome data. Nucleic Acids Res. 2017;45(4):e18–18.
43. Tillich M, Lehwark P, Pellizzer T, Ulbricht-Jones ES, Fischer A, Bock R, Greiner S. GeSeq–versatile and accurate annotation of organelle genomes. Nucleic Acids Res. 2017;45(W1):W6–11.
44. Shi L, Chen H, Jiang M, Wang L, Wu X, Huang L, Liu C. CPGAVAS2, an integrated plastome sequence annotator and analyzer. Nucleic Acids Res. 2019;47(W1):W65–73.
45. Lowe TM, Eddy SR. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. 1997;25(5):955–64.

46. Laslett D, Canback B. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. Nucleic Acids Res. 2004;32(1):11–6.
47. Benson DA, Cavanaugh M, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J. Sayers EW: GenBank. Nucleic Acids Res. 2012;41(D1):D36–42.
48. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013;30(4):772–80.
49. 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.
50. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, Von Haeseler A, Lanfear R. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. Mol Biol Evol. 2020;37(5):1530–4.
51. Kalyaanamoorthy S, Minh BQ, Wong TK, Von Haeseler A, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods. 2017;14(6):587–9.
52. Wang Z, Guo Y, Chen Y, Gao F, Xue Y, Niu Q, Li D, Smagghe G, Gai Y. Chloroplast genome sequencing of *Poa pratensis* and comparative chloroplast genomics analysis with other 29 grass species. Funct Integr Genom. 2026;26(1):44.
53. Hoang DT, Chernomor O, Von Haeseler A, Minh BQ, Vinh LS. UFBoot2: improving the ultrafast bootstrap approximation. Mol Biol Evol. 2018;35(2):518–22.
54. Anisimova M, Gascuel O. Approximate likelihood-ratio test for branches: a fast, accurate, and powerful alternative. Syst Biol. 2006;55(4):539–52.
55. Paradis E, Claude J, Strimmer K. APE: analyses of phylogenetics and evolution in R language. Bioinformatics. 2004;20(2):289–90.
56. Yu G, Smith DK, Zhu H, Guan Y, Lam TTY. ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. Methods Ecol Evol. 2017;8(1):28–36.
57. Revell LJ. phytools: an R package for phylogenetic comparative biology (and other things). Methods Ecol Evol. 2012;2:217–23.
58. Revell LJ. phytools 2.0: an updated R ecosystem for phylogenetic comparative methods (and other things). PeerJ. 2024;12:e16505.
59. Yoon SH, Ha SM, Lim J, Kwon S, Chun J. A large-scale evaluation of algorithms to calculate average nucleotide identity. Antonie Van Leeuwenhoek. 2017;110(10):1281–6.
60. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990;215(3):403–10.
61. Wang Y, Tang H, Debarry JD, Tan X, Li J, Wang X, Lee TH, Jin H, Marler B, Guo H, et al. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. Nucleic Acids Res. 2012;40(7):e49.
62. Darling AC, Mau B, Blattner FR, Perna NT. Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Res. 2004;14(7):1394–403.
63. 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.
64. Sharp PM, Li WH. An evolutionary perspective on synonymous codon usage in unicellular organisms. J Mol Evol. 1986;24(1–2):28–38.
65. Liu X, Xing H, Kong F, Zhang K, Cao Y, Guo X, Li Q, Wang J, Jing T, Zhan Y, et al. Molecular identification of F(1) hybrids of *Fraxinus mandshurica* × *Fraxinus chinensis* using SSR markers. Gene. 2025;959:149507.
66. Wright F. The 'effective number of codons' used in a gene. Gene. 1990;87(1):23–9.
67. Novembre JA. Accounting for background nucleotide composition when measuring codon usage bias. Mol Biol Evol. 2002;19(8):1390–4.
68. Sueoka N. Translation-coupled violation of Parity Rule 2 in human genes is not the cause of heterogeneity of the DNA G + C content of third codon position. Gene. 1999;238(1):53–8.
