BMC Plant Biology

RESEARCH Open Access

Plastome evolution in annual *Brachypodium* species reveals widespread heteroplasmy and chloroplast capture, lineage-specific codon usage bias, and low positive selection

Miguel Campos[1,2], John P. Vogel[3], Ernesto Pérez-Collazos[1,2] and Pilar Catalán[1,2*]

Abstract

Background Comparative genomics and plastome phylogenomics have advanced significantly in recent years, highlighting the diversity, possible admixture, and non-neutral evolution of the predominantly considered non-recombinant chloroplast genomes in angiosperms. The grass genus *Brachypodium* serves as a powerful model for studying evolutionary processes in monocots.

Results We analyzed 287 plastomes across the native circum-Mediterranean range of the three annual *Brachypodium* species (*B. distachyon*, *B. stacei*, *B. hybridum*), focusing on their structural variation, selection patterns and phylogenomic relationships. Our analyses confirmed the differentiation of the S and D plastomes, inherited respectively from the diploid progenitor species *B. stacei* and *B. distachyon*. We identified novel structural rearrangements and indels, and unique repeat motifs, along with widespread heteroplasmy, particularly in ancestral *B. hybridum*-D plastotypes. SNP diversity varied among plastotypes, reflecting population dynamics and evolutionary histories, with *B. hybridum*-D plastotypes showing the highest normalized diversity and *B. hybridum*-S the lowest. Positive selection was detected in 29 plastid genes by Tajima's neutrality test, and in nine genes by site and branch-site evolutionary models, including matK, ndhF, rbcL, and rpoC2. Phylogenomic analyses revealed well-supported clades corresponding to the S and D plastome lineages, with frequent chloroplast capture events and long-distance dispersals shaping their evolutionary trajectories.

Conclusions These findings highlight the evolutionary complexity of the *Brachypodium* panplastome, emphasizing the roles of heteroplasmy, structural variation, and adaptive evolution in shaping plastome diversity in this group of model grasses. While the ancestral plastomes of *B. hybridum*-D show evidence of heteroplasmy caused by introgressive hybridization and recombination, most of the heteroplasmic patterns in the other plastome groups can be attributed to somatic mutations. The topological congruence of phylogenies based on neutrally evolving whole plastome sequences and the majority of the plastid encoding gene sequences, including positively selected genes like ndhF, underscores the robustness of plastome-based phylogenies even under selective pressures. Together, these results shed new light on the evolutionary complexity of plastid genomes in a model polyploid system,

*Correspondence:
Pilar Catalán
pcatalan@unizar.es

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

demonstrating the interplay of structural divergence, heteroplasmy, hybridization, and selection in shaping plastome diversity. Our work thus provides a valuable comparative framework for future evolutionary, ecological, and functional genomic studies in *Brachypodium* and other grasses.

Keywords Allopolyploidy, Introgression, Plastid genomes, Positive selection, Phylogenetics

Background

The phylogenomic study of angiosperm plastomes has seen significant advances in recent years, enabling researchers to untangle the intricate patterns of chloroplast genome evolution and understand the underlying mechanisms driving genetic diversity and performance within plant species [4, 25, 81]. Pan-plastome analysis involves the exhaustive examination of all chloroplast genomes from a group of related species [39, 40, 75]. Comparative genomics and phylogenomics of plastome data have provided valuable information on the evolutionary history, adaptive potential, and lineage-specific plastid traits of land plants [78, 80]. Despite these advances, the role of heteroplasmy and positive selection in plastome evolution remains poorly explored, especially at the population level. Understanding these factors is crucial to elucidating the mechanisms that drive plastid diversity and adaptation.

Heteroplasmy, the presence of multiple types of chloroplast DNA within a single organism, can arise from biparental inheritance, somatic mutations, or introgressive hybridization, and may play a crucial role in shaping the evolutionary trajectories of plant species [5, 44]. In land plants, deviations from strict maternal inheritance patterns of plastid and mitochondrial genomes are not uncommon, and some species illustrate biparental or even paternal transmission [41, 43, 59]. However, in most angiosperms organellar inheritance is primarily maternally driven and heteroplasmy is a relatively rare event [58, 66]. Exceptions include highly hybridogenous groups, such as *Citrus* [5] or *Daucus* [58], where heteroplasmy is mainly due to introgression. This phenomenon, however, has not been extensively analyzed within a broad population-level sampling of plastomes in angiosperms. Given the hybrid allopolyploid origin of *B. hybridum* and the complex evolutionary histories of its progenitor species, exploring heteroplasmy and signals of positive selection provides a unique opportunity to understand plastome evolution, structural divergence, and the potential adaptive relevance of plastid variation in this model system. Recent large-scale phylogenomic studies of plastomes in grasses have emphasized the ability of exhaustive plastome sampling to resolve complex evolutionary relationships, particularly in groups affected by reticulation and hybridization [53]. These studies have also highlighted the importance of considering plastome-nuclear incongruence and plastid capture, both phenomena particularly relevant for understanding evolutionary dynamics within grasses. Therefore, it is of high interest to test this in our model system, *Brachypodium*.

Positive selection, or adaptive evolution, is another key feature of some plastid genes or gene codons [23, 66, 77]. Although most nuclear and organelle coding genes are under negative or neutral selection [19], which maintain protein function, positive selection peaked in several plastid proteins during the evolution of ancestral grasses [23]. Recent studies have demonstrated that it affects specific codons of Poaceae plastome genes [61], up to a third of the PACMAD tropical grass genes [55], and more specifically photosynthetic genes of temperate-steppe Stipeae grasses [34]. Investigating positive selection on plastid genes at the population and progenitor-hybrid levels in our *Brachypodium* species may reveal adaptive changes that have occurred in response to environmental pressures, contributing to our understanding of plant adaptation and evolution.

The grass genus *Brachypodium* has emerged as a model system for studying monocot functional genomics and has gained recognition for its potential use in agricultural and biofuel research [67, 74]. The genus comprises three annual species, the diploids *B. stacei* (2n = 2x = 20; x = 10) and *B. distachyon* (2n = 2x = 10; x = 5), and their derived allotetraploid *B. hybridum* (2n = 4x = 30; x = 5 + 10), which exhibit remarkable diversity in their morphology, life history traits and ecological niches [3, 8, 38, 42, 46, 47, 69, 70]. In general, *B. distachyon* mainly inhabits higher, cooler, and wetter regions of the circum-Mediterranean area, *B. stacei* is adapted to lower, warmer, and drier environments, and *B. hybridum* is found in areas with intermediate climatic features [38]. The chloroplast genomes of *B. stacei* (S plastome) and *B. distachyon* (D plastome) differ structurally from each other in the presence of some species-specific indels [63]. *B. hybridum* allotetraploids of recurrent and bidirectional origin exhibited a maternal D plastome (D-plastotypes) in the ancestral lineage (Anc) and a S plastome (S-plastotypes) in the recent western (W) and eastern (E) Mediterranean lineages [47, 63]. Ecotypes of both progenitor species, *B. distachyon* and *B. stacei*, evidenced the existence of different intraspecific chloroplast capture events [3, 17, 63].

While previous studies (e.g., [63]) have characterized plastome variation in *B. distachyon* primarily, our study extends this by incorporating broader sampling across the three annual species (*B. distachyon, B. stacei, B. hybridum*), applying a comprehensive panplastomic framework, and explicitly investigating patterns of

heteroplasmy, chloroplast capture, codon usage bias, and positive selection in progenitor and hybrid lineages providing a lineage-level comparative analysis capturing the evolutionary complexity of plastome variation within and between species. By focusing on heteroplasmy and positive selection, we aim to uncover the mechanisms driving plastome diversity and adaptation in *Brachypodium*, offering insights into the evolutionary processes shaping plastid genomes in ancestral and newly evolved plastotypes. Here we have enlarged the sampling of accessions of the three annual *Brachypodium* species in their native circum-Mediterranean region and investigated their plastomes aiming to: (i) characterize the diversity and structural variations of plastomes within and between *Brachypodium* species and plastotypes, (ii) assess the prevalence and patterns of heteroplasmy in natural populations, particularly that caused by introgression, (iii) identify signatures of positive selection acting on plastid genes, and (iv) integrate these findings to explore the deep and shallow phylogenomic relationships and evolutionary history of the *Brachypodium* plastid pangenomes and the putative phylogenetic signal of positively selected genes.

Methods

Plant material, DNA extraction and Illumina sequencing

A total of 287 assembled and annotated plastomes were used in the study [54 *B. stacei*; 111 *B. distachyon*; 122 *B. hybridum* (117 S-plastotype, 5 D-plastotype)]. Of them, 162 were generated *de novo* from newly sequenced accessions [*B. stacei* (ERS20220148 - ERS20220177, *B. distachyon* (ERS20220139 - ERS20220146), *B. hybridum* (ERS20220178 - ERS20220180)], and 125 (*B. stacei*, 26; *B. distachyon*, 87; *B. hybridum*, 12) were incorporated from previous studies [17, 46, 47, 63] (Fig. S1, Table S1).). Four of the 287 samples were collected from non-circum-Mediterranean invaded areas [*B. hybridum*-S plastotype: ABR101 and ABR 115 (South Africa), ABR135 (Uruguay), ABR137 (Western Australia) (Table S1)] and were included in the analysis to try to infer the geographic origin of their respective ancestors. The taxonomic identity of individuals was confirmed by morphological analysis following [7, 8] as well as genomic analysis of the nuclear genomes. Seeds were collected from natural populations and germplasm banks (Table S1) and grown under standard greenhouse conditions (16h light period at 20 C) at the High Polytechnic School of Huesca, Zaragoza University (Spain) and the DOE Joint Genome Institute (USA).

Total DNA was isolated from each sample using the CTAB protocol with modifications [16]. Whole genome sequencing (WGS) was performed at the DOE Joint Genome Institute (JGI) using Illumina technology. Libraries for 2x151 PE reads were constructed and sequenced using the Illumina NovaSeq S4 platform. Duplicate reads were removed based on paired sequence matching using Clumpify in BBMap [2]. BBDuk v.38.96 from BBMap was used to trim reads containing adapter sequence and G homopolymers of size 5 or more at the ends of the reads and trim reads of correct quality where quality drops below 6. BBDuk was used to remove reads that contained 1 or more 'N' bases, had an average read quality score less than 6, or had a minimum length ≤ 49 bp or 33% of the total read length. Plastome reads were filtered from the WGS data and used for plastome analysis.

Plastome assembly and annotation and plastome diversity

Whole plastome assembly of newly sequenced samples was performed using NOVOPlasty v. 4.3.1 [14]. The assembly process used a K-mer size of 39 and the entire *B. stacei* ABR114 plastome sequence (GenBank: LT558589.1) as a seed to initiate the extension and assembly of the *Brachypodium* plastome samples under study to ensure a reliable and accurate reconstruction of whole plastome sequences. Plastome assemblies available from *B. hybridum*, *B. distachyon* and *B. stacei* accessions from [63] and [46] were downloaded and added to the study (Table S1).

Plastome annotations were performed using Geneious Prime v.2023.1.2 [29] Biomatters Ltd, Auckland, NZ, https://www.geneious.com) by aligning each of the newly assembled plastomes to the reference annotated plastomes of *B. stacei* ABR114 (Accession: LT558589.1) and *B. distachyon* Bd21 (Accession: NC_011032). We also annotated both reference plastome sequences using GeSeq [72] to obtain complete annotations. We then manually curated these annotations in GeSeq to refine the precision of start and stop codons, and to establish exons and introns boundaries. Upon completing annotation curation of the *B. stacei* and *B. distachyon* reference plastomes, we transferred these refined annotations to the newly assembled plastomes. To ensure the accuracy and integrity of the transferred annotations, we conducted a posterior manual verification step.

The assembled plastomes were aligned using MAFFT v.7.031b [28]. To minimize the inclusion of spurious gaps and alignment noise, we adjusted the gap-opening penalty in MAFFT during multiple sequence alignment. This strategy reduces the occurrence of arbitrarily inserted gaps while preserving true insertion/deletion signals. In addition, all alignments were visually inspected and processed in Geneious Prime to ensure accurate homology assessment. Analyses of the number of haplotypes, the number of mutations (SNPs) and haplotype diversity (Hd) were performed using the average frequencies of all samples of each species and plastotype in DnaSP6 [60]. The haplotype number reveals the total number of unique haplotypes representing the genetic variants

present in the dataset. Haplotype diversity (Hd) was calculated to measure the probability that two randomly chosen haplotypes are different.

Heteroplasmy analysis

Plastome sequence alignments were used to detect regions of common insertions and deletions (indels) shared by *Brachypodium* species and/or plastotypes using a custom script (Indels.R). These indels were compared with plastome sequences from closely related species (*Oryza sativa*, *Diarrhena obovata*) to determine whether they represented true insertions or deletions. The indel data were then cross-referenced to determine if the detected insertions exhibited high levels of heteroplasmy, which may have resulted from introgression [63].

Following the initial assembly of all plastomes, a second round of assembly was performed with NOVOPlasty using the first sequence obtained as a reference to detect possible heteroplasmic positions, while filtering out artifactual variation. Heteroplasmy detection was performed with a Minor Allele Frequency (MAF) threshold of 0.01, allowing the identification of heteroplasmy levels greater than 1%. NOVOPlasty applies additional filtering steps to minimize the influence of plastid-like nuclear or mitochondrial sequences (NUPT and MTPT), which can confound heteroplasmy analyses. A combined forward and reverse assembly approach is used to validate haplotype continuity, and a final local reassembly step is performed to remove discordant variants that could disrupt the integrity of the organellar haplotype, labeling them as likely NUMT/MTPT-derived [14, 15]. Variant call format (VCF) files obtained from NOVOPlasty containing heteroplasmy data were loaded into Geneious Prime for further analysis. Heteroplasmic positions were aligned, and a presence-absence matrix of these positions was extracted using another custom script (parse.py). This matrix was then used to calculate the number of heteroplasmic positions shared between species and plastotypes.

Heteroplasmic positions belonging to SNVs (Single Nucleotide Variants) within the plastomes of the *B. hybridum* plastotypes D and S were analyzed to determine their compatibility with the plastome sequences of the progenitor species. For each heteroplasmic position, the minor allele was compared to the corresponding positions in the plastome sequences of *B. distachyon* and *B. stacei* to determine compatibility. Progenitor species sequences were generated as consensus sequences from the plastomes of all individuals for each species. This analysis was performed using a third custom script (Origin.R), which calculated the number of matches unique to the maternal parent, unique matches to the paternal parent, matches to both parents, and positions that did not match either parent. This test assessed whether the minor allele at the heteroplasmic loci could have been inherited from the paternal parent following [44] and [5].

Identification of SSRs and dispersed repeats

We employed the SciRoKo software [30] with MISA-mode to identify Single Sequence Repeats (SSRs) within the studied plastomes. To detect SSR motifs, we applied specific conditions for the different types of plastid SSRs, ranging from one to six nucleotides in length. For detection of hexa-, penta-, and tetranucleotide SSRs, we imposed a minimum number of three repeat units, while for tri-, di-, and mononucleotide SSRs, we used a minimum number of repeats of 4, 5 and 10 units, respectively. Furthermore, to evaluate the distribution and variations in SSR composition, we mapped the SSRs detected to the four major plastome regions (LSC, IRA, SSC, IRB) and compare them between plastomes of species and plastotypes.

Codon usage analysis

Relative synonymous codon usage (RSCU) analysis of both types of plastomes (S and D) was conducted in MEGA11 [71] to analyze potential codon bias and preferences. RSCU analysis evaluates the frequency of occurrence of synonymous codons relative to their expected frequency under a specific reference codon usage.

