Journal of Anatolian Environmental and Animal Sciences
(*Anadolu Çevre ve Hayvancılık Bilimleri Dergisi*)
https://doi.org/10.35229/jaes.1891156

JAES
Year: 11, 2026 (1-9)
AÇEH
Yıl: 11, 2026 (1-9)

ARAŞTIRMA MAKALESİ **RESEARCH PAPER**

Genome-Wide Diversity and Selection in *Poaceae* Plastomes

Yasin KAYMAZ *

Ege University, Faculty of Engineering, Department of Bioengineering, Bornova 35040 İzmir, Türkiye

Received: *17.02.2026* ***Published:*** *16.03.2026*

How to cite: **Kaymaz, Y. (2026).** Genome-wide diversity and selection in *Poaceae* plastomes. *J. Anatolian Envion. Anim. Sci.*, ***11***, 1-9. https://doi.org/10.35229/jaes.1891156

Atıf yapmak için: **Kaymaz, Y. (2026).** Poaceae plastomlarında genom çapında çeşitlilik ve seçilim. *Anadolu Çev. Hay. Bil. Derg.*, ***11***, 1-9. https://doi.org/10.35229/jaes.1891156

Abstract: Plastome provides a widely used genomic record for resolving phylogenetic relationships and locating informative variation for marker development, particularly in large, taxonomically complex plant groups such as *Poaceae*. We assembled a comparative dataset of 175 complete RefSeq plastomes representing broad phylogenetic and ecological diversity within grasses. Whole-plastome sequences were aligned and analyzed for genome-wide nucleotide diversity, Shannon entropy, gene-level variability across core coding loci, phylogenomic structure using maximum likelihood, and episodic diversifying selection across core protein-coding genes.

The sequence alignment contained extensive phylogenetic signal and showed moderate genome-wide divergence (mean π = 0.0507) distributed heterogeneously across the plastome. Sliding-window profiles identified discrete hypervariable regions, with prominent peaks near ~13.6–15.9 kb and ~105.6–107.2 kb in the reference coordinate system. Diversity was strongly compartmentalized in the single-copy regions but markedly reduced across both inverted repeats (IRb/IRa). Gene-level mapping revealed that the highest coding-sequence variability was concentrated in photosystem II genes, especially *psbI*, *psbD*, and *psbK*, whereas genes involved in transcription and translation, such as *matK* and *rps16*, were comparatively conserved. Codon-based tests detected significant evidence of gene-wide episodic diversifying selection in 13 plastid genes, led by *rpoC2* and *rpoA*, with additional signals spanning ATP synthase, PSI/PSII, cytochrome b_6f, NDH, and envelope-associated genes. Episodic selection signals in transcriptional and photosynthetic genes suggest lineage-specific adaptive episodes superimposed on pervasive purifying selection. Together, these results provide a robust phylogenomic framework, a genome-scale diversity atlas, and a prioritized set of candidate loci for barcoding and testing evolutionary hypotheses in grasses.

Keywords: Genome-wide diversity, plastome, phylogenomics, photosystem II, *Poaceae.*

***Corresponding author's:**
Yasin KAYMAZ
Ege University, Faculty of Engineering, Department of Bioengineering, Bornova 35040 İzmir, Türkiye.
✉: yasin.kaymaz@ege.edu.tr

Poaceae plastomlarında genom çapında çeşitlilik ve seçilim

Öz: Plastom, filogenetik ilişkilerin çözümlenmesi ve özellikle Poaceae gibi büyük ve taksonomik açıdan karmaşık bitki gruplarında belirteç geliştirme için bilgilendirici varyasyonların belirlenmesi amacıyla yaygın olarak kullanılan bir genomik kayıt sağlar. Çimler içinde geniş filogenetik ve ekolojik çeşitliliği temsil eden 175 tam RefSeq plastomundan oluşan karşılaştırmalı bir veri seti oluşturulmuştur. Tüm plastom dizileri hizalanarak ve genom çapında nükleotid çeşitliliği, Shannon entropisi, çekirdek kodlayıcı lokuslar boyunca gen düzeyinde değişkenlik, maksimum olabilirlik kullanılarak filogenomik yapı ve çekirdek protein kodlayan genler boyunca epizodik çeşitlendirici seçilim açısından analiz edilmiştir.

Dizi hizalaması kapsamlı bir filogenetik sinyal içerdi ve plastom boyunca heterojen biçimde dağılmış orta düzeyde genom çapında farklılaşma göstermiştir (ortalama π = 0.0507). Kayan pencere profilleri, referans koordinat sisteminde yaklaşık 13.6–15.9 kb ve 105.6–107.2 kb civarında belirgin zirvelerle ayrık hiper-değişken bölgeleri ortaya koymuştur. Çeşitlilik, tek kopyalı bölgelerde güçlü biçimde bölümlenmiş; ancak her iki ters tekrar bölgesi (IRb/IRa) boyunca belirgin biçimde azalmıştır. Gen düzeyindeki haritalama, en yüksek kodlayıcı dizi değişkenliğinin özellikle psbI, psbD ve psbK olmak üzere fotosistem II genlerinde yoğunlaştığını; buna karşılık matK ve rps16 gibi transkripsiyon ve translasyonda görev alan genlerin görece korunmuş olduğunu göstermiştir. Kodon temelli testler, rpoC2 ve rpoA başta olmak üzere 13 plastid geninde gen-genelinde epizodik çeşitlendirici seçilime ilişkin anlamlı kanıt saptamış; ayrıca ATP sentaz, PSI/PSII, sitokrom b6f, NDH ve zarfla ilişkili genleri kapsayan ek sinyaller de belirlenmiştir. Transkripsiyonel ve fotosentetik genlerdeki epizodik seçilim sinyalleri, yaygın arındırıcı seçilim üzerine binmiş soy-hattına özgü adaptif evrim dönemlerine işaret etmektedir. Birlikte değerlendirildiğinde bu sonuçlar, sağlam bir filogenomik çerçeve, genom ölçekli bir çeşitlilik atlası ve çimlerde barkodlama ile evrimsel hipotezlerin test edilmesi için önceliklendirilmiş aday lokus seti sunmaktadır.