69. Lobry JR, Sueoka N. Asymmetric directional mutation pressures in bacteria. Genome Biol. 2002;3(10):Research0058.
70. Arakawa K, Tomita M. The GC skew index: a measure of genomic compositional asymmetry and the degree of replicational selection. Evol Bioinform Online. 2007;3:159–68.
71. Zhang M, Tian M, Weng Z, Yang Y, Pan N, Shen S, Zhao H, Du H, Qu C, Yin N. Genome-Wide Identification Analysis of the 4-Coumarate: Coa Ligase (4CL) Gene Family in Brassica U's Triangle Species and Its Potential Role in the Accumulation of Flavonoids in *Brassica napus* L. Plants. 2025;14(5):714.
72. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. Mol Biol Evol. 2018;35(6):1547–9.
73. 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 Proteom Bioinf. 2010;8(1):77–80.
74. Zheng D, Zou L, Zou J, Li Q, Lu S. Refining taxonomic identification of microalgae through molecular and genetic evolution: a case study of Prorocentrum lima and Prorocentrum arenarium. Microbiol Spectr. 2024;12(5):e0236723.
75. Yu X, Liu J, Li H, Liu B, Zhao B, Ning Z. Comprehensive Analysis of Synonymous Codon Usage Bias for Complete Genomes and E2 Gene of Atypical Porcine Pestivirus. Biochem Genet. 2021;59(3):799–812.
76. 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.
77. Raghavan R, Kelkar YD, Ochman H. A selective force favoring increased G + C content in bacterial genes. Proc Natl Acad Sci U S A. 2012;109(36):14504–7.
78. Zhong Y, Feng L, Deng H, Ji X, Zhang J, Sun Y, Lin P, Qiao Y, Xie S, Wang H, et al. Genomic resequencing reveals genetic diversity, population structure, and core collection of durian germplasm. Commun Biol. 2025;8(1):1273.
79. Li LF, Yang M, Qi Y, Gao PH, Yang SW, Zhao YT, Guo JW, Wei HY, Liu JN, Zhao JR, et al. Chloroplast genome of four *Amorphophallus* species: genomic features,comparative analysis, and phylogenetic relationships among Amorphophallus species. BMC Genomics. 2024;25(1):1122.
80. Wang Y, Jiang D, Guo K, Zhao L, Meng F, Xiao J, Niu Y, Sun Y. Comparative analysis of codon usage patterns in chloroplast genomes of ten *Epimedium* species. BMC Genom Data. 2023;24(1):3.
81. Barbhuiya PA, Uddin A, Chakraborty S. Analysis of compositional properties and codon usage bias of mitochondrial CYB gene in anura, urodela and gymnophiona. Gene. 2020;751:144762.
82. Xiong Q, Zheng L, Zhang Q, Li T, Zheng L, Song L. Comparative genomic insights into ecological adaptations and evolutionary dynamics of Trebouxiophyceae algae. BMC Genomics. 2025;26(1):764.
83. Hu X, Li Y, Meng F, Duan Y, Sun M, Yang S, Liu H. Analysis of chloroplast genome characteristics and codon usage bias in 14 species of Annonaceae. Funct Integr Genomics. 2024;24(3):109.
84. Ikemura T. Correlation between the abundance of Escherichia coli transfer RNAs and the occurrence of the respective codons in its protein genes: a proposal for a synonymous codon choice that is optimal for the E. coli translational system. J Mol Biol. 1981;151(3):389–409.
85. Hershberg R, Petrov DA. Selection on codon bias. Annu Rev Genet. 2008;42:287–99.
86. Shen L, Chen S, Liang M, Qu S, Feng S, Wang D, Wang G. Comparative analysis of codon usage bias in chloroplast genomes of ten medicinal species of Rutaceae. BMC Plant Biol. 2024;24(1):424.
87. Duan Y, Wang Y, Ding W, Wang C, Meng L, Meng J, Chen N, Liu Y, Xing S. Comparative and phylogenetic analysis of the chloroplast genomes of four commonly used medicinal cultivars of *Chrysanthemums morifolium*. BMC Plant Biol. 2024;24(1):992.