Positive selection at codon-level

Protein coding sequences (CDS) were extracted using Geneious Prime and positive selection analysis was conducted on the 76 shared CDS in the 287 studied plastomes to investigate selection signatures. First, Tajima's D was calculated using the PopGenome package in R [54] to identify deviations from neutrality, where negative values suggest purifying selection or population expansion, and positive values indicate balancing or positive selection. Codon-level selection analyses were performed using different site- and branch-based selection models implemented in HyPhy [33] after removing stop codons from the alignments. The Mixed Effects Model of Evolution (MEME; [31, 32] was employed to detect episodic positive selection at individual codon sites, while BUSTED (Branch-Site Unrestricted Statistical Test for Episodic Diversification; [49] assessed gene-wide episodic diversifying selection. Pervasive positive or purifying selection was analyzed using site-based FUBAR (Fast Unbiased Bayesian AppRoximation; [48], and FEL (Fixed Effects Likelihood; [32] models. Default parameters were applied for all models unless specified, and codons under significant positive selection were identified based on a p-value < 0.05 or posterior probability > 0.9. The results of Tajima's D neutrality test for plastid genes were integrated with these findings to provide additional context

on the evolutionary selection pressures acting on the plastid genes.

Phylogenomic and evolutionary population analyses

For the reconstruction of the *Brachypodium* phylogenomic trees, we employed three separate main datasets: (1) the complete plastome sequences, (2) the concatenated coding sequences (CDS) of the 76 plastid genes, and (3) a subset of 35 CDS genes identified as under positive selection by Tajima's neutrality test and/or by evolutionary models. In addition, we also constructed individual *Brachypodium* phylogenetic trees for each of the 35 positively selected genes, several of which have been commonly used as barcoded plastid markers (e.g., matk, ndhF, rbcL, rpoC2). Each dataset was analyzed for the 287 individual samples using a maximum-likelihood (ML) approach with IQ-TREE2 [51] and *Oryza sativa* and *Diarrhena obovata* to root the trees. We performed 1000 bootstrap replicates to provide support for the inferred relationships and evaluate agreement between different data partitions with the whole-plastome tree. We performed the same ML searches on a subset of 12 samples for each data set, leaving only one representative sample from each major *Brachypodium* lineage. To test topological congruence between the trees, we performed approximately unbiased (AU) and Kishino-Hasegawa (KH) and Shimodaira-Hasegawa (SH) tests and other probability-based comparisons, including log-likelihood scores (logL), differences in log-likelihood relative to the best tree (deltaL), bootstrap proportions (bp-RELL), and expected likelihood weights (c-ELW). We were interested in testing whether alternative topologies, based primarily on genes or sets of genes under putative positive selection, were significantly incongruent with respect to the reference topology, based on whole-plastome sequences evolving under neutrality, on deep and shallow branches. The resulting trees were visualized and annotated in iTOL [37], and branch colors indicated levels of topological concordance detected across all datasets.

For population-level analyses, separate ML phylogenomic trees were constructed for the D-plastomes and S-plastomes to analyze branch lengths and clade structure at finer resolutions. We also employed Discriminant Analysis of Principal Components (DAPC) using the adegenet package in R [26, 27]. DAPC was used to identify genetic clusters based on non-redundant datasets. Principal Component Analysis (PCA) reduced genetic data to uncorrelated variables, while Linear Discriminant Analysis (LDA) identified axes that maximize separation between groups. The optimal number of clusters was determined using Ward's method, minimizing within-cluster variability, and the Akaike Information Criterion (AIC), balancing model fit and complexity. The best genetic groupings were compared to ML phylogenies using iTOL. Haplotype networks were constructed using statistical parsimony methods implemented in TCS [10] to infer relationships and were visualized with PopArt [36]. Genetic clusters identified with DAPC were compared to phylogenetic trees and haplotype networks, providing a comprehensive view of the genetic structure and diversity within and between plastome lineages.

Results

Characteristics of the plastomes of the annual *Brachypodium* species

Our comparative analysis of pan-plastomes, representing the entire distribution range of the three annual *Brachypodium* species and the *B. hybridum* plastotypes (Fig. S1) plus a few invaded ranges (Table S1), showed high structural similarity to previously described S and D plastomes [63] (Fig. 1), although our exhaustive search uncovered new evidence for small rearrangements and heteroplasmy (Fig. 1, Table S2). The full plastome lengths had mean sizes of 136,314 bp for S plastomes (136,292 *B. stacei*, 136,328 *B. hybridum* S-plastotype) and 135,158 bp for D plastomes (135,167 *B. distachyon*, 135,433 *B. hybridum* D-plastotype). The major differences affected the large single copy (LSC) region, which spanned an average of 81,251 bp in S plastomes (81,248 *B. stacei*, 81,254 *B. hybridum* plastotype S) and 79,884 bp in D plastomes (79,764 *B. distachyon*, 80,004 *B. hybridum* plastotype D), and was caused by a large insertion of 1161 bp between psaI and rbcL (including the pseudogene rpl23) present only in S plastomes. The small single copy (SSC) region [average size of 12,653 bp in S plastomes (12,645 *B. stacei*, 12,662 *B. hybridum* plastotype S), 12,635 bp in D plastomes (12,636 *B. distachyon*, 12,632 *B. hybridum* plastotype D)] and the two inverted repeat (IR) regions [21,200–21,202 bp in S plastomes, 21,193–21,195 bp in D plastomes] showed similar lengths in both plastome types, although the IRb contained a 180 bp insertion of an rps19 copy between psbA and trnH only present in D plastomes (Fig. 1, Table S2). For all four plastome regions, the mean lengths of the *B. hybridum* S and D plastotypes were slightly longer than those of their respective progenitor species.

Despite these structural differences, the overall gene content remained constant across all the plastomes. Both S and D plastomes contained an identical complement of 133 genes, including 76 protein-coding genes, seven of which were duplicated genes, 38 tRNA genes (20 non-redundant), 8 rRNA genes (4 in each IR), 4 pseudogenes (*trn*I, *rps*12a, *trn*T and *trn*I) and two hypothetical open reading frames: *ycf3* (pafI) and *ycf4*, involved in PSI assembly and PSI subunit integration, respectively [35, 50]. Due to the structural differences observed, the rpl23 pseudogene was exclusive to the S plastomes, and the rps19 duplication to the D plastomes. All plastomes

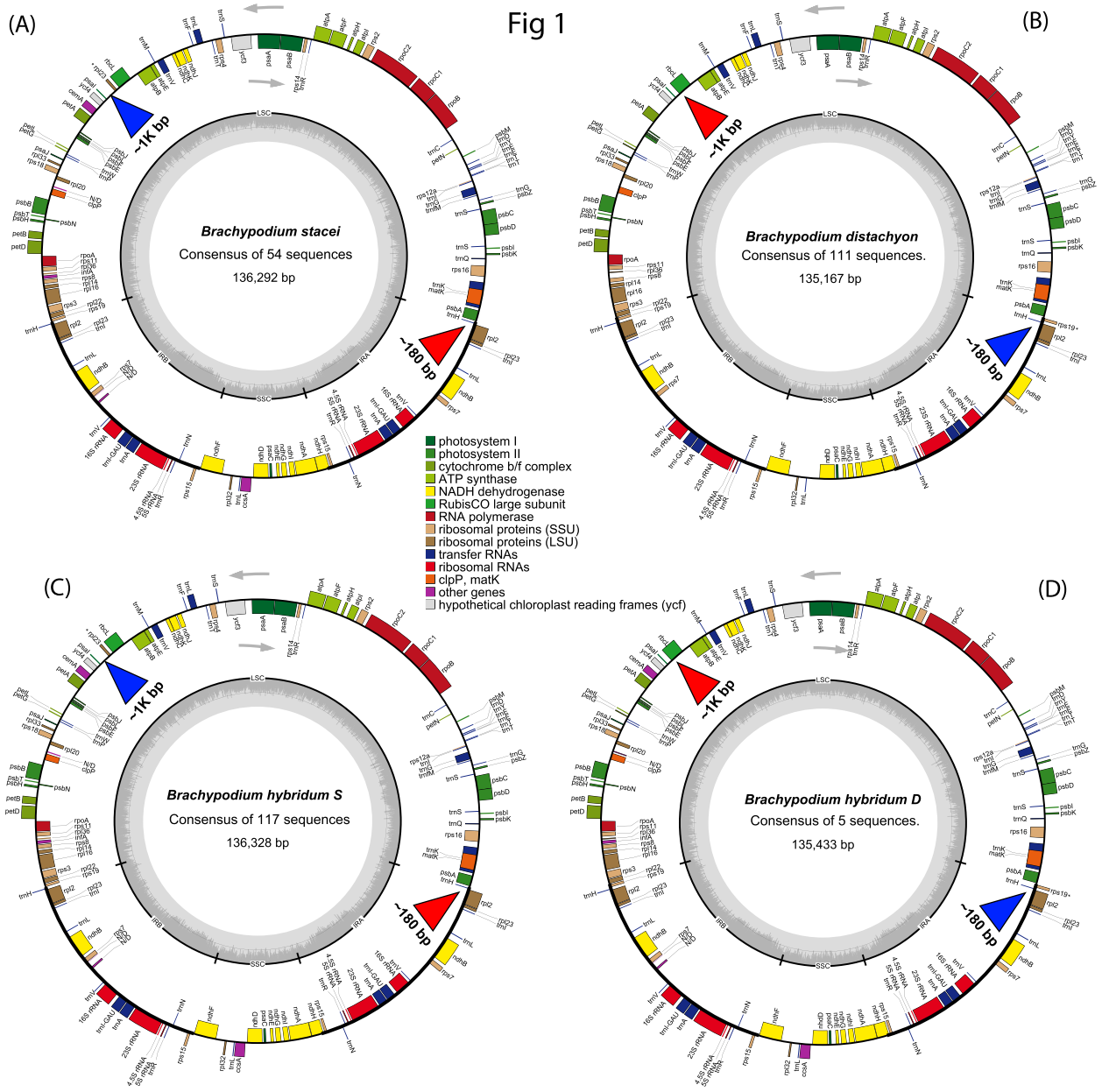

Fig. 1 Comparison of plastome maps constructed from whole-plastome consensus sequences of (**A**) *B. stacei* (54 sequences), (**B**) *B. distachyon* (111), (**C**) *B. hybridum* S-plastotype (117), and (**D**) *B. hybridum* D-plastotype (5). Maps highlight a ~1K bp indel in the LSC region containing an rpl23 pseudogene (marked with a blue triangle for insertion in S plastomes and red for deletion in D plastomes) and a 180 bp indel containing a copy of rps19 in the IRB (marked with a blue triangle for insertion in D plastomes and red for deletion in S plastomes). Small inner circles represent plastome regions (LSC, SSC, IRA, IRB) and the GC content histogram (dark and light grey circles) in each plastome group

exhibited a very similar GC content, with the LSC region averaging ~36.5%, the SSC region ~32.7%, and the IR regions showing a consistently higher GC content of approximately 44.1% (Table S2).

In addition to the structural variations, the total number of SNPs varied widely between species and plastotypes, with *B. hybridum* D displaying the highest number of SNPs per individual (10.4), and *B. hybridum* S the lowest (1.73) (Table S2). Intermediate levels of SNPs per individual were observed in *B. stacei* (4.81) and *B. distachyon* (3.86). SNPs were mainly concentrated in the LSC and SSC regions, which displayed greater variation compared to the more conserved IR regions and showed numbers per individual proportional to the respective totals. Coding regions contained fewer SNPs overall, highlighting selective constraints on protein-coding genes (Table S2).

Haplotype diversity also varied across species and plastotypes, with *B. hybridum* D showing the highest diversity (1,0), despite having only five haplotypes, suggesting strong differentiation among individuals in this plastotype (Table S2). In contrast, *B. hybridum* S had the lowest haplotype diversity (0.675), likely reflecting reduced variability in these allotetraploids, and *B. stacei* (0.915) and *B. distachyon* (0.919) exhibited similarly high levels of haplotype diversity. Tajima's D values at the level of 76 concatenated plastid CDS revealed only significant negative values for *B. hybridum* S plastomes (–2.369, $p<0.01$), suggesting possible negative selection or, more likely, population expansion in these allotetraploids, while values for the other groups were not significant (Table S2).

Indels and heteroplasmy

Indels and polymorphic position (SNV) analysis revealed multiple structural and mutational variations across all plastomes studied in the annual *Brachypodium* species (Fig. S2; Tables 1A, B, S3). Comparisons between plastome types (S and D) and between species and plastotypes identified significant structural differences, many of which were supported by alignments with reference plastomes from outgroup grasses (*Diarrhena obovata, Oryza sativa*). Key indels that distinguish S plastomes (*B. stacei* and *B. hybridum* S) from D plastomes (*B. distachyon* and *B. hybridum* D) included a 132 bp deletion in S plastomes between rps16 and psbK (indel #5) and a 369-bp deletion (including rps19) between rpl2 and psbA (indel #32), and a 1160 bp deletion in D plastomes between rbcL and psaI (indel #25). Species-or plastotype-specific indels were also observed, such as a 122 bp deletion in *B. hybridum* D (petN-trnC) (indel #13) and a 251 bp deletion in *B. distachyon* (psbZ-rps12a) (indel #8) (Fig. S2; Tables 1, S3). Noticeably, *B. hybridum* D shared five insertions with S plastomes such as a 49 bp insertion between psbA and matK (indel #1), a 69 bp insertion between matK and rps16 (indel #3), a 251 bp insertion between trnG and trnI (indel #8), a 34 bp insertion between psaA and ycf3 (indel #19), and a 56 bp insertion between ndhC and trnV (indel #22), while *B. hybridum* S did not show any shared insertions with D plastomes. Interestingly, alignment comparisons identified a 27 bp insertion in *B. hybridum* D (indel #2) that is absent in the other annual *Brachypodium* species and plastotypes but present in both outgroups, suggesting that it may represent an ancestral feature retained in *B. hybridum* D (Fig. S2; Table S3).

Heteroplasmy analysis based on SNV data provided further insights into the genetic diversity and heteroplasmic patterns across species and plastotypes (Table S4; Appendix 1). *B. stacei* exhibited the highest total number of heteroplasmic positions (8656), with an average of 157.38 positions per individual, followed by *B. distachyon* (3289 positions, 29.63 per individual) and *B. hybridum* S (3226, 27.57). In contrast, *B. hybridum* D showed the lowest total number of heteroplasmic positions (222) but had a relatively high average rate per individual (44.40). When comparing plastome types, plastomes S showed a much higher number of unique heteroplasmic positions (443) compared to plastomes D (22). Across all groups, heteroplasmic positions were evenly distributed between coding and non-coding regions, with coding sequences accounting for 51.11% of heteroplasmic positions in *B. stacei*, 48.26% in *B. hybridum* S, 49.50% in *B. distachyon*, and 42.79% in *B. hybridum* D (Table S4).

Comparative analysis of polymorphic and heteroplasmic positions in *B. hybridum* S and D plastotypes revealed key insights into introgression and inheritance patterns (Table 1C). In both plastotypes, 2163 single-nucleotide variants (SNVs) were identified. Of these, plastotype S exhibited 1134 heteroplasmic positions, with 25% matching *B. stacei* (maternal parent), 20% matching *B. distachyon* (paternal parent), and 55% showing no alignment with either parental plastome. Plastotype D had 103 heteroplasmic positions, with 31.9% matching *B. stacei* (paternal parent), 27.5% matching *B. distachyon* (maternal parent), and 40.6% not aligning with either parent.

Plastid repeat sequences

Analysis of plastid simple sequence repeats (SSRs), or microsatellite repeats, revealed significant variation in the number and distribution of repeats among *Brachypodium* plastomes. A total of 31.8–37.1 SSRs were identified across the four species and plastotypes (Table S5). Statistical analyses showed significant differences between groups ($p < 0.05$); *B. stacei* (37.1 ± 1.53) and *B. hybridum* S (36.1 ± 0.57) shared significantly higher SSR counts compared to *B. distachyon* (31.8 ± 1.55) and *B. hybridum* D (32.0 ± 0.71) (Table S5). Similar significant differences were detected in the distributions of SSRs in the main plastome regions [LSC: *B. stacei* (28.9 ± 1.37) and *B. hybridum* S (28.1 ± 0.56), vs *B. distachyon* (26.5 ± 1.36) and *B. hybridum* D (26.8 ± 0.84); SSC: *B. stacei* (4.03 ± 0.31) and *B. hybridum* S (4.00 ± 0.14), vs *B. distachyon* (3.24 ± 0.62) and *B. hybridum* D (3.20 ± 0.45); IRA: *B. stacei* (2.13 ± 0.43) and *B. hybridum* S (1.99 ± 0.17), vs *B. distachyon* (1.02 ± 0.16) and *B. hybridum* D (1.00 ± 0); and IRB: *B. stacei* (2.00 ± 0) and *B. hybridum* S (2.00 ± 0), vs *B. distachyon* (1.02 ± 0.16) and *B. hybridum* D (1.00 ± 0)] (Table S6). However, although the LSC region is proportionally longer in S plastomes (Fig. 1, Table S2), the number of its SSRs is proportionally lower compared to that in D plastomes (Fig. 2A).