Anahtar kelimeler: Genom çapında çeşitlilik, plastom, filogenomik, fotosistem II, *Poaceae.*

***Sorumlu yazar:**
Yasin KAYMAZ
Ege Üniversitesi
Mühendislik Fakültesi, Biyomühendislik Böl.
Bornova 35040 İzmir, Türkiye
✉: yasin.kaymaz@ege.edu.tr

INTRODUCTION

Poaceae (the grass family) underpins much of Earth's terrestrial productivity and human society: it includes the world's dominant staple crops (rice, wheat, maize), major forage species that sustain livestock, and many of the grasses that structure natural ecosystems from savannas to alpine meadows (Çatal, 2025; Linder, Lehmann, Archibald, Osborne, & Richardson, 2018). Because many grasses are morphologically similar and hybridization can blur species boundaries, their chloroplast genomes (plastomes), typically small, conserved, and maternally inherited in most angiosperms, are widely examined as a complementary molecular record for identifying species, clarifying evolutionary relationships, and tracking biogeographic history (Corriveau & Coleman, 1988; Daniell, Lin, Yu, & Chang, 2016; Shaw et al., 2014). Plastome data also help locate informative variation (e.g., SNPs, indels, and repeats) for developing DNA barcodes and population markers, and can illuminate photosynthesis-related evolution in grasses, including lineages associated with C4 origins and diversification. In short, plastome provides a practical, genome-scale toolkit that strengthens *Poaceae* systematics, conservation decisions, and crop/forage research by linking genetic evidence to taxonomy, ecology, and agronomic traits (Group1 et al., 2009; Kress & Erickson, 2007).

Plastomes are foundational resources for plant systematics, phylogeography, and evolutionary genomics due to their relatively conserved gene content, predominantly uniparental inheritance, and manageable genome size (Daniell et al., 2016). Despite broad conservation in structure, typically a quadripartite organization comprising large and small single-copy regions (LSC and SSC) separated by a pair of inverted repeats (IRs), plastomes exhibit substantial heterogeneity in substitution rate and indel dynamics across loci and lineages (Jansen & Ruhlman, 2012). This heterogeneity impacts phylogenetic resolution, marker choice for species identification, and inference of molecular adaptation (Shaw et al., 2014).

Large multi-species plastome datasets enable questions that are difficult to address with smaller sampling: first, how nucleotide diversity is distributed along the plastome at the genome scale, secondly, whether hypervariable regions cluster near structural boundaries and intergenic intervals, and third, how selective constraint varies among functional gene categories (photosynthesis, transcription/translation, and housekeeping) (Daniell et al., 2016; Shaw et al., 2014). While many studies report plastome trees and descriptive genome features, fewer integrate genome-wide variability landscapes with codon-based selection models across large taxon sets (Duvall, Burke, & Clark, 2019; Saarela et al., 2018).

Here, we analyze 175 complete plastomes using a whole-plastome alignment combined with gene-resolved codon alignments derived from GenBank annotations. We infer a maximum-likelihood phylogenomic backbone; quantify plastome-wide variability using nucleotide diversity and entropy to identify statistically supported diversity hotspots; and evaluate selective pressures on core protein-coding genes using dN/dS-based and episodic-selection frameworks (Ben Murrell et al., 2012). This integrative approach yields a high-resolution atlas of plastome evolution and provides practical candidate regions for downstream marker development.

MATERIAL AND METHOD

Dataset acquisition and metadata: We analyzed 175 complete plastomes identified by NCBI RefSeq/GenBank accessions (Table 1). GenBank records were downloaded via NCBI Entrez using the efetch utility.

Whole-plastome alignment and occupancy filtering: Genomes were aligned using MAFFT (v7.526) (Katoh, Misawa, Kuma, & Miyata, 2002). We computed alignment quality metrics (gap fraction per genome, ungapped length distribution). To reduce artifacts from sparse columns, we retained usable sites defined as alignment columns with ≥50% non-gap occupancy.

Genome-wide diversity and hotspot detection: For each retained site, we computed nucleotide diversity (π) and Shannon entropy using A/C/G/T-only characters. Sliding-window means (window 1000 bp; step 500 bp) were calculated across the retained coordinate space, and windows were mapped onto a reference genome coordinate system by removing gaps from the first sequence in the alignment (NC_058870.1). The top-ranked windows were reported as candidate hypervariable regions.

Phylogenetic inference: Maximum-likelihood phylogenies were inferred with IQ-TREE3 (v3.0.1) using ModelFinder for substitution model selection and support via 1000 ultrafast bootstrap replicates and 1000 SH-aLRT replicates (Wong et al., 2025). Analyses were performed on both the full alignment and an optional trimmed alignment generated by trimAl (v1.5.1) (gap threshold 0.5) to assess robustness (Capella-Gutiérrez, Silla-Martínez, & Gabaldón, 2009). The analysis used automatic parallelization with a maximum of 56 threads.

Gene extraction and codon alignments: Protein-coding sequences were extracted from GenBank annotations (CDS features) in GFF format. Genes were designated core if present in ≥90% of genomes. For each core gene, amino acid sequences were aligned with MAFFT (v7.526) (Katoh et al., 2002) and back-translated to codon alignments, preserving reading frame and enabling codon-based models.

Selection analyses: We performed gene-level selection inference using HyPhy (v2.5.93) (Pond, Frost, &

Muse, 2004). BUSTED was used to test for gene-wide episodic selection on at least one branch (Murrell et al., 2015); MEME was used to identify sites under episodic diversifying selection (B. Murrell et al., 2012). For each gene, test statistics and p-values were summarized into tables for multiple-testing-aware interpretation.

Table 1. The list of species and their accession IDs included in this study.