88. Huang H, Shi C, Liu Y, Mao SY, Gao LZ. Thirteen *Camellia* chloroplast genome sequences determined by high-throughput sequencing: genome structure and phylogenetic relationships. BMC Evol Biol. 2014;14:151.
89. Provan J, Powell W, Hollingsworth PM. Chloroplast microsatellites: new tools for studies in plant ecology and evolution. Trends Ecol Evol. 2001;16(3):142–7.
90. Melotto-Passarin DM, Tambarussi EV, Dressano K, De Martin VF, Carrer H. Characterization of chloroplast DNA microsatellites from Saccharum spp and related species. Genet Mol Res. 2011;10(3):2024–33.
91. Xu Y, Chen S, Chen S, Wei X, Shang H, Zhang Q, Zhang J. Genome-wide development of SSR molecular markers for modern sugarcane cultivars. Front Plant Sci. 2025;16:1573967.
92. Morris AB, Shaw J. Markers in time and space: A review of the last decade of plant phylogeographic approaches. Mol Ecol. 2018;27(10):2317–33.
93. Du C, Che Y, Zhang Z, He F, Zhang X, Han Y, Yang Z, Wang J, Qi T, Lan Y, et al. Identification and expression analysis of cinnamyl CoA reductase gene family and function of StCCR6 in potato (*Solanum tuberosum* L). Front Plant Sci. 2025;16:1638073.
94. Nei M, Gojobori T. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol Biol Evol. 1986;3(5):418–26.
95. Martín M, Sabater B. Plastid *ndh* genes in plant evolution. Plant Physiol Biochem. 2010;48(8):636–45.
96. Chris Blazier J, Guisinger MM, Jansen RK. Recent loss of plastid-encoded ndh genes within *Erodium* (Geraniaceae). Plant Mol Biol. 2011;76(3–5):263–72.

97. Ifuku K, Endo T, Shikanai T, Aro EM. Structure of the chloroplast NADH dehydrogenase-like complex: nomenclature for nuclear-encoded subunits. Plant Cell Physiol. 2011;52(9):1560–8.
98. Rousseau-Gueutin M, Huang X, Higginson E, Ayliffe M, Day A, Timmis JN. Potential functional replacement of the plastidic acetyl-CoA carboxylase subunit (accD) gene by recent transfers to the nucleus in some angiosperm lineages. Plant Physiol. 2013;161(4):1918–29.
99. Cheng X, Shi C, Yang T, Ge Q, Kress WJ, Liu X. Unveiling the conserved nature of *Heliconia* chloroplast genomes: insights from the assembly and analysis of four complete chloroplast genomes. Front Plant Sci. 2024;15:1535549.
100. Song F, Li T, Burgess KS, Feng Y, Ge XJ. Complete plastome sequencing resolves taxonomic relationships among species of *Calligonum* L. (Polygonaceae) in China. BMC Plant Biol. 2020;20(1):261.
101. Chen N, Sha LN, Dong ZZ, Tang C, Wang Y, Kang HY, Zhang HQ, Yan XB, Zhou YH, Fan X. Complete structure and variation of the chloroplast genome of *Agropyron cristatum* (L.) Gaertn. Gene. 2018;640:86–96.
102. Li Q-j, Liu Y, Wang A-h, Chen Q-f, Wang J-m, Peng L, Yang Y. Plastome comparison and phylogenomics of *Fagopyrum* (Polygonaceae): insights into sequence differences between *Fagopyrum* and its related taxa. BMC Plant Biol. 2022;22(1):339.
103. Walker JF, Walker-Hale N, Vargas OM, Larson DA, Stull GW. Characterizing gene tree conflict in plastome-inferred phylogenies. PeerJ. 2019;7:e7747.
104. Stegemann S, Keuthe M, Greiner S, Bock R. Horizontal transfer of chloroplast genomes between plant species. Proc Natl Acad Sci. 2012;109(7):2434–8.

Publisher's note