All four plastome groups showed significant differences among them for the SSR type category (Table S5). The most abundant SSR type across all plastomes was

Table 1 Heteroplasmy analysis of *Brachypodium* plastomes based on identified insertions and heteroplasmic positions

(A)						
Plastome type		Insertion				
	Indel No.	Start	End	Insert Size	Insertion region	Heteroplasmic positions
Plastome S	7	13116	13230	14	114	23 (1)
Plastome S	9	17060	17176	16	116	16 (4)
Plastome S	11	17542	17660	18	118	7 (1)
Plastome S	12	17725	17844	19	119	17 (3)
Plastome S	21	48193	48308	15	115	3 (2)
Plastome S	25	56865	58125	1160	1260	72 (68)
Plastome S	26	60337	60448	11	111	3 (0)
Plastome S	28	66032	66143	11	111	8 (3)
Plastome S	30	105805	105916	11	111	6(2)
						Total 155 (84)
Plastome D	5	5970	6202	132	232	0 (0)
Plastome D	6	12862	12993	31	131	0 (0)
						Total 0 (0)
(B)						
Species or plastotypes		Insertion				
	Indel N	Start	End	Insert Size	Insertion region	Heteroplasmic positions
B. stacei	16	27823	27834	11	111	12 (1)
B. distachyon	10	17334	17352	18	118	4 (1)
B. distachyon	15	19801	19824	23	123	1 (0)
B. distachyon	27	63659	63669	10	110	3 (0)
						Total 8 (1)
B. hybridum S	31	105978	106001	23	123	2 (0)
B. hybridum D	17	33041	33060	19	119	1 (0)
B. hybridum D	29	105803	105817	14	114	1 (0)
						Total 2 (0)
B. hybridum D + *B. stacei* + *B. hybridum* S	1	1551	1600	49	149	50 (16)
	3	3388	3457	69	169	51 (24)
	8	13976	14227	251	351	90 (66)
	19	42263	42297	34	134	10 (3)
	22	50746	50802	56	156	8 (4)
						Total 209 (113)
(C)						
				Origin		
	Total SNVs	Hp positions		*B. stacei*	*B. distachyon*	None
B. hybridum D	2163	103		31.9% (paternal)	27.5% (maternal)	40.6%
B. hybridum S	2163	1134		25.0% (maternal)	20.0% (paternal)	55.0%

Comparative genomic analysis of all *Brachypodium* plastomes studied at (**A**) plastome-type level (plastome S vs plastome D) and (**B**) species and plastotype levels (*B. stacei*, *B. distachyon*, *B.hybridum*-S, *B. hybridum*-D). For each insertion, the indel number (see Table S3), start and end position, insert size, insertion region (insert size + 100bp), and associated heteroplasmic positions detected within a range of 50 base pair (bp) upstream and downstream of each insertion are indicated; values in parentheses indicate heteroplasmic positions found within the insertion region. (**C**) Summary of polymorphic and heteroplasmic positions detected in the comparative plastome compatibility analysis of *B. hybridum* plastotypes S and D and their parental species, including the total number of polymorphic positions (SNVs) and heteroplasmic positions (Hp positions), and their origin, indicating the compatibility of these heteroplasmic positions with the sequences of the maternal and paternal parent species, expressed as percentages of matches with plastomes of *B. stacei* only, *B. distachyon* only, or neither parent (none)

the mononucleotide repeat, accounting for 14.9–20.1 repeats on average (Figs. 2B, C; Table S5). *B. stacei* and *B. hybridum* S had significantly higher mononucleotide repeat counts (20.1 ± 1.51 and 19.2 ± 0.71, respectively) compared to *B. distachyon* (14.9 ± 1.37) and *B. hybridum* D (16.4 ± 0.89). Tetranucleotide repeats were the second most frequent type, ranging from 8.96 ± 0.33 in *B. hybridum* S to 9.98 ± 0.16 in *B. distachyon*. Hexanucleotide repeats (AAAAAG, AAGGGT) were unique to *B. distachyon* and *B. hybridum* D, whereas the pentanucleotide repeat AAATT and the tetranucleotide AATG were

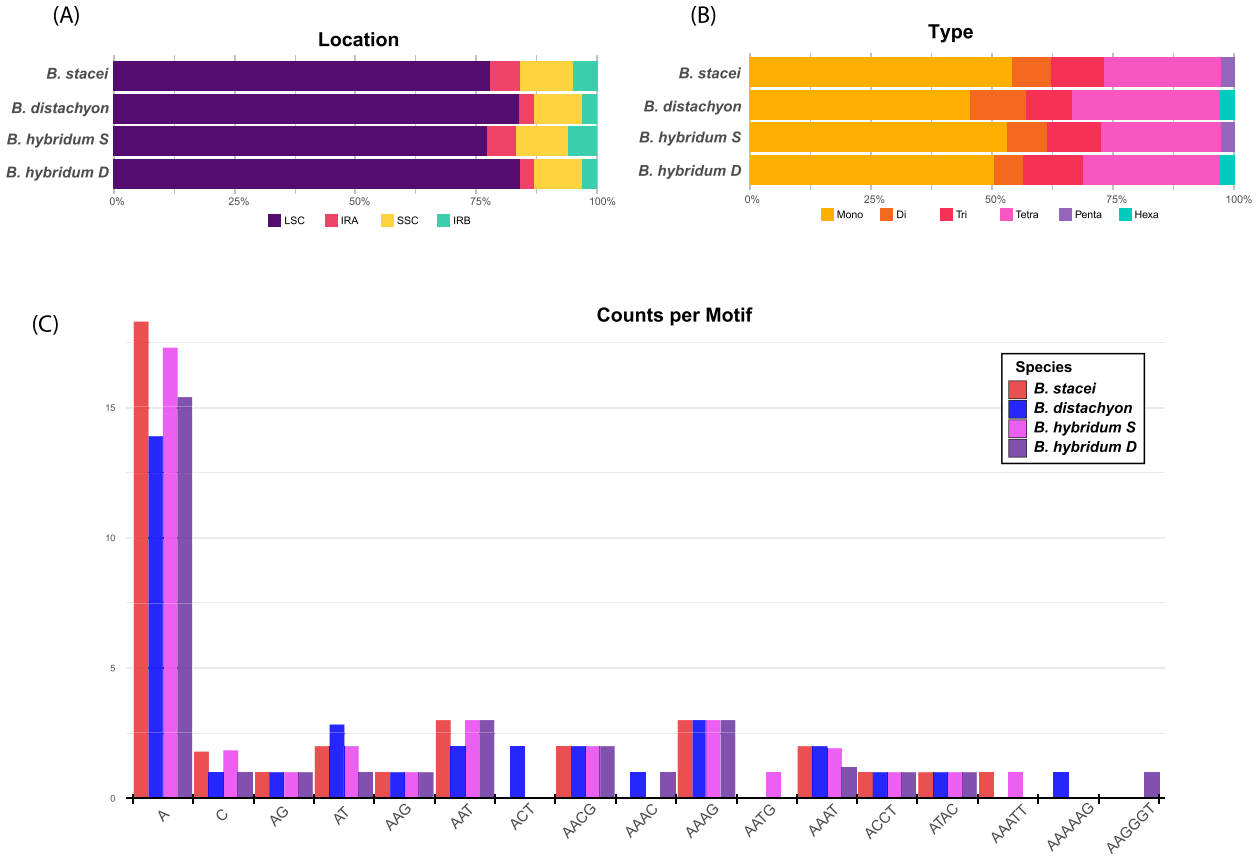

Fig. 2 Analysis of unique simple sequence repeats (SSRs) in the studied plastomes of the four *Brachypodium* species and plastotype groups. (**A**) Distribution of plastid SSRs across the plastome regions of each group, (**B**) occurrence of the six types of SSRs in those plastomes, and (**C**) numerical counts per motif in each group

unique to *B. stacei* and *B. hybridum* S (Figs. 2B, C; Table S5).

Amino acidic frequencies and codon usage

Amino acid percentage and relative synonymous codon usage (RSCU) analyses were performed on the 76 protein-coding sequences from the four *Brachypodium* plastome groups (species and plastotypes). The total number of codons of the 76 protein-coding genes was similar among groups (19,759 *B. stacei*, 19,761 *B. distachyon*, 19,761 *B. hybridum* S, 19,764 *B. hybridum* D (Fig. S3; Table S6). However, the percentage of amino acids encoded in those protein genes varied, especially in *B. distachyon* compared with the other groups, which were more similar to each other. Thus, B. *distachyon* showed higher proportions of arginine (R), asparagine (N), cysteine (C), histidine (H), phenylalanine (F), serine (S), tryptophan (W), and tyrosine (Y) in its genes than *B. stacei*, *B. hybridum* S, and *B. hybridum* D, and lower proportions of alanine (A), glutamic acid (E), isoleucine (I), lysine (K), leucine (L), methionine (M), proline (P), and valine (V) than them (Table S6). Among the encoded amino acids, leucine (L) was the most frequently represented in the genes of both the S and *B. hybridum*-D plastomes (average codon counts of 2475.9 *B. stacei*, 2475.9 *B. hybridum*-S, 2725.5 *B. hybridum* D; approximately 12.53–13.79% of their total codons), and serine (S) in those of *B. distachyon* (2255.4; 11.4%). In contrast, cysteine (C) was the least represented amino acid in the genes of the S and *B. hybridum*-D plastomes (285 *B. stacei*, 284.9 *B. hybridum* S, 320.2 in *B. hybridum* D; 1.44–1.62.44.62%) and methionine (M) in the *B. distachyon* genes (220; 1.11%) (Table S6). The RSCU values revealed variability in codon preference among synonymous codons encoding the same amino acid (Fig. S3; Table S6). For amino acids with multiple synonymous codons, RSCU values indicate the degree of bias in codon usage (values greater than 1 denote preferred codons, while values below 1 indicate underrepresented codons). Most protein-coding genes in the *Brachypodium* plastomes utilize the standard start codon (AUG; RSCU = 1.00). However, alternative start codons were observed in specific genes (e. g., GUG in rps19, UUC in rps12a, ACG in rpl2, and UAC in rps15) in all groups, consistent with findings from other plastome studies such as those of algae *Chlamydomonas* or the palm genus *Bactris* [9, 12,

62]. Similarly, all three stop codons (UAA, UAG, and UGA) were identified in the analysed plastomes, with UAA being the most frequently used termination signal (Table S6).

Selection pressure

To investigate the possible existence of selection pressure on protein-coding genes from the studied *Brachypodium* plastomes, we first performed neutrality tests at the CDS codon level using Tajima's D statistic (76 coding genes). Most genes (43) did not show significant D values for any group, indicating that they evolved neutrally. In contrast, 33 showed significant positive or negative values for any of the groups (Tables 2, S7). Of these, five genes revealed a predominance of purifying selection for *Brachypodium* or particular species or plastotypes (matK, ndhH, psbA, rpoC1, rps12) (Table S7); in contrast, 29 genes showed significant positive selection D values for *Brachypodium* (atpB, atpF, atpI, ccsA, infA, ndhA, ndhB, ndhC, ndhF, ndhG, ndhI, ndhK, petB, psaC, psbB, psbH, rbcL, rpl14, rpl16, rpl20, rpoA, rpoB, rpoC2, rps14, rps2, rps3, rps7, ycf3, ycf4) (Tables 2, S7). Interestingly, four genes showed significant positive selection for *Brachypodium* but significant negative selection for species or plastotypes (ndhF, rbcL, rpoB, rpoC2) (Table S7). The pairwise correlation between nucleotide diversity (π) and Tajima's D values across all studied *Brachypodium* genes and plastomes were significantly but modestly associated (R^2=0.235, p=7.798x10^{-6}; Fig. S4), indicating limited overall association between these parameters. However, certain genes, such as ndhF, rps2, and rpoC2, exhibited high nucleotide diversity and high D values, suggesting that these loci may be hotspots for diversification in *Brachypodium* plastomes.

Furthermore, we analyzed putative codon-level selection for the 76 protein-coding genes in *Brachypodium* plastomes using four evolutionary models. We tested for episodic positive selection on (i) the entire *Brachypodium* tree topology and particular branches of it (clade S, *B.hybridum*-D, *B.distachyon*) (see Fig. 3) using the BUSTED model, and (ii) the site-based MEME model, and for widespread positive selection across the phylogeny using the (iii) FUBAR and (iv) FEL site-based models. BUSTED revealed significant LRT values for the unconstrained model only at the full *Brachypodium* tree level, while for the individual clades (S lineage, *B. hybridum*-D, and *B. distachyon*), the detected selection was not statistically significant ($p > 0.05$) (Table 3). In all cases, the predominant selection mode indicated that most codon sites are under purifying selection, with only a small fraction showing evidence of episodic diversifying selection in *Brachypodium* (dN/dS = 34,133.479, 0.006%; $p < 0.000$). FEL analysis revealed that 70 codons out of 37 genes were under significant purifying selection in all individuals ($p < 0.05$), while some codons showed this for particular groups (*B. hybridum* S: 2; *B. hybridum* D: 1) and there was no evidence of positive selection in any case (Tables 2, S8). MEME detected 5 codon sites under significant positive selection for all *Brachypodium* plastomes (cemA 97, matK 204, ndhF 642 and 651, rpoC2 1482). One of them was also found to be significant for the plastomes of *B. stacei* (ndhF 642) and another for those of *B. distachyon* (rpoC2 1482), whereas no signals of episodic diversifying selection were detected in *B. hybridum* S or *B. hybridum* D plastomes (Tables 2, S9A). FUBAR analysis identified 12 codons of the studied genes that were under diversifying selection; nine of them showed significant positive selection in all *Brachypodium* plastomes (atpA 119, cemA 97, matK 204, ndhF 642, ndhK 238, psaB 417, rbcL 43 and 95, rps11 37), and two additional codons in *B. stacei* plastomes (ndhF 651, rpl23 2) (Tables 2, S9B). A summary of the statistical test and evolutionary models that detected significant diversifying selection for genes and/or codon sites in all *Brachypodium* plastomes or in those of some species or plastotypes suggested that 35 genes may be under positive selection, especially when considering all the samples studied; however, those selected by two or more methods were only seven (cemA, matK, ndhF, rbcL, rpoA, rpoC2 and rps12) (Table 2).