No:	Accession	Species	No:	Accession	Species
1	NC_058870.1	*Achnatherum pekinense*	89	NC_039983.1	*Froesiochloa boutelouoides*
2	NC_015820.1	*Acidosasa purpurea*	90	NC_024718.1	*Gaoligongshania megalothyrsa*
3	NC_068107.1	*Acrachne racemosa*	91	NC_035051.1	*Garnotia tenella*
4	NC_024831.1	*Aegilops bicornis*	92	NC_036686.1	*Gastridium ventricosum*
5	NC_042858.1	*Aeluropus lagopoides*	93	NC_062076.1	*Gelidocalamus kunishii*
6	NC_059819.1	*Agenium leptocladum*	94	NC_072533.1	*Gelidocalamus zixingensis*
7	NC_037162.1	*Agrostis gigantea*	95	NC_035046.1	*Germainia capitata*
8	NC_035517.1	*Alloeochaete namuliensis*	96	NC_050765.1	*Gigantochloa albociliata*
9	NC_047228.1	*Alopecurus aequalis*	97	NC_050406.1	*Glyceria arkansana*
10	NC_043932.1	*Ampelocalamus melicoideus*	98	NC_029749.1	*Guadua angustifolia*
11	NC_030619.1	*Amphicarpum muhlenbergianum*	99	NC_061343.1	*Guaduella macrostachys*
12	NC_035520.1	*Amphipogon caricinus*	100	NC_036687.1	*Gynerium sagittatum*
13	NC_087655.1	*Anadelphia scyphofera*	101	NC_036688.1	*Halopyrum mucronatum*
14	NC_086835.1	*Anatherum africanum*	102	NC_035027.1	*Heteropogon contortus*
15	NC_035030.1	*Andropogon abyssinicus*	103	NC_036124.1	*Hildaea pallens*
16	NC_035010.1	*Andropogon fastigiatus*	104	NC_043941.1	*Himalayacalamus collaris*
17	NC_014062.1	*Anomochloa marantoidea*	105	NC_044487.1	*Hitchcockella baronii*
18	NC_060728.1	*Aristida adscensionis*	106	NC_036689.1	*Holcus lanatus*
19	NC_035048.1	*Arthraxon hispidus*	107	NC_036125.1	*Homolepis aturensis*
20	NC_023934.1	*Arundinaria appalachiana*	108	NC_056985.1	*Hordeum vulgare*
21	NC_030620.1	*Arundinella deppeana*	109	NC_036691.1	*Humbertochloa bambusiuscula*
22	NC_037077.1	*Arundo donax*	110	NC_058302.1	*Hygroryza aristata*
23	NC_042839.1	*Astrebla lappacea*	111	NC_087669.1	*Hyparrhenia anamesa*
24	NC_043840.1	*Australopyrum retrofractum*	112	NC_030487.1	*Imperata cylindrica*
25	NC_044171.1	*Avena strigosa*	113	NC_079918.1	*Indocalamus emeiensis*
26	NC_046490.1	*Axonopus compressus*	114	NC_067633.1	*Indosasa crassiflora*
27	NC_050773.1	*Bambusa basihirsuta*	115	NC_035530.1	*Isachne albens*
28	NC_085343.1	*Bonia levigata*	116	NC_030488.1	*Ischaemum afrum*
29	NC_030621.1	*Bothriochloa alta*	117	NC_059831.1	*Iseilema anthephoroides*
30	NC_068108.1	*Bouteloua dactyloides*	118	NC_085716.1	*Kampochloa brachyphylla*
31	NC_033879.1	*Brachiaria fragrans*	119	NC_058788.1	*Kengyilia grandiglumis*
32	NC_036836.1	*Brachypodium hybridum*	120	NC_035009.1	*Kerriochloa siamensis*
33	NC_067043.1	*Briza media*	121	NC_072345.1	*Koeleria macrantha*
34	NC_054212.1	*Bromus catharticus*	122	NC_037168.1	*Lamarckia aurea*
35	NC_067046.1	*Calamagrostis epigeios*	123	NC_036123.1	*Lasiacis nigra*
36	NC_066044.1	*Campeiostachys calcicola*	124	NC_024106.1	*Lecomtella madagascariensis*
37	NC_030622.1	*Capillipedium venustum*	125	NC_034766.1	*Leersia japonica*
38	NC_036711.1	*Catapodium rigidum*	126	NC_058993.1	*Leptagrostis schimperiana*
39	NC_057588.1	*Cenchrus centrasiaticus*	127	NC_033863.1	*Leptaspis banksii*
40	NC_036712.1	*Chaetium bromoides*	128	NC_035532.1	*Nematopoa longipes*
41	NC_031299.1	*Chasechloa egregia*	129	NC_031333.1	*Oryza sativa*
42	NC_035522.1	*Chasmanthium laxum*	130	NC_030494.1	*Paspalidium geminatum*
43	NC_027184.1	*Chikusichloa aquatica*	131	NC_030777.1	*Saccharum arundinaceum*
44	NC_053872.1	*Chimonobambusa hejiangensis*	132	NC_036118.1	*Saccharum hildebrandtii*
45	NC_024714.1	*Chimonocalamus longiusculus*	133	NC_087642.1	*Schizachyrium nodulosum*
46	NC_032033.1	*Chloris truncata*	134	NC_068071.1	*Schizostachyum auriculatum*
47	NC_046489.1	*Chrysopogon aciculatus*	135	NC_061348.1	*Scrotochloa urceolata*
48	NC_042672.1	*Chusquea culeou*	136	NC_021761.1	*Secale cereale*
49	NC_060390.1	*Cleistogenes caespitosa*	137	NC_082098.1	*Sinobambusa rubroligula*
50	NC_035732.1	*Coelachne africana*	138	NC_008602.1	*Sorghum bicolor*
51	NC_062353.1	*Coleanthus subtilis*	139	NC_064989.1	*Spodiopogon sagittifolius*
52	NC_037165.1	*Connorochloa tenuis*	140	NC_068109.1	*Sporobolus aculeatus*
53	NC_035523.1	*Crinipes abyssinicus*	141	NC_030499.1	*Steinchisma laxum*
54	NC_085717.1	*Ctenium aromaticum*	142	NC_036704.1	*Stenotaphrum secundatum*
55	NC_042144.1	*Cymbopogon citratus*	143	NC_072322.1	*Stipa aliena*
56	NC_034680.1	*Cynodon dactylon*	144	NC_036112.1	*Stipagrostis hirtigluma*
57	NC_036714.1	*Dactyloctenium aegyptium*	145	NC_062958.1	*Stipellula capensis*
58	NC_030502.1	*Danthoniopsis dinteri*	146	NC_033862.1	*Streptochaeta spicata*
59	NC_050762.1	*Dendrocalamus bambusoides*	147	NC_036683.1	*Streptogyna americana*
60	NC_080989.1	*Deschampsia nubigena*	148	NC_036116.1	*Streptostachys asperifolia*
61	NC_063471.1	*Desmostachya bipinnata*	149	NC_035535.1	*Styppeiochloa gynoglossa*
62	NC_035525.1	*Dichaetaria wightii*	150	NC_037160.1	*Taeniatherum caput-medusae*
63	NC_030623.1	*Dichanthelium acuminatum*	151	NC_085342.1	*Temochloa liliana*
64	NC_087665.1	*Diheteropogon amplectens*	152	NC_024724.1	*Thamnocalamus spathiflorus*
65	NC_035020.1	*Dimeria ornithopoda*	153	NC_059838.1	*Themeda anathera*
66	NC_062405.1	*Dinebra chinensis*	154	NC_043837.1	*Thinopyrum bessarabicum*
67	NC_068112.1	*Diplachne fusca*	155	NC_030616.1	*Thyridolepis xerophila*
68	NC_035526.1	*Dregeochloa pumilla*	156	NC_042830.1	*Tragus australianus*
69	NC_037167.1	*Drepanostachyum falcatum*	157	NC_036705.1	*Tribolium hispidum*
70	NC_032383.1	*Echinochloa colona*	158	NC_036706.1	*Tricholaena monachne*
71	NC_036715.1	*Ehrharta bulbosa*	159	NC_067610.1	*Trichoneura ciliata*
72	NC_056927.1	*Eleusine coracana*	160	NC_042860.1	*Triodia basedowii*
73	NC_087668.1	*Elymandra archaelymandra*	161	NC_060399.1	*Tripogon bromoides*
74	NC_061050.1	*Elymus atratus*	162	NC_037087.1	*Tripsacum dactyloides*
75	NC_035527.1	*Elytrophorus globularis*	163	NC_036115.1	*Tristachya humbertii*
76	NC_042837.1	*Enneapogon caerulescens*	164	NC_002762.1	*Triticum aestivum*
77	NC_068113.1	*Enteropogon dolichostachyus*	165	NC_036709.1	*Uniola paniculata*
78	NC_036126.1	*Entolasia imbricata*	166	NC_030067.1	*Urochloa brizantha*
79	NC_059727.1	*Eragrostis atrovirens*	167	NC_042238.1	*Urochondra setulosa*
80	NC_035028.1	*Eremochloa ciliaris*	168	NC_036710.1	*Vaseyochloa multinervosa*
81	NC_059826.1	*Eremopogon foveolatus*	169	NC_030618.1	*Whiteochloa capillipes*
82	NC_041557.1	*Eriachne agrostidea*	170	NC_043894.1	*Yushania brevipaniculata*
83	NC_030624.1	*Eriochloa meyeriana*	171	NC_001666.2	*Zea mays*
84	NC_029883.1	*Eriochrysis laxa*	172	NC_030500.1	*Zeugites pittieri*
85	NC_035031.1	*Eulalia siamensis*	173	NC_037170.1	Zingeria biebersteiniana
86	NC_036685.1	*Eustachys glauca*	174	NC_029401.1	*Zizania latifolia*
87	NC_043891.1	*Fargesia albocerea*	175	NC_053873.1	*Zoysia matrella*
88	NC_043937.1	*Fargesia fungosa*			