Phylogenomic analysis

The ML phylogenomic tree based on 287 whole-plastome *Brachypodium* sequences, revealed strongly supported key relationships between different species and plastotypes (Figs. 3, 4 and 5). Most major lineages showed strong bootstrap support values above 90%, while branches with bootstrap values below 80% were considered weakly supported and collapsed in the visual representations. The whole-plastome *Brachypodium* tree, rooted with *Oryza sativa* and *Diarrhena obovata*, showed the splitting of the crown node into two strongly supported sister clades corresponding, respectively, to the main S and D plastomes (Fig. 3). Bootstrap support analyses further highlighted that the major clades and subclades identified in the whole-plastome tree (deep branches) were also present in the CDS- (76 concatenated genes) and the Positive selection- (35 concatenated genes) based trees, as well as in any of the 35 independent gene trees which are potentially under positive selection (Table 2), including some of the most widely used in plastome phylogenies of angiosperms (e. g., matK, ndhF, rbcL, rpoC2) (Figs. 3, S5). However, at shallow branches, the CDS, positive selection, and independent gene trees lacked support or collapsed into polytomies for many lineages, as a result of population similarity and reduced plastomic variation within recent S- and D-groups (Fig. S5A). Pairwise topological congruence tests between the whole-plastome tree (H0) and different alternative trees

Table 2 *Brachypodium* plastid genes under potential positive selection identified by Tajima's D neutrality test and three evolutionary selection models: generalized site-based FEL (Fixed Effects Likelihood) model, episodic site-based MEME (Mixed Effects Model of Evolution) model, and generalized site-based FUBAR (Fast Unbiased Bayesian Approximation) model (see Tables S7, S8, S9)

Gene	Tajima's D	FEL	MEME	FUBAR	# Methods
atpA	1.620	No	No	**All (1)**	**1**
atpB	**2.20***	No	No	No	**1**
atpE	0.250	No	No	No	0
atpF	**2.600***	No	No	No	**1**
atpH	−0.790	No	No	No	0
atpI	**2.120***	No	No	No	**1**
ccsA	**2.860***	No	No	No	**1**
cemA	0.410	No	**All (1)**	**All (1)**	**2**
clpP	1.570	No	No	No	0
infA	**2.920***	No	No	No	**1**
matK	1.600	**No**	**All (1)**	**All (1)**	**2**
ndhA	**2.710***	No	No	No	**1**
ndhB	**2.710***	No	No	No	**1**
ndhC	**3.380***	No	No	No	**1**
ndhD	1.530	No	No	No	0
ndhE	1.070	No	No	No	0
ndhF	**2.960***	No	**All (2)**	**All (1), *B. stacei* (1)**	**3**
ndhG	**2.280***	No	No	No	**1**
ndhH	1.840	No	No	No	0
ndhI	**2.260***	No	No	No	**1**
ndhJ	0.790	No	No	No	0
ndhK	**3.250***	No	No	**All (1)**	**1**
petA	−0.150	No	No	No	0
petB	**3.360***	No	No	No	**1**
petD	−0.030	No	No	No	0
petG	0.000	No	No	No	0
petL	0.000	No	No	No	0
petN	0.000	No	No	No	0
psaA	2.070	No	No	No	0
psaB	1.600	No	No	**All (1)**	**1**
psaC	**2.600***	No	No	No	**1**
psaI	1.920	No	No	No	0
psaJ	1.920	No	No	No	0
psbA	−0.740	No	No	No	0
psbB	**2.830***	No	No	No	**1**
psbC	1.080	No	No	No	0
psbD	1.870	No	No	No	0
psbE	0.710	No	No	No	0
psbF	0.000	No	No	No	0
psbH	**3.050***	No	No	No	**1**
psbI	0.000	No	No	No	0
psbJ	1.920	No	No	No	0
psbK	0.640	No	No	No	0
psbL	1.920	No	No	No	0
psbM	0.000	No	No	No	0
psbN	0.000	No	No	No	0
psbT	1.870	No	No	No	0
psbZ	0.000	No	No	No	0
rbcL	**2.690***	No	No	**All (2)**	**2**
rpl14	**3.030***	No	No	No	**1**
rpl16	**2.650***	No	No	No	**1**

Table 2 (continued)

Gene	Tajima's D	FEL	MEME	FUBAR	# Methods
rpl2	1.920	No	No	No	0
rpl20	**2.800***	No	No	No	**1**
rpl22	0.970	No	No	No	0
rpl23	0.680	No	No	***B. stacei* (1)**	**1**
rpl32	2.090	No	No	No	0
rpl33	0.130	No	No	No	0
rpl36	1.920	No	No	No	0
rpoA	**3.140***	No	**All**	No	**2**
rpoB	**2.220***	No	No	No	**1**
rpoC1	2.030	No	No	No	0
rpoC2	**3.310***	No	**All (1)**	No	**2**
rps2	**3.740***	No	No	No	**1**
rps3	**2.990***	No	No	No	**1**
rps4	0.870	No	No	No	0
rps7	**2.600***	No	No	No	**1**
rps8	2.070	No	No	No	0
rps11	1.570	No	No	**All (1)**	**1**
rps12	−2.050*	No	No	No	0
rps14	**2.600***	No	No	No	**1**
rps15	0.630	No	No	No	0
rps16	1.120	No	No	No	0
rps18	2.080	No	No	No	0
rps19	0.980	No	No	No	0
ycf3	**2.120***	No	No	No	**1**
ycf4	**2.180***	No	No	No	**1**

The table reports the results for all 76 genes and the affected groups (*B. stacei*, *B. distachyon*, *B. hybridum* S, *B. hybridum* D) or all species across the phylogeny (All). Significant positive and negative Tajima's D values indicate positive and purifying selection, respectively. The number of codons affected by positive selection is shown in parentheses for each model. Genes identified under positive selection by one or more methods (# of methods) are highlighted in bold

(H1: CDS, positive selection, independent gene trees) based on Kishino-Hasekawa (KH), Shimodaira-Hasegawa (SH), and Approximately Unbiased (AU) approaches for the 287 samples studied revealed that all alternative topologies were significantly incongruent except those from the CDS and Positive selection data sets for KH (Table S10A). However, tests performed with pruned trees computed for 12 samples, leaving only a single representative tip per major lineage (Fig. S5B), showed that the CDS and Positive-selection topologies, and those of some independent genes (e. g., ndhF, rpoC2) were significantly congruent with that of the whole-plastome tree for all KH, SH and AU tests (Table S10B).

The separate phylogenies of the D and S groups yielded different evolutionary scenarios for the plastomes of the progenitor species and their *B. hybridum*-derived plastotypes (Figs. 4 and 5). Within clade D, the ancestral lineage D, present exclusively in all *B. hybridum*-D plastotypes, diverged early from the common ancestor, and the remaining lineages of the *B. distachyon* progenitor showed the consecutive splits of the recently evolved lineages Sicily, EDF+, EDF+Spain, S+ and T+, although the Sicilian samples (CSR_6/CSR_7 and LPA3.2) were sister to two separate EDF+ subclades (Figs. 3 and 4). In contrast, within clade S, lineages of the two taxa diverged recently within the same subclades, and those of the *B. hybridum*-S plastotypes were likely derived from those of the progenitor *B. stacei* (Figs. 3 and 5). The topology of clade S represented consecutive splits of a more ancestral lineage *B. stacei* + *B. hybridum*-S from the Canary Islands (CI), Balearic Islands (BI) and Spain+, followed by those of *B. stacei* + *B. hybridum*-S W Med I+, *B. stacei* + *B. hybridum*-S E Med I (Israel)+, *B. hybridum*-S E Med II, B. stacei W Med II, *B. stacei* + *B. hybridum*-S W-E Med I, and *B. stacei* + *B. hybridum*-S W-E Med II, but with several geographical mismatches (Figs. 3 and 5). All *B. hybridum*-S samples from the invaded regions belonged to the more recently evolved lineage *B. stacei* + *B. hybridum*-S W-E Med II (Fig. 5); however, the low resolution of this clade prevented the identification of the geographic origins of its ancestors. A Discriminant Analysis of Principal Components (DAPC) of plastomes, which captured more than 90% of the variance in the first five axes for the D plastomes and up to the first twenty axes for S plastomes, recovered clusters for the D and S plastome that were consistent with clades in the phylogenomic tree, based on Ward and AIC statistics (Figs. S6A, S6B). Similarly, statistical parsimony haplotype networks constructed

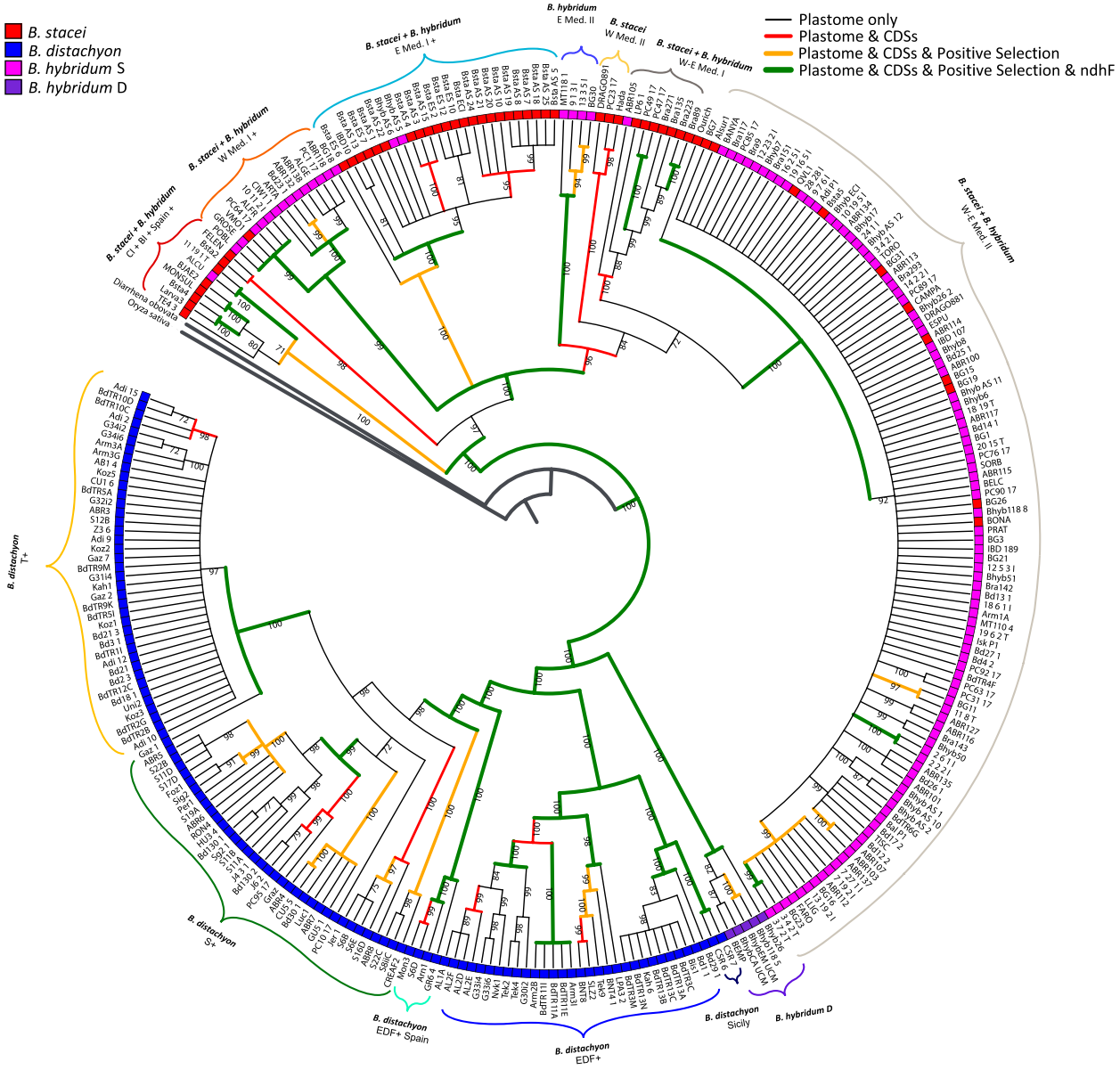

Fig. 3 Maximum likelihood (ML) plastome phylogenomic tree based on complete plastome sequences of 287 annual *Brachypodium* species and plastotypes, using *Diarrhena obovata* and *Oryza sativa* as outgroups. Colors next to the tips indicate species and plastotypes (*B. distachyon*, blue; *B. stacei*, red; *B. hybridum* D, mauve; *B. hybridum* S, purple). Thick green branches indicate congruent topologies in four plastome trees [Whole-Plastome (1), CDSs (2), Positive selection (3), ndhF (4)], yellow in three plastome trees (1–3), and red in two trees (1–2). Branches with bootstrap support <70% were collapsed. The abbreviations below the tip labels indicate the species/plastotype identity and assignment to the geographic or nuclear group, based on previous phylogenomic studies [3, 46, 47, 63]. CI = Canary Islands; BI = Balearic Islands; WMed = Western Mediterranean; EMed = Eastern Mediterranean; EDF = Extremely Delayed Flowering Group; S = Spain Group; T = Turkey Group; for any group + indicates more others

separately for D and S plastomes reflected the respective phylogenomic subtrees (Figs. 4 and 5), differentiating haplotype groups that corresponded to DACP clusters. However, while in clade D haplotype network groups and DACP clusters matched the constitutions of main lineages of the subtree and were supported by many shared branch mutations and by long DACP 1 and 2 plot distances (Figs. 4, S6A), in clade S the DAPC cluster 3 was not completely consistent with the lineages of the subtree and amalgamated several haplotype groups as they were separated by short DACP plot distances (Figs. 5, S6B).

When the topological divergences of the major lineages of the D and S plastome subtrees were contrasted with those obtained from nuclear-based trees for the same samples [3, 17, 46, 47] several discordances were observed in the D and S groups. Within the best-studied D group, polyphyletic samples from Sicily in the plastome tree (pt) were resolved as monophyletic and belonging to

Table 3 Analysis of potential (gene-wide) positive selection on *Brachypodium* plastid genes according to the Branch-site Unrestricted Statistical Test for Episodic Diversification (BUSTED) model

BUSTED		*Brachypodium*		S lineage		*B. distachyon*		*B. hybridum* D	
Model	**Selection Mode**	**dN/dS**	**Proportion, %**	**dN/dS**	**Proportion, %**	**dN/dS**	**Proportion, %**	**dN/dS**	**Proportion, %**
Constrained model dN/dS > 1 not allowed	Negative selection	0.000	80.536	0.000	81.012	0.063	56.431	0.205	91.303
	Negative/Neutral evolution	1.000	15.861	0.001	0.000	0.125	34.170	0.859	2.234
	Neutral evolution	1.000	3.603	1.000	18.988	1.000	9.400	1.000	6.462
Unconstrained model dN/dS > 1 allowed	Negative selection	0.056	71.956	0.000	0.975	0.141	79.111	0.198	93.453
	Negative evolution	0.516	28.038	0.000	90.086	0.186	20.720	0.886	0.911
	Diversifying selection	34133.479	0.006	2.309	8.939	15.346	0.169	1.411	5.556
Likelihood Ratio Test		*P*-value = 0.000		*P*-value = 0.090		*P*-value = 0.070		*P*-value = 0.500	

The selection modes identified by BUSTED, based on the constrained model (dN/dS > 1 not allowed; negative or neutral selection mode), and the unconstrained model (dN/dS > 1 allowed; negative, neutral, or diversifying selection mode), were tested on the whole-plastome *Brachypodium* tree (see Fig. 3). We tested potential positive selection on the full topology (*Brachypodium*) and for specific monophyletic clades of the tree [S lineage (*B. stacei* + *B. hybridum*-S), *B. distachyon*, *B. hybridum*-D] that were treated, respectively, as foreground branches, while the remaining clades were considered as background branches. For each group, the dN/dS ratio and the proportion of branches under the particular selection mode are indicated. The test compares both models (constrained vs. unconstrained) through a likelihood ratio test (LRT); if the unconstrained alternative model is significantly better than the constrained null model (p-value < 0.05), positive selection is statistically detectable

an early divergent lineage of *B. distachyon* in the nuclear tree (nt), several samples nested within the Sicily-EDF+ clade in pt were nested within the T+ clade in the nuclear topology, samples from the EDF+Spain clade, sister to the Sicily-EDF+ clade in pt, were sister to the T+S+ clade in nt, and several samples of the sister S+ and T+ clades in pt were nested within the other clade in the nuclear topology (Fig. 4; [17]). Within the less explored group of S plastomes, several discordances between plastid and nuclear phylogenies were detected, indicating possible cases of chloroplast capture. In particular, *B. stacei* samples from the Canary Islands, which form a single, well-defined clade in nt, fell into two divergent clades *B. stacei+B.hybridum* CI+BI+Spain+ and *B.stacei* W Med II in pt, samples from Sicily and the Balearic Islands belonging to the ancestral western Mediterranean B. stacei BI+SI clade in nt were nested within the *B.stacei+B. hybridum*-S E Med I and the recently evolved *B.stacei+B. hybridum*-S W-E Med II clades in pt, and those from a young Ibero-Maghrebian clade in nt were nested within the separate B. stacei CI+BI+Spain+, *B.stacei* W Med II and B. stacei + B. hybridum-S W-E Med II clades in pt (Fig. 5; [3]). Similarly, *B. hybridum*-S samples from the eastern Mediterranean of the B. *stacei+B.hybridum*-S E Med (Israel) I+ clade in nt were nested within the *B.stacei+B.hybridum*-S W-E Med II clade in pt (Fig. 5, [47]).