RESULTS

The plastome dataset was assembled to capture broad phylogenetic and ecological representation within the focal plant group while maintaining consistent data quality for comparative analyses. We restricted sampling to records with RefSeq accessions, which provides a standardized, curated set of complete plastome assemblies suitable for downstream alignment-based inference. The final panel includes both economically important taxa (major crop lineages and their relatives) and wild species spanning multiple genera, enabling assessment of genome-wide diversity patterns (LSC/SSC vs. IR compartments), gene-level variability in core coding loci, and evolutionary signal across a wide taxonomic breadth. This strategy balances comparability (uniform genome type and annotation quality) with diversity (broad lineage coverage), increasing power to detect conserved versus rapidly evolving regions and to generalize patterns of plastome variation across the sampled taxa.

The dataset contained 175 plastomes aligned to 356,558 columns. The mean ungapped sequence length was ~137.8 kb (range ~114.8-152.9 kb), consistent with typical plastome sizes. After filtering for ≥50% occupancy, 138,148 sites were retained for diversity analyses. Among these retained sites, 84,164 were variable, indicating substantial phylogenetic signal across the alignment.

Genome-wide diversity landscape and hypervariable regions: Mean nucleotide diversity across retained sites was π = 0.0507, with strong spatial heterogeneity. Sliding-window analysis identified discrete hypervariable regions. The highest-diversity windows mapped (in the reference coordinate system of NC_058870.1) span approximately 13.6-15.9 kb and 105.6-107.2 kb, with additional peaks distributed across early- and mid-genome segments. These regions represent candidate loci for marker development and may coincide with rapidly evolving intergenic spacers and boundary-adjacent genes in plastomes.

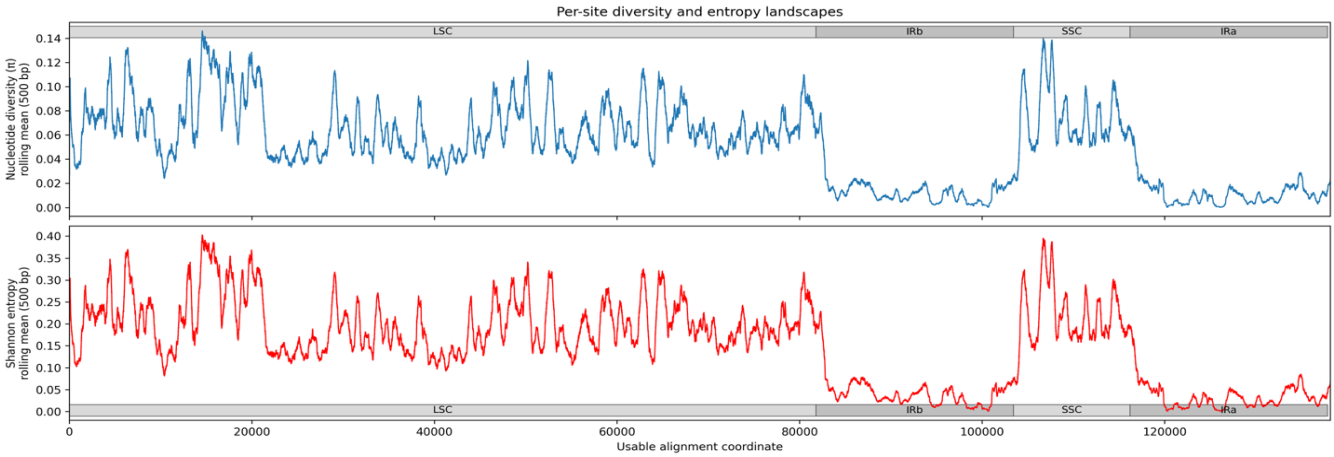

Figure 1. Per-site nucleotide diversity and entropy across the plastome alignment. The top panel shows nucleotide diversity (π) calculated per alignment column and smoothed with a 500-bp rolling mean. The bottom panel shows per-site Shannon entropy (red), also smoothed with a 500-bp rolling mean. The x-axis represents the usable alignment coordinate (i.e., the columns retained after filtering). Gray horizontal bars indicate the approximate boundaries of major plastome regions-large single-copy (LSC), inverted repeat B (IRb), small single-copy (SSC), and inverted repeat A (IRa)-mapped onto the alignment coordinate system.

Genome-wide sliding-window profiles of nucleotide diversity (π) and Shannon entropy revealed a strongly structured landscape of variation across the plastome alignment, with pronounced differences among the major plastome compartments (**Figure 1**). Both metrics were generally elevated through portions of the single-copy regions (LSC and SSC), where divergence peaks formed discrete hotspots separated by long conserved intervals, indicating that polymorphism is concentrated at a subset of alignment positions rather than evenly distributed across the genome. In contrast, π and entropy dropped markedly across both inverted repeats (IRb and IRa), producing the lowest and flattest variability baselines in the plastome. This reduction is biologically expected because the two IR copies are maintained in near identity by intramolecular recombination and gene conversion, which homogenize sequence differences between repeats and slow the accumulation of substitutions. In addition, the IR is enriched for highly conserved rRNA genes and essential housekeeping loci, which are typically under strong purifying selection and further contribute to reduced variability. The combination of a low-variation IR background and higher, punctuated variation in the LSC/SSC supports a model in which plastome evolution is compartmentalized: repeat-mediated homogenization and functional constraint dampen diversity in the IR, whereas the single-copy regions tolerate more lineage-specific substitutions and indels, yielding localized peaks that are candidate markers for population discrimination and phylogenetic signal.

Gene-level mapping of diversity showed that the strongest signals of coding-sequence variation were concentrated in photosystem II (PSII) genes, particularly those encoding core and stabilizing PSII subunits (Table 2). The highest mean π and entropy occurred in *psbI* (mean π = 0.0867), *psbD* (0.0850), and *psbK* (0.0766), all of which encode PSII membrane components that contribute to reaction-center integrity, complex stability, and assembly/repair. Specifically, *psbD* encodes the D2 reaction-center core protein, which pairs with D1 (*PsbA*) to form the PSII reaction center and coordinates key cofactors for electron transfer; elevated variation in *psbD* (and to a lesser extent *psbA*, the D1 protein) is consistent with divergence accumulating at a subset of codons while preserving essential photochemical function. *psbC* (encoding CP43, a core chlorophyll-binding antenna protein) showed moderate mean variability but high maxima, indicating localized hotspots within an otherwise conserved light-harvesting core. Several small PSII subunits (*psbM*, *psbZ*/ycf9) exhibited low-to-moderate

mean diversity yet occasional high-variability sites, consistent with their roles in maintaining PSII architecture, dimerization, and supercomplex organization, where limited residues may tolerate change without disrupting overall structure. In contrast, genes primarily involved in plastome gene expression, *matK* (a group II intron maturase required for splicing of multiple plastid introns) and *rps16* (a small ribosomal subunit protein important for plastid translation), were comparatively conserved in mean π/entropy, supporting stronger purifying constraints across most sites. Overall, the concordant ranking of π and entropy across genes indicates that the principal coding diversity in this dataset is driven by PSII-related loci, with variation occurring mainly as site-specific peaks rather than uniformly elevated divergence across entire genes.

Table 2. Gene-level summaries of nucleotide diversity and Shannon entropy across core plastome protein-coding genes. Per-site π and Shannon entropy were calculated along the multiple-sequence alignment and mapped to coding sequences (CDS) using the annotated reference plastome. For each gene, values were summarized across all mapped alignment sites within the CDS. Reference CDS length (bp) indicates the CDS length in the reference annotation. Only genes with ≥50 mapped sites were retained.

Gene	Number of sites	π (mean)	π (median)	π (max)	Entropy (mean)	Entropy (median)	Entropy (max)	Reference CDS length (bp)
psbI	111	0.0867	0.0454	0.6558	0.2496	0.1774	1.6732	111
psbD	1062	0.085	0.0351	0.6868	0.2505	0.1422	1.7461	1062
psbK	186	0.0766	0.0345	0.587	0.2273	0.1277	1.4035	186
psbC	1422	0.0394	0.0115	0.607	0.1144	0.0513	1.5522	1422
psbM	198	0.0308	0	0.4832	0.0857	0	1.0094	198
psbA	1062	0.012	0	0.4531	0.0408	0	0.9515	1062
matK	1536	0.0107	0	0.4883	0.0346	0	0.9831	1536
rps16	279	0.0062	0	0.4786	0.0167	0	0.9689	279
psbZ	189	0.0015	0	0.2155	0.0044	0	0.5374	189

Phylogenomic inference: Maximum-likelihood phylogenies inferred from the whole-plastome alignment produced a strongly supported backbone. The tree provides a robust framework for downstream comparative analyses, including lineage-specific rate and selection tests.

In the maximum-likelihood phylogeny inferred from the whole-plastome alignment, the topology broadly matches the accepted deep structure of Poaceae and is strongly supported at major nodes (generally ≥95–100% support). Basal positions are occupied by early-diverging grasses (*Anomochloa marantoidea*, then *Streptochaeta spicata*, followed by *Leptaspis banksii* and *Guaduella macrostachys*), consistent with their status as lineages that split off before the main radiation of core grasses. The remaining taxa separate into the two principal radiations, a BOP clade (*Bambusoideae-Oryzoideae-Pooideae*; 100% support) and a PACMAD clade (100% support). Within BOP, a well-resolved *Pooideae* assemblage groups cool-season, largely C3 temperate grasses and cereals, including a *Triticeae*-dominated cluster (*Triticum*, *Hordeum*, *Secale*, *Aegilops*, *Thinopyrum*, *Kengyilia*) together with other pooids (e.g., *Brachypodium*, *Bromus*, *Avena*, *Stipa*), while an *Oryzoideae* clade unites wetland-associated lineages (*Oryza, Leersia*, *Hygroryza*, *Zizania*, *Chikusichloa*). The Bambusoideae component forms a coherent bamboo grouping spanning temperate bamboos (*Fargesia*, *Yushania*, *Gelidocalamus*, *Indocalamus*, *Thamnocalamus*) and tropical woody bamboos (*Bambusa*, *Dendrocalamus*, *Gigantochloa*, *Guadua*), reflecting shared plastome history within bamboos. Within PACMAD, the earliest split places an *Aristidoideae* pair (*Aristida adscensionis* and *Stipagrostis hirtigluma*) as sister to the rest, and the remaining taxa partition into a strongly supported *Chloridoideae*-dominated clade (arid/saline and open-habitat grasses such as *Cynodon*, *Eleusine*, *Chloris*, *Zoysia*, *Sporobolus*, *Eragrostis*, *Triodia*) versus a *Panicoideae*-centered clade. The *Panicoideae*-centered portion further resolves an *Andropogoneae*/grassland-dominant C4 assemblage (*Zea*, *Tripsacum*, *Sorghum*, *Saccharum*, *Imperata*, *Andropogon*, *Themeda*, *Cymbopogon*, *Hyparrhenia*) and a distinct *Paniceae*/*Paspaleae* grouping (*Urochloa*, *Brachiaria*, *Echinochloa*, *Cenchrus*, *Paspalidium*, *Stenotaphrum*), together capturing major ecological transitions in grasses (temperate C3 vs warm-season C4 lineages, and forest/woody bamboo diversification). As these relationships are inferred from maternally inherited plastomes, they primarily reflect plastid lineage history and can differ from the species tree in groups prone to hybridization, but the recovered deep clades provide a biologically coherent framework for interpreting plastome evolution across the sampled *Poaceae*.