Discussion

Advances over previous studies of *Brachypodium* plastomes

Our comprehensive study expands on previous work [63] by incorporating broader sampling across the native circum-Mediterranean range, using a panplastomic framework that integrates heteroplasmy detection, codon usage analysis, and positive selection tests. Unlike previous analyses focused primarily on plastome diversity and phylogeny of *B. distachyon*, our study provides a lineage-level comparison between the plastotypes of *B. distachyon*, *B. stacei*, and *B. hybridum*, capturing the deep and shallow evolutionary processes that shape plastome variation. Importantly, we provide new insights into the role of heteroplasmy and chloroplast capture in recent and ancient lineages, and we assess the functional evolutionary pressures acting on plastid genes

Features and evolution of the panplastome of the annual *Brachypodium* complex species

Our comparative genomics, statistical and phylogenomic analysis of 287 *Brachypodium* plastomes developed this study have provided a comprehensive framework on the constitution and evolutionary dynamics of the panplastome of the three annual *Brachypodium* species, a model system for allopolyploidy. Our extensive sampling has covered the geographic distributions of the three species and plastotypes (Table S1). All the sequenced plastomes are classified within the four main S- (*B. stacei*, *B. hybridum*-S) and D- (*B. distachyon*, *B. hybridum*-D) plastome groups of *Brachypodium* (Fig. 1, Table S2). Although the plastomes of the *B. hybridum*-derived plastotypes were similar to those of the respective progenitor species, they tended to be larger (Fig. 1, Table S2), due to insertions (Fig. S2, Tables 1, S3), which could be related to their greater tendency towards heteroplasmic introgression [63].

Interestingly, although the length of S plastomes is longer than that of D plastomes, especially in the LSC region containing the ~1,1 Kbp insertion (Fig. 1, Table 1), diversity rates are higher in D plastomes of the *B. hybridum*-D lineage for all parameters analyzed after normalization

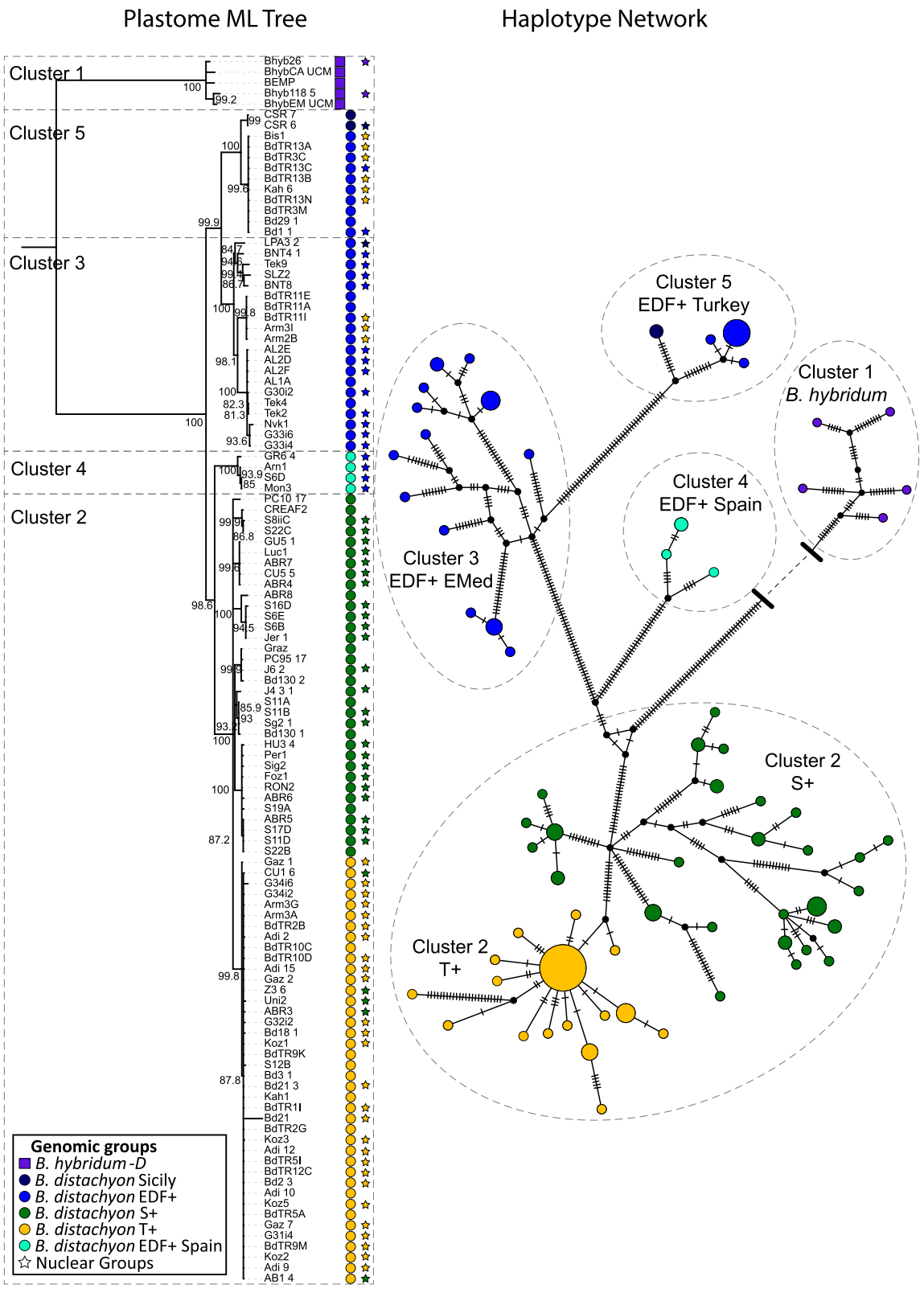

Fig. 4 Maximum likelihood (ML) phylogenomic subtree of *Brachypodium* D-plastomes (see Fig. 3) showing lineage divergences. A statistical haplotype network is shown next to the subtree. Color codes for tree lineages and network groups are shown in the chart. DAPC groups (see Fig. S6A) are indicated for tree clades and network groups. Symbols, circle: *B. distachyon* plastome, square: *B. hybridum* D plastome, star: *B. distachyon* and *B. hybridum* D nucleus. Color discrepancy between circle/square and star represents topological cytonuclear discordance, suggesting chloroplast capture. Bootstrap support values greater than 80 are indicated on branches

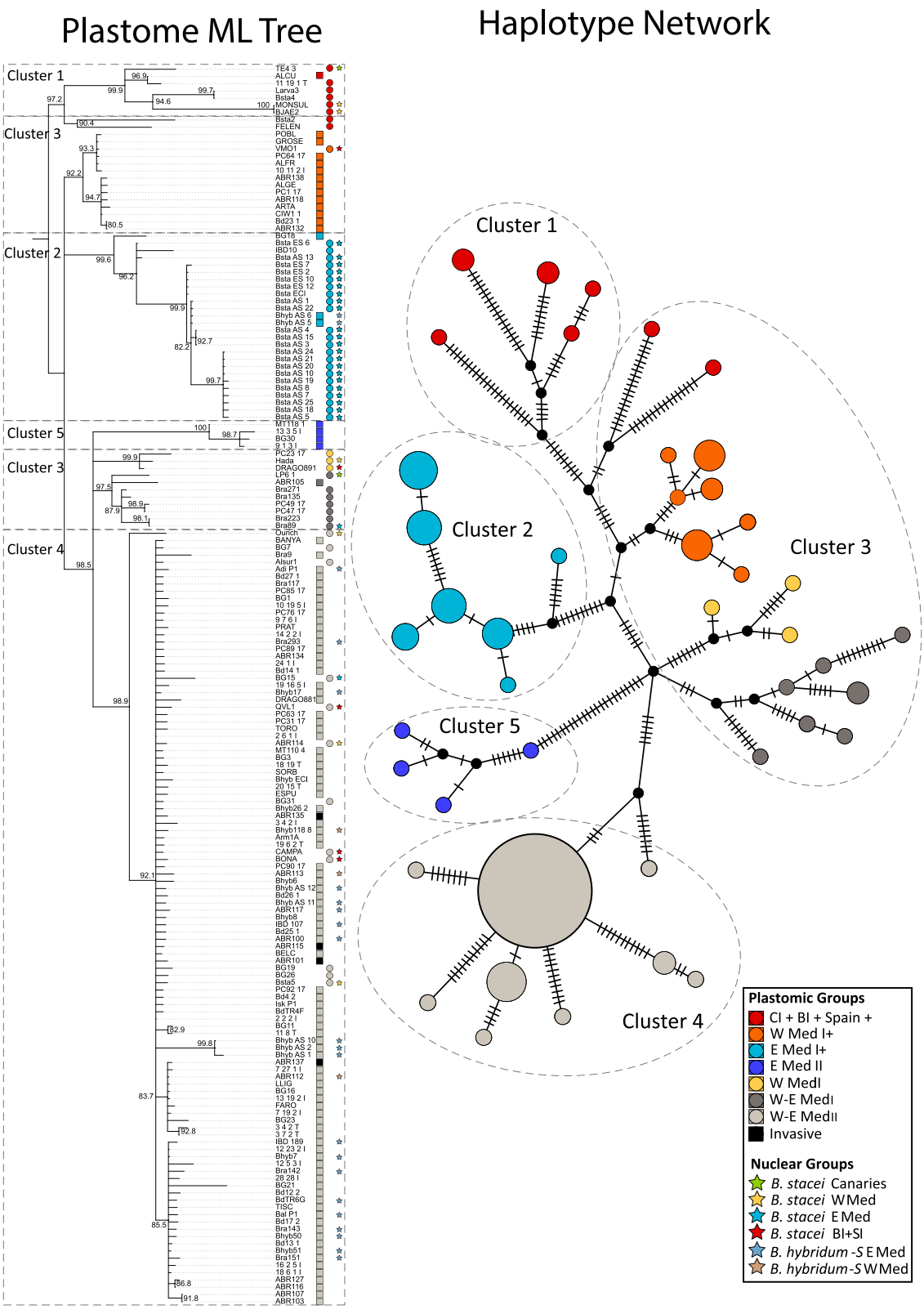

Fig. 5 Maximum likelihood (ML) phylogenomic subtree of *Brachypodium* S-plastomes (see Fig. 3) showing lineage divergences. A statistical haplotype network is shown next to the subtree. Color codes for tree lineages and network groups are shown in the chart. DAPC groups (see Fig. S6B) are indicated for tree clades and network groups. Symbols, circle: *B. stacei* plastome, square: *B. hybridum* S plastome, star: *B. stacei* and *B. hybridum* S nucleus. Color discrepancy between circle/square and star represents topological cytonuclear discordance, suggesting chloroplast capture. Bootstrap support values greater than 80, considered strong support, are indicated on branches

tests for sampling size (Table S2). Five wild samples from central and southern Iberian Peninsula that share this plastome-type, each present a different plastome (Tables S1, S2). The *B. hybridum*-D plastotype constitute the most ancestral radiation of the three annual *Brachypodium* taxa (1.4 Ma; [17, 65]). Our large sampling efforts indicate that this ancestral group is restricted to a reduced distribution area in the western Mediterranean region (Fig. S1). The few wild *B. hybridum*-D plastotypes are the only ones that show inheritance of D-plastomes from *B. distachyon*-like maternal parents, an unusual occurrence compared to the predominant inheritance of S-plastome from *B.stacei*-like maternal parents observed in the remaining 120 wild *B. hybridum*-S plastotypes (Fig. S1, Tables S1, S2). The high differentiation of *B. hybridum*-D plastomes is consistent with that of its nuclear subgenomes (D and S) that diverged much earlier than the respective genomes of the current progenitor species (*B. distachyon*, *B. stacei*) and the recently derived *B.hybridum*-S subgenomes [3, 17, 47]. Furthermore, the ancestral *B. hybridum*-D genome showed a unique nuclear 5S rDNA family and a trend towards reduction (diploidization) in rDNA genes, moving away from the additive pattern of distachyon-type and staceitype 35S and 5S rDNA genes observed in the recent *B. hybridum*-S reference genome [73]. Therefore, the special plastome features of the *B. hybridum*-D plastotypes are consistent with those of their nuclear subgenomes, suggesting that they have arisen as a consequence of evolution since the hybridization and WGD, or were acquired from unknown (and probably extinct) progenitor species different from extant *B. distachyon* and *B. stacei* species.

The patterns of SNP diversity observed among plastomes of the annual *Brachypodium* species reflect complex evolutionary processes. While *B. distachyon* plastomes exhibit a higher total SNP count (429) compared to *B. stacei* (260) (Table S2), the normalized SNP count per individual (3.9 for *B. distachyon* vs. 4.9 for *B. stacei*) may be influenced by differences in population dynamics and sampling strategies. The wider ecological range and greater geographic diversity associated with *B. distachyon* [17, 38] suggest that its plastome diversity is distributed across a broader population pool, diluting the average diversity per individual. In contrast, *B. stacei*, with a more restricted distribution, may exhibit localized diversity hotspots that are more apparent in smaller population samples. The lower diversity of *B. hybridum*-S plastotypes (202 total SNPs, 1.8 SNPs per individual) compared to both progenitor species reflects its recent maternal inheritance from *B. stacei* and its limited evolutionary time for divergence from its origin [47]. In contrast, *B. hybridum*-D plastotypes show a unique pattern, with the lowest total SNP count (52) but the highest normalized SNP count per individual (10.4). This disparity aligns with their highly localized distribution and small sample size, indicating considerable diversity of the ancestral population that likely experienced bottleneck effects and potential selective pressures acting on these plastomes. These patterns are consistent with the hypothesis that *B. hybridum*-D plastotypes represent an ancient lineage subject to prolonged isolation and local adaptation. Overall, these findings reveal a balanced interplay of evolutionary forces shaping plastome diversity within the *Brachypodium* complex taxa. The relatively higher normalized SNP diversity in *B. stacei* plastomes highlights the influence of population structure and local adaptation, while the broad distribution and ecological flexibility of *B. distachyon* likely contributes to its overall diversity [3, 38]. In the case of *B. hybridum*, the contrasting patterns between the S and D plastotypes emphasize the combined effects of age and ancestral diversity, population bottlenecks, and local selection on plastome evolution in allopolyploid lineages [47, 65].

Analysis of microsatellite SSRs revealed the presence of both common and exclusive repeat motifs. These patterns reflect the overall SSR abundance, with *B. stacei* (37.1 ± 1.53) and *B. hybridum-S* (36.1 ± 0.57) plastomes showing significantly higher SSR counts than *B. distachyon* (31.80 ± 1.55) and *B. hybridum-D* (32.0 ± 0.71) plastomes (Fig. 2, Table S5), supporting the higher haplotype diversity of S plastomes (excluding *B. hybridum*-D) (Table S2). Some SSRs were unique to specific *Brachypodium* plastome groups (e.g., *B. distachyon*: ACT; *B. hybridum*-D: AAGGGT; *B. hybridum-S*: AAATT), albeit at low frequencies, whereas the predominant poly-A monomers were more abundant in S plastomes (Table S6), contributing to the genetic distinction between them.

Amino acid frequencies in the 76 protein-coding genes and the Relative Synonymous Codon Usage (RSCU) analyses provided insights into the amino acid composition and codon preferences of plastid genes among the analyzed *Brachypodium* plastomes. *B. distachyon* plastid genes showed distinct amino acid profiles and different preferences in codon usage compared to those in the other three plastome groups (Table S6). For example, *B. distachyon* protein plastid genes had higher proportions of C, F, H, N, R, S, W, and Y and lower proportions of A, E, I, K, L, M, P and V than protein genes in the other three groups. Interestingly, the amino acid content of the *B. hybridum*-D genes was more similar to that of S plastomes than to that of its D-type plastome (Table S6). This suggests possible recent mutations in the amino acid composition of protein-coding genes in the common ancestor of the *B. distachyon* clade, which was estimated to have split approximately 1 Ma ago [17]. *B. distachyon* genes also preferentially use the stop codon UGA over UAG, which was used by genes from the other three groups, GCA (encoding A) over GCU, and GUU

(encoding V) over GUA. Furthermore, plastid genes from D plastomes (*B. distachyon* and *B. hybridum*-D) showed a preference for ACA (encoding threonine) and CCA (encoding proline), whereas those from S plastomes (*B. stacei* and *B. hybridum* S) preferred ACU and CCU, respectively (Table S6). Amino acid content and codon usage bias can influence protein configuration and gene expression efficiency, and potential adaptive responses to environmental conditions [45, 57]. *B. distachyon* shows the highest niche breath of the three species, and a better adaptation to mesic climates than *B. stacei* or *B. hybridum* [38]; however, the potential adaptive advantages of its plastid proteins over the others remain to be evaluated. By contrast, the conservation of codon usage patterns in different chloroplast genomes suggests an evolutionary advantage in maintaining essential chloroplast functions [13, 57].