Gene content and codon-based selection: Extraction from GenBank annotations yielded a set of core protein-coding genes present in ≥90% of genomes. Gene-wide episodic selection was assessed across core plastid protein-coding genes using BUSTED with synonymous rate variation. After Benjamini-Hochberg correction, 13 genes showed significant evidence of episodic diversifying selection ($q \leq 0.05$) (**Figure 3**). The strongest signal was detected in *rpoC2* ($p = 4.99 \times 10^{-11}$; $q = 1.50 \times 10^{-9}$) and *rpoA* ($p = 2.24 \times 10^{-6}$; $q = 3.36 \times 10^{-5}$), followed by additional significant loci spanning the ATP synthase, photosystem I/II, cytochrome b_6f, NDH complex, and envelope-associated genes (*atpH*, *rbcL*, *psbI*, *ndhD*, *psbB*, *cemA*, *psbD*, *psaA*, *ndhJ*, *petA*, *ndhG*; $q = 1.0 \times 10^{-4}$ - 3.5×10^{-2}). In contrast, most ribosomal protein genes and several photosystem-associated loci did not show gene-wide evidence for episodic selection ($q \approx 0.3$-0.5), consistent with pervasive purifying constraint across the plastome.

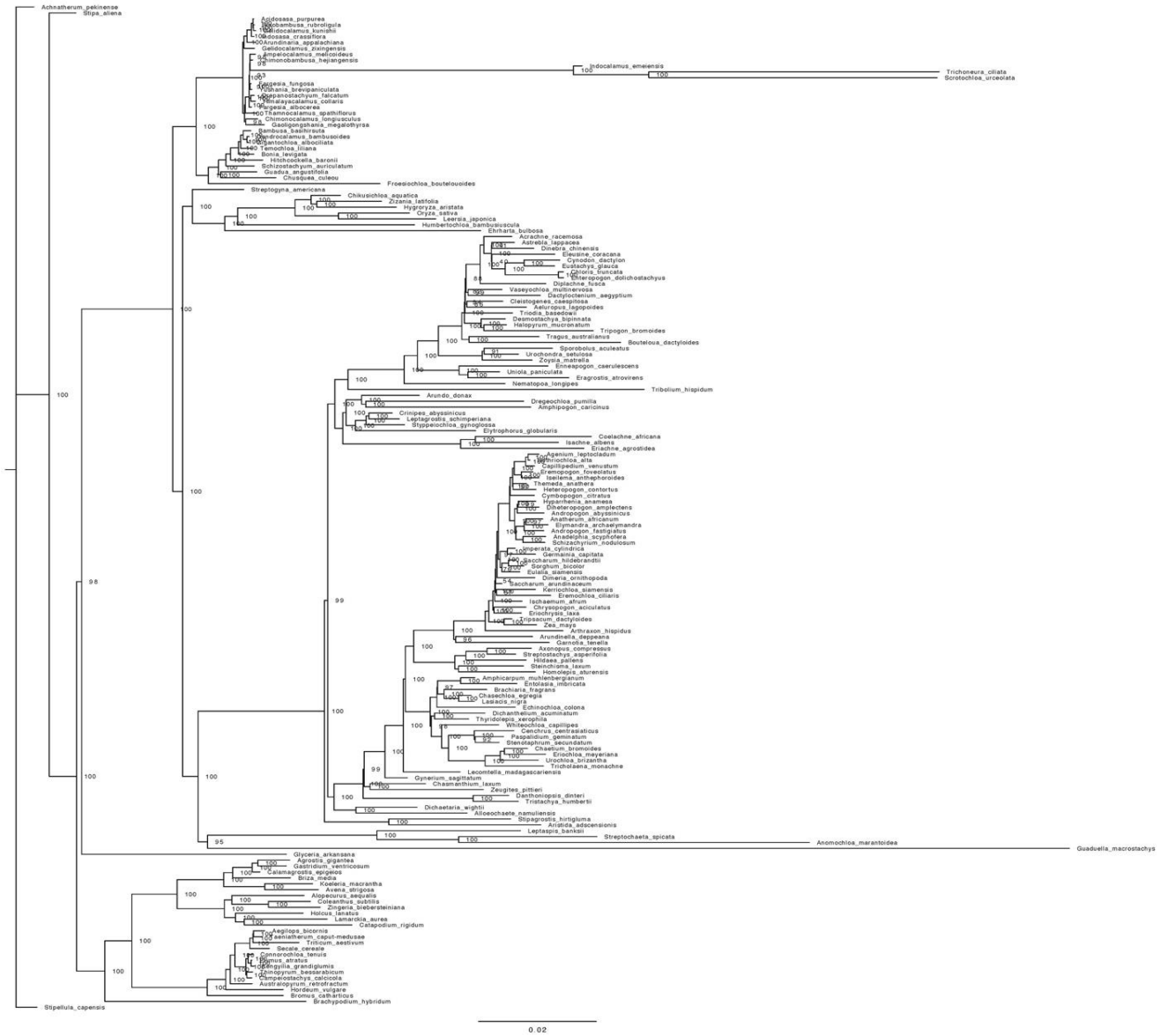

Figure 2. Maximum-likelihood phylogeny of *Poaceae* inferred from whole plastomes. The tree was reconstructed using the complete plastome alignment of the sampled grass species (tips labeled by species name), and branch lengths are proportional to the estimated number of substitutions per site. Node values indicate ultrafast bootstrap support (UFBoot; %), with major nodes showing strong support. The topology resolves the principal grass radiations, including early-diverging lineages (e.g., *Anomochloa*, *Streptochaeta*, *Leptaspis*) and the two major clades of core *Poaceae*, BOP (*Bambusoideae-Oryzoideae-Pooideae*) and PACMAD, within which major subfamily-level groupings (e.g., *Bambusoideae*, *Oryzoideae*, *Pooideae*, *Chloridoideae*, *Panicoideae*/*Andropogoneae*) are recovered.

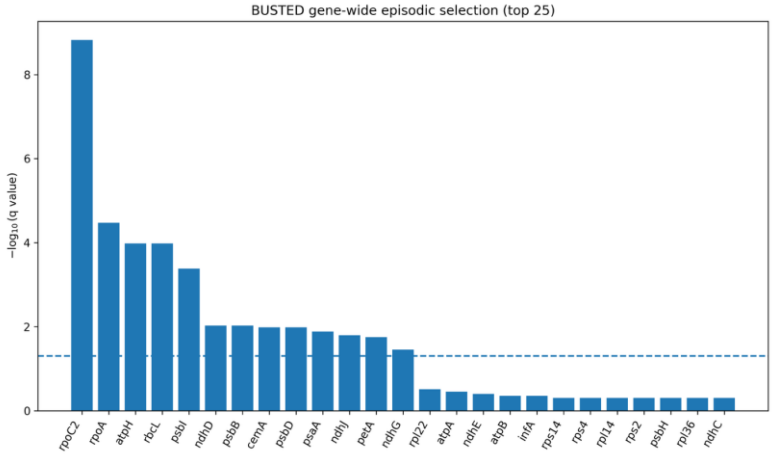

Figure 3. Gene-wide episodic selection across plastome protein-coding genes (BUSTED). Bar plot shows the strength of evidence for episodic diversifying selection inferred with the BUSTED model for the top-ranked genes, expressed as −log10(q-value) after Benjamini–Hochberg false discovery rate (FDR) correction. Genes are ordered from most to least significant (lowest to highest q-value). The dashed horizontal line marks the significance threshold (q = 0.05). Genes above the line indicate loci with significant gene-wide evidence of episodic selection under the branch-site model, whereas genes below the line are not significant after multiple testing correction.

DISCUSSION

By combining dense plastome sampling with genome-wide diversity mapping, phylogenomics, and codon-based tests of selection, this study provides an integrated view of plastome evolution across *Poaceae*. Large-scale plastome phylogenomics has repeatedly shown that whole plastomes can strongly stabilize deep grass relationships relative to sparse-locus datasets, while enabling finer-scale comparisons of rate and character distribution across the genome (Saarela et al., 2018). Our broad taxon coverage and standardized RefSeq-only filtering extend this comparative framework and provide a stable backbone suitable for downstream molecular-evolutionary inference, while also motivating caution because plastome-wide analyses can be sensitive to alignment properties (e.g., gap treatment) in very large matrices (Orton et al., 2021).

In our study, the central result is the strong spatial heterogeneity of plastome variation, as the diversity is concentrated into discrete hotspots separated by long conserved intervals. This pattern matches the broader observation that plastome variability often clusters into a limited set of rapidly evolving regions rather than being evenly distributed across the genome (Li, Kuo, Pryer, & Rothfels, 2016). These windows, therefore, provide a practical shortlist for marker development and targeted follow-up, especially when primer design and taxon-specific validation are prioritized across multiple *Poaceae* subfamilies.

Diversity is also strongly compartmentalized among plastome regions, as both nucleotide diversity and entropy metrics drop sharply across inverted repeats relative to the single-copy regions. Reduced substitution rates in the IR compared with LSC/SSC have been documented across land plants, and natural experiments in which genes move into the IR show rate deceleration consistent with the structural and homogenizing effects of the repeat (Li et al., 2016). Recent experimental work removing one IR copy in tobacco further demonstrates functional consequences of IR architecture, including gene-dosage effects and altered plastid genome copy number, supporting the view that IR retention reflects both mutational/structural buffering and functional constraints on expression capacity (Krämer et al., 2024). Together, these comparative and experimental lines of evidence align with our low-variation IR baseline and higher, punctuated variation across LSC/SSC, and they reinforce a practical division of labor. As a result, IR segments are useful for conserved anchors and robust alignments across deep divergence, whereas single-copy hotspots provide greater discriminatory power for shallow divergences.

Our gene-level inspections suggested that coding-sequence variability is concentrated in a subset of photosystem II loci, while genes involving transcription and translation remain comparatively conserved. In grasses, plastome-wide screens have similarly emphasized that most plastid genes are under purifying selection, but that selective pressures can differ substantially among functional modules and may become detectable during major ecological or photosynthetic transitions, including C3–C4-associated changes (Piot, Hackel, Christin, & Besnard, 2018). A plausible synthesis is that PSII genes accommodate limited, site-specific divergence while retaining strong constraint at photochemically essential positions, producing elevated maxima within otherwise conserved proteins.

The whole-plastome maximum-likelihood phylogeny recovers the accepted deep structure of *Poaceae*, consistent with published grass plastome phylogenies (Saarela et al., 2018). However, plastome topologies can shift under extensively larger datasets, and plastome history can differ from species history where hybridization and introgression have occurred, making cytonuclear discordance a key interpretive caveat (Orton et al., 2021; Zhang, Zeng, & Li, 2012). Accordingly, the plastome tree is best viewed as a strong scaffold for plastid evolution and comparative rate/selection analyses, while nuclear data are needed to resolve reticulate histories in problem clades.

Codon-based tests reinforce this overall picture of constraint with punctuated departures. Most of the plastome genes show patterns consistent with pervasive purifying selection. On the other hand, we also detected gene-wide evidence of episodic diversifying selection in a subset of loci. Episodic-selection frameworks were designed to detect transient or lineage-restricted selection that can be missed by gene-wide average dN/dS summaries (Murrell et al., 2015). Prior *Poaceae*-focused analyses likewise report that selection intensity varies among plastid genes and that some photosynthesis-related loci (especially *rbcL*) show stronger signals during evolutionary transitions in the PACMAD clade (Piot et al., 2018). In our dataset, the concentration of significant signals in transcriptional machinery (*rpo* genes) and energy-transduction pathways is consistent with the hypothesis that adaptive episodes may accompany lineage-specific shifts in plastome gene regulation and photosynthetic performance, even against a background of strong functional constraint.

Beyond interpretation, the combined outputs provide an actionable toolkit: (i) ranked hypervariable windows for marker design, (ii) a robust plastome phylogeny for mapping lineage-specific rate shifts, and (iii) a short list of candidate genes for deeper site- and branch-level follow-up (e.g., MEME-like site tests, structural mapping of variable residues, or clade-partitioned selection models). Future work should also integrate ecological

metadata (habitat, climate, photosynthetic type) and nuclear loci to test whether the detected rate/selection shifts correlate with major ecological transitions or reflect cytonuclear conflict and introgression in reticulate groups.

ACKNOWLEDGEMENTS

The author gratefully acknowledges Prof. Dr. M. Bahattin Tanyolaç for his valuable guidance, insightful advice, and support throughout this study.

Conflict of Interest: All the authors declare no conflict of interest.

Data availability: All the genome files are publicly available through the NCBI genome database. Multiple sequence alignment and phylogenetic tree files are available upon reasonable requests.