Heteroplasmy of the annual *Brachypodium* plastomes

The rare occurrence of heteroplasmy in angiosperms, caused by biparental inheritance or introgressive hybridization, is explained by genomic conflict between the two plastomes that results in the sorting out of homoplasmy in few generations, while the prevalence of maternal plastome inheritance may respond to the excess of mutation load in the paternal plastomes compared to maternal ones [11, 66]. Despite its rarity, heteroplasmy has been detected in some highly hybridogenous angiosperm groups (e. g. *Citrus*; [5] and also in the ancestral *B. hybridum* D plastotype [17]. Our extensive sampling of 287 annual *Brachypodium* plastomes has identified specific heteroplasmic patterns in the major plastome groups (Tables 1, S4; Appendix 1) that may reflect introgressions or mutations. The high divergence of *B. hybridum*-D plastome sequences from the remaining D sequences of *B. distachyon* (Fig. S5) is accompanied by their higher number of insertions, larger insertion sizes, and higher rates of heteroplasmy per individual (Tables 1, S3, S4). Within the two allotetraploid *B. hybridum* plastotype types, *B. hybridum*-D constituted the plastome type (D) with more foreign content from the other plastome type (S). This heteroplasmy, apparently caused by introgressive hybridization, may be a consequence of their ancestral origin (~1.4 Ma; [17], as these individuals have had more time for interplastomic recombinations than younger *B. hybridum*-S plastotypes (0.14–0.13 Ma; [17, 47] that show less foreign content of D plastomes (Tables 1, S3, S4). Notably, introgressive hybridization appears to have preferentially transferred sequences from S plastomes to D plastomes in *B. hybridum*-D, but not vice versa (Table S3). This asymmetry could reflect differences in recombination efficiency [52, 68], plastome stability [21, 22], or selective pressures favouring S-derived insertions into D plastomes [1]. However, although the observed patterns are consistent with introgressive hybridization and possible plastid recombination, the detected heteroplasmic diversity may also reflect retained ancestral plastomic polymorphism or results of chloroplast capture events, which are known to occur in hybridogenous species [5, 13] as alternative explanations.

In addition to heteroplasmy arising from introgressive hybridization and recombination, our analyses identified significant heteroplasmic patterns across all plastome groups, which can be attributed to mutations and other molecular mechanisms (Tables 1, S4; Appendix 1). These patterns suggest that heteroplasmy in *Brachypodium* plastomes also includes mutational heteroplasmy caused by somatic mutations or errors during plastome replication and repair [19, 66]. Interestingly, heteroplasmic diversity also revealed marked differences between plastome types. S plastomes showed a substantially higher number of unique heteroplasmic positions (443) compared to D plastomes (22) (Table S4). This asymmetry could be a result of intrinsic differences in plastome size and mutation rates, as S plastomes are characterized by a larger LSC region, which may facilitate a higher probability of mutational activity [22]. A relatively balanced percentage of these heteroplasmic positions occurred in coding and non-coding regions, with coding regions accounting for 51.11% in *B. stacei*, 48.26% in *B. hybridum*-S, 49.50% in *B. distachyon*, and 42.79% in *B. hybridum*-D. These values indicate that the proportion of heteroplasmic positions between coding and non-coding regions is relatively even across plastomes, without a clear bias toward one category. Similarly, the overall distribution of heteroplasmic positions between S and D plastomes follows a comparable pattern, suggesting that heteroplasmy is widespread but does not strongly differentiate between plastome types in terms of coding vs. non-coding impact (Table S4). Furthermore, the limited heteroplasmic diversity in D plastomes might reflect a combination of a higher mutation-purging capacity or stronger purifying selection for heteroplasmic positions in coding regions that may have occurred over a longer time span since their respective divergences from their common ancestor (*B. hybridum*-D, ~1.4 Ma; *B. distachyon*, ~1 Ma; [17] compared to the recent radiation of the crown group *B. stacei* (~0.256 Ma; [3] and its derived lineage *B. hybridum*-S (~0.13 Ma; [47]. Regarding the comparative analysis of plastome compatibility of *B. hybridum* S and D plastotypes and their parental species, the percentages of SNV and heteroplasmic positions matches revealed that approximately half of the positions did not align with those of either parent in both cases and thus could have resulted from somatic mutations (Table 1C). Furthermore, while the percentages of matches of *B. hybridum*-S were slightly higher with the maternal parent (*B. stacei*), those of *B. hybridum*-D were higher with the

paternal parent (*B. stacei*), underlining the complexity of heteroplasmy in the evolution of these *Brachypodium* plastomes. The functional implications of these heteroplasmic positions remain an open question, in particular in coding regions that could be under selective pressures. While previous studies identified the existence of heteroplasmy in a limited number of *B. hybridum*-D samples [17], our expanded sampling uncovers heteroplasmic patterns across all plastome groups, indicating a broader prevalence of plastid introgression and somatic mutation effects in *Brachypodium*.

Phylogenomics reveal life-history traits and geographic structure within D and S plastomes coupled with long-distance dispersals and frequent chloroplast capture events

The increased sample size of this study has recovered new evolutionary relationships for annual *Brachypodium* species, mainly in the under-investigated S-plastome clade, and has strengthened those of the D-plastome clade with our extended sampling (Figs. 3, 4 and 5). Furthermore, the contrasting phylogenetic placement of several samples in the plastome-based D- and S-trees, compared to available data on nuclear-based trees, suggests widespread chloroplast capture events not only in the progenitor species *B. distachyon* and *B. stacei* but also in the *B. hybridum*-S plastotypes (Figs. 4 and 5; [3, 17, 46, 47, 63]).

The expanded whole-plastome phylogenomic tree of annual *Brachypodium* species and plastotypes revealed contrasting evolutionary patterns for the *Brachypodium* D- and S-plastome lineages (Figs. 3, 4 and 5). Within the ancient D-genome clade, most divergences are structured by flowering time (Fig. 4; [18, 17]) and geography, a typical feature of plastid genomes [20, 56, 76]. The early split of the ancestral lineage *B. hybridum*-D (~1.4 Ma; [17] is corroborated by new analyses and a few additional samples support its restricted distribution to the south-central Iberian Peninsula, constituting the historical remnants of this archaic and genetically isolated allotetraploid [65]. The highly divergent lineage *B. hybridum*-D is the only group that showed no evidence of chloroplast capture with any other clade, supporting its tendency towards speciation [17]. The *B. distachyon* (~0.93 Ma, [63] lineages Sicily-EDF+-Turkey, Sicily-EDF+-East Med, EDF+-Spain, S+ and T+, supported by well-characterized haplotypes, showed general coherence with flowering time and geography. However, the few mismatches (e.g., Sicilian samples) and large cytonuclear topological discordances (Fig. 4) confirm previously reported cases of long-distance dispersals (LDD) followed by inter-clade crossovers and chloroplast captures across all lineages, extended here to samples from the ancestral Sicilian lineage showing more recent EDF+ plastomes (Fig. 4; [17]).

Within the recently evolved S genome clade (~0.25 Ma; [3]) (Fig. 5), the subtle geographic structure of the most ancestral *B. stacei* islands and the most recent western and eastern Mediterranean lineages, and their derived western and eastern *B. hybridum*-S lineages (~0.14–0.13 Ma; [17, 47], corroborate the independent origins of these two young allotetraploids from their respective local maternal progenitor species [47]. However, in most S lineages there is evidence of frequent LDD to the other region, obscured by the low resolution of the most recent clades with spatially separated samples sharing identical plastomes (Fig. 5; Table S2). Our analyses detected several cases of cytonuclear discordance in the S lineages, particularly in *B. stacei* samples from the Canary, Baleric and Sicily Islands, and from the East Mediterranean (Israel), and in *B. hybridum*-S samples from East Mediterranean (Israel), where plastid phylogenies deviate from nuclear-based classifications (Fig. 5; [3, 46]). This pattern corroborates the existence of chloroplast capture events not only in *B. stacei*, as previously reported [3], but for the first time in *B. hybridum*-S. These plastid captures were observed mostly in Eastern (*B. hybridum*-S East I+) Mediterranean lineages where several individuals from Israel cluster with geographically distinct western Mediterranean lineages despite their nuclear affinity to a different maternal group (Fig. 5). The plastid captures resulting from repeated introgressions and random backcrosses could be highly feasible for these genomically close individuals of the S clade. The expanded plastome phylogenies presented here improve on previous frameworks [63], resolving new clades within the S plastome lineage and confirming widespread chloroplast capture events in both progenitor and hybrid species.

Positive selection, functional adaptation and phylogenetic signal

The existence of positive selection in plastid genes of Poaceae has been mentioned in previous studies, which showed that it expanded in ancestral grasses [23] and affects a variable number of genes in the plastomes of current grasses [34, 61, 55]. The number of plastid genes under positive selection varies mainly according to the phylogenetic depth of the sampling, from 65 loci in representatives of main Poaceae lineages [61], to 25 in representatives of the tropical PACMAD clade [55], and up to 11 in representatives of temperate subtribes Stipeae-Ampelodesmae [34]. Within these genes, the number of codons showing significant diversifying selection also varied depending on the phylogenetic breath and taxon sampling, being higher at the family level and smaller at tribal level (e. g., rbcL: 47 codons in Poaceae, 10 in PACMAD, 5 in Stipeae; [34, 55, 61].

Our study of 287 plastomes provides further evidence for the possible existence of this phenomenon in annual

Brachypodium species and plastotypes, uncovering signatures of positive selection in up to 29 genes, based on the Tajima D neutrality test, and up to 9 genes, based on site and branch site evolutionary models, summarizing 35 protein-coding genes (Tables 2, S7, S8, S9) detected in previous grass studies. By employing various estimates (including Tajima's D, BUSTED, MEME, FEL and FUBAR) we detected evidence for episodic or pervasive positive selection, or significant deviation from neutrality towards balancing selection (Tables 2 and 3, S7, S8, S9). However, alternative approaches yielded different results. Tajima's D detected the highest number of genes under potential positive selection (29) considering all *Brachypodium* plastomes studied (Tables 2, S7). However, most of these genes evolved neutrally or even under purifying selection at the species or plastotype level (Table S7). These differences may have been caused by the artificial *Brachypodium* "population", which includes samples from two highly divergent and non-sister lineages, *B. stacei* and *B. distachyon* [64]. The deviation from neutrality of Tajima's D test for the 29 loci could have been caused by an excess of intermediate-frequency alleles, thus favoring balancing selection. Of the four evolutionary models tested, the FEL model did not detect any codon sites under significant positive selection for any loci and group, and only marginally for one codon (matK 203) across all plastomes ($p = 0.08$; Table S8), whereas BUSTED detected a very low proportion of branches under positive selection across the entire *Brachypodium* tree (<1%; Table 3). MEME detected five codon sites of cemA (97), matK (204), ndhF (642, 651), and rpoC2 (1482) affected by positive selection, and FUBAR nine (atpA 119, cemA 97, matK 204, ndhF 642, ndhK 238, psaB 417, rbcL 43 and 95, rps11 37) (Tables 2, S9). Overall, evolutionary selection models only recovered weak evidence of positive selection in the *Brachypodium* phylogeny and groups (Tables 2 and 3, S9), highlighting some common genes and codon sites that experienced it [matK (204), ndhF (642); (Tables 2 and 3)]. Our results agree with previous studies of grass plastomes that detected the highest number of codons under significant positive selection in the genes matK, ndhF, rbcL and rpoC2 (Table 2; [34, 55, 61]) but also detected positive selection in rps11, only found before in Stipeae [34]. The positive selective pressure detected at these codons has been attributed to functional adaptive innovations [34, 55, 61]. Even if some of their encoded proteins perform crucial physiological functions, such as transcription (rpoC2) and translation (rps11) processes, splicing factor (matK), chlororespiration (ndhF, ndhK), and carboxylation-oxidation (rbcL), some of the positively affected amino acid changes (Tables 2, S9) are not directly related to their catalytic efficiencies, or their functions have not been clearly determined yet.

It has been argued that positive selection acting on these plastid genes may compromise their value for phylogenetic reconstruction [61]. We assessed the extent of positive selection on the phylogenetic signal of these genes in our *Brachypodium* evolutionary framework by comparing the null hypothesis of the whole-plastome neutral tree with alternative topologies based on the 76 concatenated protein-coding loci (CDS), the 35 concatenated loci under potential positive selection, and each of these 35 individual genes, using KH, SH, and AU topological congruence tests (Table S10) and topological contrast (Figs. 3, S5). The more robust AU test discriminated any alternative topologies as incongruent at shallow branches, while it revealed that CDS, positive selection, and ndhF and rpoC2 trees were significantly congruent with respect to the whole-plastome tree at deep nodes (Table S10). These findings indicate that some plastid genes, discarded for phylogenetic reconstruction as potentially misleading [61], such as ndhF, recover topologies concordant with that of the full plastome tree (Table S10B, Fig. S5) for *Brachypodium* lineages as divergent as 10.2 Ma [17, 64]. The ndhF gene has been a classic barcode plastid gene for the phylogenetic reconstruction of several grass lineages, such as Poaceae [79] and Pooideae [6]. Recent phylogenomic reconstructions based on single-copy nuclear genes and whole plastome sequences have recovered the topology of most grass subfamilies and tribes, as previously established by ndhF analysis [24]. Despite the prevalence of positive selection, phylogenetic analyses based on these genes (e.g., matK, ndhF, rbcL) maintained congruent topologies at the S and D clades and species and plastotype lineages' levels with the whole-plastome phylogenomic tree (Figs. 3, S5B), although some incongruences were observed at the shallow individual level (Figs. 3, S5A).

Conclusions

Our extensive study of the *Brachypodium* panplastome provides the first lineage-level exploration of plastome diversity, structural evolution, heteroplasmy, and adaptive selection in three annual species and their plastotypes. Through exhaustive sampling across their native circum-Mediterranean range, we identify novel structural rearrangements and deletions, unique repeat motifs, substantial biases in amino acid composition and codon usage, and widespread heteroplasmy in the plastomes of the parental species *B. distachyon* and *B. stacei* and the D and S plastotypes of *B. hybridum*. While the ancestral plastomes of *B. hybridum*-D show evidence of heteroplasmy caused by introgressive hybridization and recombination, most of the heteroplasmic patterns in the other plastome groups can be attributed to somatic mutations. Analysis of topological discordances between plastome and nucleus —which were expected due to their distinct

allelic constitutions (haploid vs. diploid-polyploid) and inheritance patterns (maternal vs. biparental)— revealed, however, specific inconsistencies consistent with chloroplast capture in the recently evolved *B. distachyon, B. stacei*, and *B. hybridum*-S groups, while highlighting the genetic isolation of the ancestral clade *B. hybridum*-D. Positive codon selection was detected in nine plastid genes within *Brachypodium*, although topologies reconstructed from some of these genes (e.g., ndhF, rpoC2) showed phylogenetic signal, supporting its continued usefulness in evolutionary studies.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12870-026-08570-2.

Supplementary Material 1. Fig. S1: Geographical distribution of the 283 out of the 287 studied samples of annual *Brachypodium* species and plastotypes representative of its entire native Mediterranean distribution area. *B. distachyon* (blue; 2n=2x=10), *B. stacei* (red; 2n=2x=20), *B. hybridum* S (pink; 2n=4x=30), *B. hybridum* D (violet; 2n=4x=30). Four samples from invaded regions (South Africa, Uruguay, W Australia) are not represented. Fig. S2: Plastome insertions of *B. stacei*, *B. hybridum* S, *B. hybridum* D, and *B. distachyon* plastomes. Indel 1: 49 bp (start position 1,551); Indel 2: 27 bp (1,627); Indel 3: 69 bp (3,388); Indel 4: 20 bp (5,867); Indel 5: 132 bp (6,020); Indel 6: 31 bp (12,912); Indel 7: 14 bp (13,166); Indel 8: 251 bp (13,976); Indel 9: 16 bp (17,110); Indel 10: 18 bp (17,334); Indel 11: 18 bp (17,592); Indel 12: 19 bp (17,775); Indel 13: 122 bp (18,331); Indel 14: 14 bp (19,434); Indel 15: 23 bp (19,801); Indel 16: 11 bp (27,823); Indel 17: 19 bp (33,041); Indel 18: 30 bp (36,371); Indel 19: 34 bp (42,263); Indel 20: 24 bp (46,613); Indel 21: 15 bp (48,243); Indel 22: 56 bp (50,746); Indel 23: 39 bp (50,960); Indel 24: 16 bp (56,847); Indel 25: 1160 bp (56,915); Indel 26: 11 bp (60,387); Indel 27: 10 bp (63,659); Indel 28: 11 bp (66,082); Indel 29: 14 bp (105,803); Indel 30: 11 bp (105,855); Indel 31: 23 bp (105,978); Indel 32: 369 bp (137,038); Indel 33: 20 bp (137,387). Fig. S3: (A) Codon usage analysis of 76 protein-coding genes from plastomes of the annual *Brachypodium* species and plastotypes, showing variations for 20 amino acids and stop codons. (B) The histogram below each amino acid illustrates codon usage frequencies within group [*B. stacei* (Bs), *B. distachyon* (Bd), *B. hybridum* S (BhS), *B. hybridum* D (BhD)]. Codon colours in the column plot (B) correspond to those shown above (A). RSCU (Relative Synonymous Codon Usage) values are provided for each codon (see Table S6). *: stop codons; A: alanine; C: cysteine; D: aspartic acid; E: glutamic acid; F: phenylalanine; G: glycine; H: histidine; I: isoleucine; K: lysine; L: leucine; M: methionine; N: asparagine; P: proline; Q: glutamine; R: arginine; S: serine; T: threonine; V: valine; W: tryptophan; Y: tyrosine. Fig. S4: Correlation between nucleotide diversity (π) and Tajima's D values in the 76 plastid protein coding genes represented in a density plot. Fig. S5: Maximum likelihood trees constructed with IQtree for all 35 independent gene trees under positive selection (see Table 2) including (A) all sampled individuals (287) and (B) pruned to one individual per clade (12 samples). Branches showing bootstrap values <80 were collapsed. Fig. S6: Discriminant analysis of principal components (DAPC) for the D (A) and S (B) plastomes of *Brachypodium*, showing the total variance explained by each axis, the AIC value used to select the optimal number of clusters, and a DAPC plot presenting a minimum spanning network connecting the identified clusters.

Supplementary Material 2. Table S1: Whole plastome sequence data of the annual *Brachypodium* complex species included in this study. Information on Species, Accesion code, Location (including germplasm bank of seed origin), geographical coordinates, and sequenced accession number (SRA) is provided for each sample. Newly sequenced plastome data are highlighted in bold. Germplasm banks: ABERS (Aberystwyth University, UK), INIA (Instituto Nacional de Investigacion y Tecnologia Agraria, Spain); JGI (Joint Genome Institute, USA); LU (Lanzhou University); TAU (Tel Aviv University, Israel); UCM (Universidad Complutense de Madrid, Spain); UJA (Universidad de Jaen, Spain); UMASS (University of Massachusetts; USA); UZ (Universidad de Zaragoza, Spain). Table S2: Summary of plastome characteristics and diversity of the whole plastome, and Tajima's D values at 76 plastid concatenated CDSs, for species and plastotypes of the annual *Brachypodium* complex, normalized by sample size. Plastome size, GC content, SNP distribution, and haplotype diversity are given for each species and plastotype. Numbers in parentheses indicate values normalized per individual. Significance of Tajima's D value is indicated by asterisks (**, p<0.01). LSC: long single copy region; SSC: short single copy region; IRA: inverted repeat region A; IRB: inverted repeat region B. Table S3: Plastome insertions and deletions detected in the studied *Brachypodium* plastomes by comparative analysis with reference plastomes of closely related grasses (*Oryza sativa, Diarrhena obovata*). The table shows the indel number, start position, size (bp), indel type [insertion or deletion with respect to plastome types (S or D) and species and plastotypes (*B. stacei, B. distachyon, B. hybridum*-S, *B. hybridum*-D)] and sequence for each indel (see Fig. S2). Possible heteroplasmic indels caused by introgression are shown in bold. Table S4: Total number of heteroplasmic positions at the group level (*B. stacei*, *B. hybridum*-S, *B. distachyon*, *B. hybridum*-D), with the average rate per individual in parentheses, and for coding and non-coding regions, and the plastome type level (S and D). Table S5: Analysis of repetitive simple sequence repeats (SSR) in the studied *Brachypodium* plastomes. The location of repetitive elements in the four regions of the plastomes, the type of repetitive element (no. of nucleotides of the repeat unit), and the count of distinct SSR motifs (repeat unit types) are indicated for each group (species and plastotypes). Values correspond to estimates of mean ± SD. n, sampling size. Superscripts letters denote Tukey pairwise comparisons between groups for each category; groups with same letter do not differ significantly ($P < 0.05$). Table S6: Amino acid composition and codon usage data from the four *Brachypodium* plastome groups studied (species and plastotypes) showing the total percentage of the amino acid in the 76 plastid-coding genes (%) and, for each of the 64 different codons, the number of times each codon type was found (Count), its relative synonymous codon usage (RSCU) and its frequency (Proportion) in the 76 protein-coding genes analyzed. aa, amino acids. Table S7: Tajima's D neutrality test statistic for 76 plastid genes in the plastomes of the annual *Brachypodium* species under study. Tajima's D values are indicated for each gene and group (*Brachypodium*, *B. stacei*, *B. distachyon*, *B. hybridum* S, *B. hybridum* D). Negative values suggest purifying selection, whereas positive values indicate balancing selection. Genes and groups potentially under significant positive or negative selection are highlighted in bold and marker with an asterisk (*). n, sampling size. Table S8: Analysis of potential generalized selection based on codon sites of *Brachypodium* plastid genes according to the FEL (Fixed Effects Likelihood) model. Results are shown for each codon analyzed: genes and their codon sites, group (All, or specific species or plastotypes if identified individually), model selection parameters (Alpha, Beta), likelihood ratio test (LRT) values, type of selection (Positive or Negative) and corresponding p-values. Significant selection results at $p \leq 0.05$ are highlighted in bold. Table S9: Analysis of potential episodic codon-site-based positive selection of *Brachypodium* plastid genes according to the Mixed Effects Model of Evolution (MEME) model and of generalized codon-site-based selection according to the Fast Unbiased Bayesian Approximation (FUBAR) model. (A) MEME results include the gene, codon site, and the p-value resulting from the likelihood ratio test for each codon and group. Cases of significant positive selection ($p \leq 0.05$) are highlighted in bold. (B) FUBAR results show the gene, codon site, selection parameters (Alpha, Beta), and posterior probability (PP) values for positive selection. PP values >0.9 supporting positive selection are highlighted in bold. Table S10: Summary of pairwise statistical tests for topological congruence of *Brachypodium* plastome and plastid phylogenetic trees for (A) full tree (287 samples studied), and (B) pruned tree (12 samples, keeping a single representative for each of the major clades of the full plastome tree) (see Figs. 3, 4, 5, S5). Plastome: tree based on whole plastomes; CDSs: tree based on 76 concatenated coding genes; Positive selection: tree based on 35 coding genes under potential positive selection; and individual trees of each of these 35 genes. Log-likelihood scores (logL), differences in log-likelihood relative to the best tree (deltaL), bootstrap proportions (bp-RELL), expected likelihood weights (c-ELW), and p-values from Kishino–Hasegawa (p-KH), Shimodaira–Hasegawa (p-SH), and approximately unbiased (p-AU) tests. These tests compared alternative topologies (H1) to the reference topology of the Plastome tree (H0). P-values <0.05 indicate significant incongruence of the alternative hypothesis with the null hypothesis. Alternative topologies congruent

with the Plastome tree based on AU test are highlighted in bold.

Supplementary Material 3. Appendix 1. Heteroplasmic positions identified in the studied *Brachypodium* plastome alignment indicating the unique identifier of the sample (Accession), the start position of the alignment where the heteroplasmic position was detected (Start), the length of the heteroplasmic fragment (Length) containing the position, the major nucleotide at that position (RefNucl) (a dash indicates an insertion), the alternative nucleotide at that position (AltNucl), and the minor allele frequency of that heteroplasmic position (MinAF).

Authors' contributions

PC, JV and EP obtained funding for the project. PC and MC designed the study. PC collected samples. MC, EP, JV and PC conducted and supervised the analyses. PC and MC wrote the original draft, and all authors reviewed it.

Funding

This research was supported by the Spanish Ministry of Science and Innovation (Grants No. PID2019-108195GB-I00 and PID2022-140074NB-I00), and the Spanish Aragon Government-European Social Fund Bioflora (Grant No. A01-23R) to PC and EP. The work (proposal:https://doi.org/10.46936/10.25585/60001143) conducted by the U.S. Department of Energy Joint Genome Institute (https://ror.org/04xm1d337), a DOE Office of Science User Facility, is supported by the Office of Science of the U.S. Department of Energy operated under Contract No. DE-AC02-05CH11231. MCC was supported by a Spanish Ministry of Science and Innovation PhD fellowship.

Data availability

Plastome Illumina paired-end sequencing data generated for this project are available on the European Nucleotide Archive, project number PRJEB76383. Documents containing the Plastome sequences and custom scripts designed for the analyses performed in this study are available on Github (https://github.com/Bioflora/Brachy_panplastome).

Declarations

Ethics approval and consent to participate

Seed sampling from natural populations was carried out in accordance with applicable local, regional, and national legislation in the respective countries of origin. For collections made after October 12, 2014, access to genetic resources complied with the Nagoya Protocol on Access to Genetic Resources and the Fair and Equitable Sharing of Benefits Arising from their Utilization, including in European Union Member States where the Protocol is in force. The collectors (co-authors of this study) or collaborating researchers who provided seed material obtained the necessary collection permits and/or authorizations. Accessions from germplasm banks were obtained in accordance with the regulations and material transfer agreements of the respective institutions.
The species studied are neither protected nor endangered in the sampled regions, and sampling did not cover protected areas unless the corresponding authorization had been granted.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

[1]Departamento de Ciencias Agrarias y del Medio Natural, Escuela Politécnica Superior de Huesca, Universidad de Zaragoza, C/Carretera de Cuarte Km 1, Huesca E-22071, Spain
[2]Grupo de Bioquímica, Biofísica y Biología Computacional (BIFI, UNIZAR), Unidad Asociada al CSIC, Zaragoza, Spain
[3]U.S. Department of Energy Joint Genome Institute, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA

Received: 24 February 2025 / Accepted: 9 March 2026
Published online: 24 March 2026

References

1. Blazier CC, Guisinger MM, Jansen RK. Recent loss of plastid-encoded *ndh* genes within *Erodium* (Geraniaceae). Plant Mol Biol. 2011;76(3–5):263–72. https://doi.org/10.1007/s11103-011-9753-5.
2. Bushnell B. BBMap: a fast, accurate, splice-aware aligner. Berkeley: Lawrence Berkeley National Lab.(LBNL); 2014. https://sourceforge.net/projects/bbmap/.
3. Campos M, Pérez-Collazos E, Díaz-Pérez A, López-Alvarez D, Oumouloud A, Mur LAJ, Vogel JP, Catalán P. Repeated migration, interbreeding and bottleneck shaped the phylogeography of the selfing grass Brachypodium stacei. Mol Ecol 2024;e17513.
4. Cao Z, Yang L, Xin Y, Xu W, Li Q, Zhang H, et al. Comparative and phylogenetic analysis of complete chloroplast genomes from seven *Neocinnamomum* taxa (Lauraceae). Front Plant Sci. 2023;14:1–12. https://doi.org/10.3389/fpls.2023.1205051.
5. Carbonell-Caballero J, Alonso R, Ibañez V, Terol J, Talon M, Dopazo J. A phylogenetic analysis of 34 chloroplast genomes elucidates the relationships between wild and domestic species within the genus *citrus*. Mol Biol Evol. 2015;32(8):2015–35. https://doi.org/10.1093/molbev/msv082.
6. Catalán P, Kellogg EA, Olmstead RG. Phylogeny of poaceae subfamily pooideae based on chloroplast*ndhF* gene sequences. Mol Phylogenet Evol. 1997;8(2):150–66. https://doi.org/10.1006/mpev.1997.0416.
7. Catalán P, López-Álvarez D, Bellosta C, Villar L. Updated taxonomic descriptions, iconography, and habitat preferences of *Brachypodium distachyon*, *B. stacei*, and *B. hybridum* (Poaceae). An Jard Bot Madrid. 2016;73(1):e028. https://doi.org/10.3989/ajbm.2428.
8. Catalán P, Müller J, Hasterok R, Jenkins G, Mur LAJ, Langdon T, et al. Evolution and taxonomic split of the model grass *Brachypodium distachyon*. Ann Bot. 2012;109(2):385–405. https://doi.org/10.1093/aob/mcr294.
9. Chen X, Kindle KL, Sternai DB. Site of translation Initiation in Chlamydomonas Chloroplasts. Society. 1995;7:1295–305.
10. Clement M, Posada D, Crandall KA. TCS: a computer program to estimate gene genealogies. Mol Ecol. 2000;9:1657–9.
11. Crosby K, Smith DR. Does the mode of plastid inheritance influence plastid genome architecture? PLoS One. 2012;7(9):1–8. https://doi.org/10.1371/journal.pone.0046260.
12. Da Silva RS, Clementi CR, Balsanelli E, de Baura VA, de Souza EM, de Freitas Fraga HP, et al. The plastome sequence of Bactris gasipaes and evolutionary analysis in tribe Cocoseae (Arecaceae). PLoS One. 2021;16(8 August):1–15. https://doi.org/10.1371/journal.pone.0256373.
13. Daniell H, Lin C-S, Yu M, Chang W-J. Chloroplast genomes: diversity, evolution, and applications in genetic engineering. Genome Biol. 2016;17:1–29.
14. Dierckxsens N, Mardulyn P, Smits G. NOVOPlasty : de novo assembly of organelle genomes from whole genome data. 2017;45(4). https://doi.org/10.1093/nar/gkw955.
15. Dierckxsens N, Mardulyn P, Smits G. Unraveling heteroplasmy patterns with NOVOPlasty. NAR Genom Bioinform. 2020;2(1):1–10. https://doi.org/10.1093/nargab/lqz011.
16. Doyle JJ, Doyle JL. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem Bull. 1987;19(1):11–5.
17. Gordon SP, Contreras-Moreira B, Levy JJ, Djamei A, Czedik-Eysenberg A, Tartaglio VS, et al. Gradual polyploid genome evolution revealed by pan-genomic analysis of *Brachypodium hybridum* and its diploid progenitors. Nat Commun. 2020;11(1):1–16. https://doi.org/10.1038/s41467-020-17302-5.
18. Gordon SP, Contreras-Moreira B, Woods DP, Des Marais DL, Burgess D, Shu S, et al. Extensive gene content variation in theBrachypodium distachyon pan-genome correlates with population structure. Nat Commun. 2017;8:2184. https://doi.org/10.1038/s41467-017-02292-8.
19. Greiner S, Bock R. Tuning a ménage à trois: co-evolution and co-adaptation of nuclear and organellar genomes in plants. Bioessays. 2013;35(4):354–65. https://doi.org/10.1002/bies.201200137.
20. Greiner S, Sobanski J, Bock R. Why are most organelle genomes transmitted maternally? Bioessays. 2015;37(1):80–94. https://doi.org/10.1002/bies.201400110.
21. Greiner S, Wang X, Herrmann RG, Rauwolf U, Mayer K, Haberer G, et al. The complete nucleotide sequences of the 5 genetically distinct plastid genomes of Oenothera, subsection Oenothera: II. a microevolutionary view using bioinformatics and formal genetic data. Mol Biol Evol. 2008;25(9):2019–30. https://doi.org/10.1093/molbev/msn149.
22. Greiner S, Wang X, Rauwolf U, Silber MV, Mayer K, Meurer J, et al. The complete nucleotide sequences of the five genetically distinct plastid genomes of Oenothera, subsection Oenothera: I. sequence evaluation and plastome