Code availability: All the analysis codes are available through the GitHub repository https://github.com/yasinkaymaz/cpGenomics

REFERENCES

Capella-Gutiérrez, S., Silla-Martínez, J.M., & Gabaldón, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics,* ***25***(15), 1972-1973. https://doi.org/10.1093/bioinformatics/btp348

Çatal, M. İ. (2025). Botanical Composition and Pasture Quality Assessment of Sıraköy Pasture in Çamlıhemşin (Rize, Türkiye). [Çamlıhemşin (Rize, Türkiye) Sıraköy Merasının Botanik Kompozisyonu ve Mera Kalitesinin Değerlendirilmesi]. *Journal of Anatolian Environmental and Animal Sciences,* ***10***(5), 712-717. https://doi.org/10.35229/jaes.1724457

Corriveau, J.L., & Coleman, A.W. (1988). Rapid Screening Method to Detect Potential Biparental Inheritance of Plastid DNA and Results for Over 200 Angiosperm Species. *American Journal of Botany,* ***75***(10), 1443-1458. https://doi.org/10.1002/j.1537-2197.1988.tb11219.x

Daniell, H., Lin, C.-S., Yu, M., & Chang, W.-J. (2016). Chloroplast genomes: diversity, evolution, and applications in genetic engineering. *Genome Biology,* ***17***(1), 134. https://doi.org/10.1186/s13059-016-1004-2

Duvall, M.R., Burke, S.V., & Clark, D.C. (2019). Plastome phylogenomics of Poaceae: alternate topologies depend on alignment gaps. *Botanical Journal of the Linnean Society,* ***192***(1), 9-20. https://doi.org/10.1093/botlinnean/boz060

Group1, C.P.W., Hollingsworth, P.M., Forrest, L.L., Spouge, J.L., Hajibabaei, M., Ratnasingham, S., ..., & Little, D.P. (2009). A DNA barcode for land plants. *Proceedings of the National Academy of Sciences,* ***106***(31), 12794-12797. https://doi.org/10.1073/pnas.0905845106

Jansen, R.K., & Ruhlman, T.A. (2012). Plastid Genomes of Seed Plants. In R. Bock & V. Knoop (Eds.), *Genomics of Chloroplasts and Mitochondria* (pp. 103-126). Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-007-2920-9

Katoh, K., Misawa, K., Kuma, K., & Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res,* ***30***(14), 3059-3066. https://doi.org/10.1093/nar/gkf436

Krämer, C., Boehm, C.R., Liu, J., Ting, M.K. Y., Hertle, A.P., Forner, J., ..., & Bock, R. (2024). Removal of the large inverted repeat from the plastid genome reveals gene dosage effects and leads to increased genome copy number. *Nature Plants,* ***10***(6), 923-935. https://doi.org/10.1038/s41477-024-01709-9

Kress, W.J., & Erickson, D.L. (2007). A Two-Locus Global DNA Barcode for Land Plants: The Coding rbcL Gene Complements the Non-Coding trnH-psbA Spacer Region. *PLoS One,* ***2***(6), e508. https://doi.org/10.1371/journal.pone.0000508

Li, F.W., Kuo, L.Y., Pryer, K.M., & Rothfels, C.J. (2016). Genes Translocated into the Plastid Inverted Repeat Show Decelerated Substitution Rates and Elevated GC Content. *Genome Biology and Evolution,* ***8***(8), 2452-2458. https://doi.org/10.1093/gbe/evw167

Linder, H. P., Lehmann, C. E. R., Archibald, S., Osborne, C. P., & Richardson, D. M. (2018). Global grass (Poaceae) success underpinned by traits facilitating colonization, persistence and habitat transformation. *Biol Rev Camb Philos Soc,* ***93***(2), 1125-1144. https://doi.org/10.1111/brv.12388

Murrell, B., Weaver, S., Smith, M. D., Wertheim, J. O., Murrell, S., Aylward, A., ..., & Kosakovsky Pond, S.L. (2015). Gene-Wide Identification of Episodic Selection. *Molecular Biology and Evolution,* ***32***(5), 1365-1371. https://doi.org/10.1093/molbev/msv035

Murrell, B., Wertheim, J.O., Moola, S., Weighill, T., Scheffler, K., & Kosakovsky Pond, S.L. (2012). Detecting individual sites subject to episodic diversifying selection. *PLoS Genet,* ***8***(7), e1002764. https://doi.org/10.1371/journal.pgen.1002764

Orton, L.M., Barberá, P., Nissenbaum, M.P., Peterson, P.M., Quintanar, A., Soreng, R.J., & Duvall, M.R. (2021). A 313 plastome phylogenomic analysis of Pooideae: Exploring relationships among the largest subfamily of grasses. *Molecular Phylogenetics and Evolution,* ***159***, 107110. https://doi.org/10.1016/j.ympev.2021.107110

Piot, A., Hackel, J., Christin, P.A., & Besnard, G. (2018). One-third of the plastid genes evolved under positive selection in PACMAD grasses.

Planta, ***247***(1), 255-266. https://doi.org/10.1007/s00425-017-2781-x

Pond, S.L.K., Frost, S.D.W., & Muse, S.V. (2004). HyPhy: hypothesis testing using phylogenies. *Bioinformatics,* ***21***(5), 676-679. https://doi.org/10.1093/bioinformatics/bti079

Saarela, J.M., Burke, S.V., Wysocki, W.P., Barrett, M.D., Clark, L.G., Craine, J.M., ..., & Duvall, M.R. (2018). A 250 plastome phylogeny of the grass family (Poaceae): topological support under different data partitions. *PeerJ,* ***6***, e4299. https://doi.org/10.7717/peerj.4299

Shaw, J., Shafer, H.L., Leonard, O.R., Kovach, M.J., Schorr, M., & Morris, A.B. (2014). Chloroplast DNA sequence utility for the lowest phylogenetic and phylogeographic inferences in angiosperms: the tortoise and the hare IV. *American Journal of Botany,* ***101***(11), 1987-2004. https://doi.org/10.3732/ajb.1400398

Wong, T.K.F., Ly-Trong, N., Ren, H., Baños, H., Roger, A.J., Susko, E., ..., & Quang, B. (2025). *IQ-TREE 3: Phylogenomic Inference Software using Complex Evolutionary Models*. https://doi.org/10.32942/X2P62N

Zhang, Y.X., Zeng, C.X., & Li, D.Z. (2012). Complex evolutionn Arundinarieae (Poaceae: Bambusoideae): Incongruence between plastid and nuclear GBSSI gene phylogenies. *Molecular Phylogenetics and Evolution,* ***63***(3), 777-797. https://doi.org/10.1016/j.ympev.2012.02.023