evolution. Nucleic Acids Res. 2008;36(7):2366–78. https://doi.org/10.1093/nar/gkn081.
23. Hasegawa M, Zhong B, Zhong Y. Adaptive evolution of chloroplast genomes in ancestral grasses. Plant Signal Behav. 2009;4(7):623–4. https://doi.org/10.4161/psb.4.7.8914.
24. Iii GPWG, Arthan W, Baker WJ, Barrett MD, Barrett RL, Bennetzen JL, et al. A nuclear phylogenomic tree of grasses (Poaceae) recovers current classification despite gene tree incongruence. New Phytol. 2025;245(2):818–34.
25. Jiang D, Cai X, Gong M, Xia M, Xing H, Dong S, et al. Complete chloroplast genomes provide insights into evolution and phylogeny of Zingiber (Zingiberaceae). BMC Genomics. 2023;24(1):1–18. https://doi.org/10.1186/s12864-023-09115-9.
26. Jombart T. Adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics. 2008;24(11):1403–5.
27. Jombart T, Collins C. A tutorial for discriminant analysis of principal components (DAPC) using adegenet 2.0. 0. London: Imperial College London, MRC Centre for Outbreak Analysis and Modelling; 2015.
28. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013;30:772–80. https://doi.org/10.1093/molbev/mst010.
29. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, et al. Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics. 2012;28(12):1647–9. https://doi.org/10.1093/bioinformatics/bts199.
30. Kofler R, Schlötterer C, Lelley T. Sciroko: a new tool for whole genome microsatellite search and investigation. Bioinformatics. 2007;23(13):1683–5. https://doi.org/10.1093/bioinformatics/btm157.
31. Kosakovsky Pond SL, Frost SDW. Not so different after all: a comparison of methods for detecting amino acid sites under selection. Mol Biol Evol. 2005;22(5):1208–22. https://doi.org/10.1093/molbev/msi105.
32. Kosakovsky Pond SL, Frost SDW, Muse SV. HyPhy: hypothesis testing using phylogenies. Bioinformatics. 2005;21(5):676–9. https://doi.org/10.1093/bioinformatics/bti079.
33. Kosakovsky Pond SL, Poon AFY, Velazquez R, Weaver S, Hepler NL, Murrell B, et al. HyPhy 2.5—a customizable platform for evolutionary hypothesis testing using phylogenies. Mol Biol Evol. 2020;37(1):295–9.
34. Krawczyk K, Myszczyński K, Nobis M, Sawicki J. Insights into adaptive evolution of plastomes in *Stipa* L. (Poaceae). BMC Plant Biol. 2022;22(1):1–16. https://doi.org/10.1186/s12870-022-03923-z.
35. Krech K, Ruf S, Masduki FF, Thiele W, Bednarczyk D, Albus CA, et al. The plastid genome-encoded Ycf4 protein functions as a nonessential assembly factor for photosystem I in higher plants. Plant Physiol. 2012;159(2):579–91. https://doi.org/10.1104/pp.112.196642.
36. Leigh JW, Bryant D. POPART: full-feature software for haplotype network construction. Methods Ecol Evol. 2015;6(9):1110–6.
37. Letunic I, Bork P. Interactive tree of life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Res. 2021;49(W1):W293–6. https://doi.org/10.1093/NAR/GKAB301.
38. López-Alvarez D, Manzaneda AJ, Rey PJ, Giraldo P, Benavente E, Allainguillaume J, et al. Environmental niche variation and evolutionary diversification of the *Brachypodium distachyon* grass complex species in their native circum-Mediterranean range. Am J Bot. 2015;102(7):1073–88. https://doi.org/10.3732/ajb.1500128.
39. Lu R, Sen Hu K, Zhang FJ, Sun XQ, Chen M, Zhang YM. Pan-Plastome of Greater Yam (Dioscorea alata) in China: intraspecific genetic variation, comparative genomics, and phylogenetic analyses. Intern J Mol Sci. 2023;24(4). https://doi.org/10.3390/ijms24043341.
40. Magdy M, Ou L, Yu H, Chen R, Zhou Y, Hassan H, Feng B, Taitano N, van der Knaap E, Zou X, Li F, Ouyang B. Pan-plastome approach empowers the assessment of genetic variation in cultivated Capsicum species. Horticult Res. 2019;6(1). https://doi.org/10.1038/s41438-019-0191-x.
41. Mason-Gamer RJ, Holsinger KE, Jansen RK. Chloroplast DNA haplotype variation within and among populations of *Coreopsis grandiflora* (Asteraceae). Mol Biol Evol. 1995;12(3):371–81. https://doi.org/10.1093/oxfordjournals.molbev.a040228.
42. Minadakis N, Williams H, Horvath R, Caković D, Stritt C, Thieme M, et al. The demographic history of the wild crop relative *Brachypodium distachyon* is shaped by distinct past and present ecological niches. Peer Community J. 2023;3:e84. https://doi.org/10.24072/pcjournal.319.
43. Mogensen HL. Invited special paper: the hows and whys of cytoplasmic inheritance in seed plants. Am J Bot. 1996;83(3):383–404.
44. Moreira CD, Gmitter FG, Grosser JW, Huang S, Ortega VM, Chase CD. Inheritance of organelle DNA sequences in a Citrus-Poncirus intergeneric cross. J Hered. 2002;93(3):174–8. https://doi.org/10.1093/jhered/93.3.174.
45. Morton BR. The role of context-dependent mutations in generating compositional and codon usage bias in grass chloroplast DNA. J Mol Evol. 2003;56(5):616–29. https://doi.org/10.1007/s00239-002-2430-1.
46. Mu W, Li K, Yang J, Breiman A, Yang Y, Wu Y, Shuang W, Zhu M, Liu J, Nevo E, Catalan P. Scattered differentiation of unlinked loci across the genome underlines ecological divergence of the selfing grass Brachypodium stacei. Proc Natl Acad Sci. 2023;120(45). https://doi.org/10.1101/2023.06.06.543844.
47. Mu W, Li K, Yang Y, Breiman A, Yang J, Wu Y, et al. Subgenomic stability of progenitor genomes during repeated allotetraploid origins of the same grass *Brachypodium hybridum*. Mol Biol Evol. 2023;40(12):1–20. https://doi.org/10.1093/molbev/msad259.
48. Murrell B, Moola S, Mabona A, Weighill T, Sheward D, Kosakovsky Pond SL, et al. FUBAR: a fast, unconstrained bayesian AppRoximation for inferring selection. Mol Biol Evol. 2013;30(5):1196–205. https://doi.org/10.1093/molbev/mst030.
49. Murrell B, Wertheim JO, Moola S, Weighill T, Scheffler K, Kosakovsky Pond SL. Detecting individual sites subject to episodic diversifying selection. PLoS Genet 2012;8(7). https://doi.org/10.1371/journal.pgen.1002764.
50. Naver H, Boudreau E, Rochaix JD. Functional studies of Ycf3: its role in assembly of photosystem I and interactions with some of its subunits. Plant Cell. 2001;13(12):2731–45. https://doi.org/10.1105/tpc.13.12.2731.
51. Nguyen L-T, Schmidt HA, Von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol. 2015;32(1):268–74.
52. Oldenburg DJ, Bendich AJ. Most chloroplast DNA of maize seedlings in linear molecules with defined ends and branched forms. J Mol Biol. 2004;335(4):953–70. https://doi.org/10.1016/j.jmb.2003.11.020.
53. Orton LM, Barberá P, Nissenbaum MP, Peterson PM, Quintanar A, Soreng RJ, Duvall MR. A 313 plastome phylogenomic analysis of Pooideae: Exploring relationships among the largest subfamily of grasses. Mol Phylogenet Evol. 2021;159(August 2020). https://doi.org/10.1016/j.ympev.2021.107110.
54. Pfeifer B, Wittelsbürger U, Ramos-Onsins SE, Lercher MJ. PopGenome: an efficient Swiss army knife for population genomic analyses in R. Mol Biol Evol. 2014;31(7):1929–36.
55. Piot A, Hackel J, Christin PA, Besnard G. One-third of the plastid genes evolved under positive selection in PACMAD grasses. Planta. 2018;247(1):255–66. https://doi.org/10.1007/s00425-017-2781-x.
56. 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. https://doi.org/10.1016/S0169-5347(00)02097-8.
57. Rahmawati A, Volkaert HA, Dinarti D, Maskromo I, Hatta ANNL, Sudarsono S. Complete chloroplast genome sequences of coconut cv. Kopyor green dwarf and comparative genome analysis to oil palm, date palm, sago palm, and miniature sugar palm BT - oil crop genomics (H. Tombuloglu, T. Unver, G. Tombuloglu, & K. R. Hakeem (eds.); pp. 189–216). Springer International Publishing; 2021. https://doi.org/10.1007/978-3-030-70420-9_10.
58. Ramsey AJ, McCauley DE, Mandel JR. Heteroplasmy and patterns of cytonuclear linkage disequilibrium in wild carrot. Integr Comp Biol. 2019;59(4):1005–15. https://doi.org/10.1093/icb/icz102.
59. Reboud X, Zeyl C. Organelle inheritance in plants. Heredity. 1994;72(2):132–40. https://doi.org/10.1038/hdy.1994.19.
60. Rozas J, Ferrer-Mata A, Sanchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE, et al. DnaSP 6: DNA sequence polymorphism analysis of large data sets. Mol Biol Evol. 2017;34(12):3299–302. https://doi.org/10.1093/molbev/msx248.
61. Saarela JM, Burke SV, Wysocki WP, Barrett MD, Clark LG, Craine JM, Peterson PM, Soreng RJ, Vorontsova MS, Duvall MR. A 250 plastome phylogeny of the grass family (Poaceae): Topological support under different data partitions. PeerJ. 2018;(2):1–71. https://doi.org/10.7717/peerj.4299.
62. Sadhu L, Kumar K, Kumar S, Dass A, Pathak R, Bhardwaj A, et al. Chloroplasts evolved an additional layer of translational regulation based on non-AUG start codons for proteins with different turnover rates. Sci Rep. 2023;13(1):1–15. https://doi.org/10.1038/s41598-022-27347-9.
63. Sancho R, Cantalapiedra CP, López-Alvarez D, Gordon SP, Vogel JP, Catalán P, et al. Comparative plastome genomics and phylogenomics of *Brachypodium*: flowering time signatures, introgression and recombination in recently diverged ecotypes. New Phytol. 2018;218(4):1631–44. https://doi.org/10.1111/nph.14926.

64. Sancho R, Inda LA, Díaz-Pérez A, Des Marais DL, Gordon SP, Vogel JP, et al. Tracking the ancestry of known and 'ghost' homeologous subgenomes in model grass *Brachypodium* polyploids. Plant J. 2022;109(6):1535–58. https://doi.org/10.1111/tpj.15650.
65. Scarlett VT, Lovell JT, Shao M, Phillips J, Shu S, Goodstein DM, Jenkins J, Grimwood J, Barry K, Chalhoub B, Hasterok R, Catala P, Vogel JP, Lusinska J, Schmutz J. Multiple origins , one evolutionary trajectory : gradual evolution characterizes distinct lineages of allotetraploid Brachypodium. 2022.
66. Schneider A. Organelle inheritance: understanding the basis of plastid transmission for transgenic engineering. J Mitochondria Plast Endosymbiosis. 2023;1(1). https://doi.org/10.1080/28347056.2023.2261790.
67. Scholthof KBG, Irigoyen S, Catalan P, Mandadi KK. *Brachypodium*: a monocot grass model genus for plant biology. Plant Cell. 2018;30(8):1673–94. https://doi.org/10.1105/tpc.18.00083.
68. Sheppard AE, Ayliffe MA, Blatch L, Day A, Delaney SK, Khairul-Fahmy N, et al. Transfer of plastid DNA to the nucleus is elevated during male gametogenesis in tobacco. Plant Physiol. 2008;148(1):328–36. https://doi.org/10.1104/pp.108.119107.
69. Shiposha V, Catalán P, Olonova M, Marques I. Genetic structure and diversity of the selfing model grass Brachypodium stacei (Poaceae) in Western Mediterranean: Out of the Iberian Peninsula and into the islands. PeerJ. 2016(9). https://doi.org/10.7717/peerj.2407.
70. Shiposha V, Marques I, López-Alvarez D, Manzaneda AJ, Hernandez P, Olonova M, et al. Multiple founder events explain the genetic diversity and structure of the model allopolyploid grass Brachypodium hybridum in the Iberian Peninsula hotspot. Ann Bot. 2020;125(4):625–38. https://doi.org/10.1093/aob/mcz169.
71. Tamura K, Stecher G, Kumar S. MEGA11: molecular evolutionary genetics analysis version 11. Mol Biol Evol. 2021;38(7):3022–7. https://doi.org/10.1093/MOLBEV/MSAB120.
72. Tillich M, Lehwark P, Pellizzer T, Ulbricht-jones ES, Fischer A, Bock R, et al. GeSeq – versatile and accurate annotation of organelle genomes. Nucleic Acids Res. 2017;45(May):6–11. https://doi.org/10.1093/nar/gkx391.
73. Trunova D, Borowska-Zuchowska N, Mykhailyk S, Xia K, Zhu Y, Sancho R, Rojek-Jelonek M, Garcia S, Wang K, Catalan P, Kovarik A, Hasterok R, Kolano B. Does time matter? Intraspecific diversity of ribosomal RNA genes in lineages of the allopolyploid model grass Brachypodium hybridum with different evolutionary ages. BMC Plant Biol. 2024;24(1). https://doi.org/10.1186/s12870-024-05658-5.
74. Vogel JP, Garvin DF, Mockler TC, Schmutz J, Rokhsar D, Bevan MW, et al. Genome sequencing and analysis of the model grass *Brachypodium distachyon*. Nature. 2010;463(7282):763–8. https://doi.org/10.1038/nature08747.
75. Wang J, Liao X, Gu C, Xiang K, Wang J, Li S, Tembrock LR, Wu Z, He W. The Asian lotus (Nelumbo nucifera) pan-plastome: diversity and divergence in a living fossil grown for seed, rhizome, and aesthetics. Ornamental Plant Res. 2022;2. https://doi.org/10.48130/OPR-2022-0002.
76. Wolfe KH, Li W-H, Sharp PM. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. Proc Natl Acad Sci USA. 1987;84(24):9054–8.
77. Wu H, Li DZ, Ma PF. Unprecedented variation pattern of plastid genomes and the potential role in adaptive evolution in *Poales*. BMC Biol. 2024;22(1):97. https://doi.org/10.1186/s12915-024-01890-5.
78. Wu H, Yang JB, Liu JX, Li DZ, Ma PF. Organelle phylogenomics and extensive conflicting phylogenetic signals in the monocot order poales. Front Plant Sci 2022;12(January). https://doi.org/10.3389/fpls.2021.824672.
79. Wysocki WP, Clark LG, Attigala L, Ruiz-Sanchez E, Duvall MR. Evolution of the bamboos (Bambusoideae; Poaceae): a full plastome phylogenomic analysis genome evolution and evolutionary systems biology. BMC Evol Biol. 2015;15(1):1–12. https://doi.org/10.1186/s12862-015-0321-5.
80. Yang T, Sahu SK, Yang L, Liu Y, Mu W, Liu X, et al. Comparative analyses of 3,654 plastid genomes unravel insights into evolutionary dynamics and phylogenetic discordance of green plants. Front Plant Sci. 2022;13:1–13. https://doi.org/10.3389/fpls.2022.808156.
81. Zhang Z, Zhang DS, Zou L, Yao CY. Comparison of chloroplast genomes and phylogenomics in the *Ficus sarmentosa* complex (Moraceae). PLoS One. 2022;17(12 December):1–22. https://doi.org/10.1371/journal.pone.0279849.

Publisher's Note