BMC Plant Biology

RESEARCH

Open Access

The presence and impact of G-quadruplexes in plant chloroplast DNA

Michaela Dobrovolná[1,2], Georgie Middleton[3,4], Stefan Bidula[3,4], Vratislav Peška[1], Lionel Guittat[5], Jean-Louis Mergny[5*] and Václav Brázda[1,2*]

Abstract

Background Chloroplasts are essential organelles in the regulation of oxygen metabolism in algae and plants, and therefore play a central role in the physiology of oxygen-dependent life. Chloroplasts typically contain a circular DNA molecule with replication and inheritance largely independent of the nuclear genome. Putative G-quadruplex–forming sequences (PQS) have emerged as important non-canonical DNA features in various genomes; however, their distribution and potential roles in chloroplast DNA (cpDNA) remain poorly understood. In this study, we used the G4Hunter algorithm to perform a comprehensive, large-scale analysis of PQS across all publicly available chloroplast genomes in the NCBI database ($n = 7,187$).

Results Our analysis provides a systematic characterization of cpDNA with respect to PQS localisation, frequency, conservation, and variation across major evolutionary lineages. The average frequency of PQS in cpDNA genomes was 0.873 PQS per kilobase pair (kbp), ranging from 0.018 PQS/kbp in Sonderella linearis to 5.011 PQS/kbp in Paradoxia multiseta. Marked lineage-specific differences were observed, with the highest PQS frequencies in lycophytes and the lowest in red algae. PQS were unevenly distributed along cpDNA sequences, with enrichment depending on motif length and genomic context. Specifically, PQS were preferentially located within transposable elements, replication origins, tRNA genes, 3′ untranslated regions (3′UTRs), and repeat regions, rather than being randomly dispersed.

Conclusions Our findings reveal non-random patterns of PQS distribution across chloroplast genomes and identify genomic regions with consistent enrichment of these motifs. While these observations are consistent with the possibility that PQS could contribute to structural or regulatory processes in cpDNA, their functional relevance remains to be experimentally validated. This work provides a comprehensive resource and establishes a foundation for future studies aimed at elucidating the biological roles of G-quadruplex structures in chloroplast genomes.

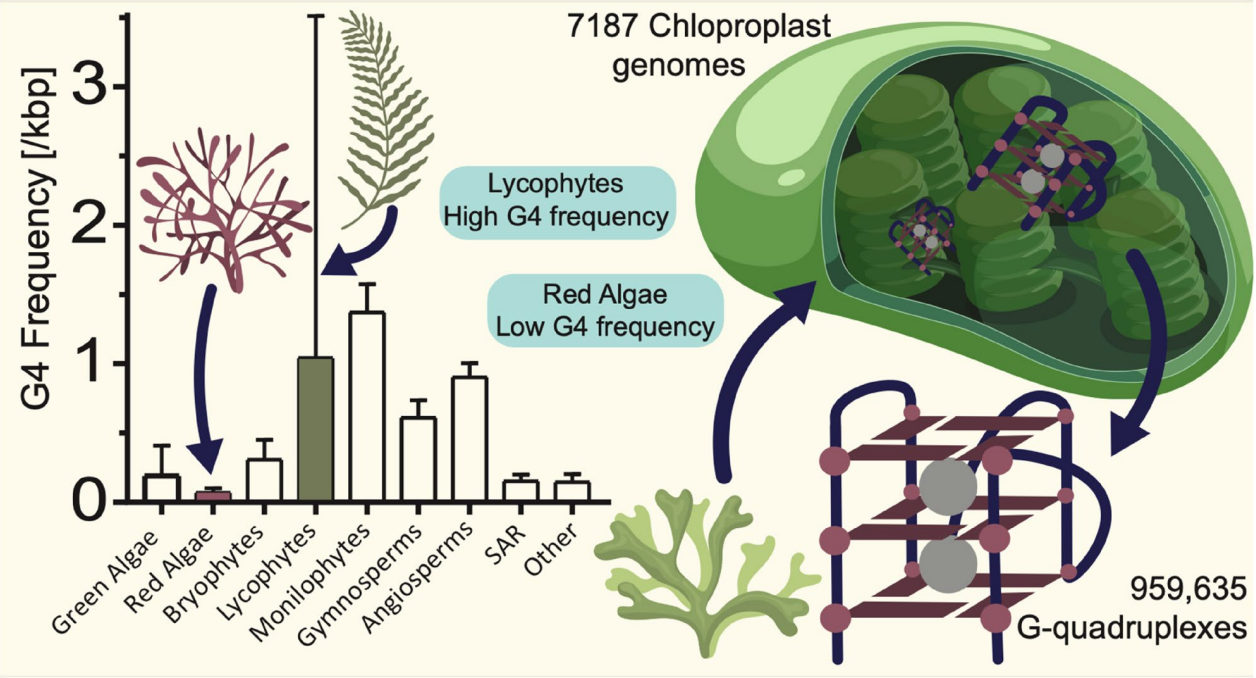
Keywords G-quadruplex, G4 frequency, Chloroplast DNA, Sequence prediction

*Correspondence:
Jean-Louis Mergny
jean-louis.mergny@polytechnique.edu
Václav Brázda
vaclav@ibp.cz

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

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

Graphical Abstract



Introduction

Chloroplasts are essential intracellular organelles in plants. Beyond their well-known role in photosynthesis, chloroplasts contribute to the biosynthesis of amino acids, nucleotides, fatty acids, phytohormones, vitamins, porphyrins, isoprenoids, and starch [1, 2]. Interestingly, some of these biosynthetic pathways, such as fatty acid synthesis, also occur in the cytosol of the cell. This means that the same molecules can be synthesized in parallel at different locations within the cell, ensuring greater flexibility and efficiency in producing these important substances [3]. Most of the proteins present in chloroplasts are encoded by nuclear DNA, which reflects the evolutionary shift from their cyanobacterial ancestors [4]. Chloroplasts originated from a symbiotic relationship between a primitive eukaryotic cell and a photosynthetic cyanobacterium, leading to the integration of the cyanobacterial genome into the host cell and the eventual formation of the modern chloroplast [1, 5]. Over time, chloroplast DNA (cpDNA) has undergone significant reduction and rearrangement, resulting in the highly specialized genomes seen in contemporary plant chloroplasts [6]. The nuclear-encoded proteins are synthesized as precursors on cytosolic ribosomes and then imported into chloroplasts, where they are incorporated into the metabolic pathways [1]. This intricate process illustrates the ongoing collaboration between the nuclear and plastid genomes to maintain chloroplast function. These versatile functions underscore the importance of chloroplasts in maintaining plant health and development [7].

cpDNA typically ranges in size from 100 to 247 kb in most plants, although both smaller and larger genomes can be found. The chloroplast genome contains 110–130 genes, including those essential for plant metabolism, such as those involved in photosynthesis, transcription, and ribosomal RNA production [8–10]. Of these, 80–90 are protein-coding genes, 30–31 are tRNA genes, and 4 are rRNA genes [10]. Plant cells typically harbour multiple chloroplasts, and multiple copies of cpDNA in each chloroplast, with the number of copies varying based on cell type and its differentiation state [11]. In higher plants, the cpDNA is generally organized into two long inverted repeats (IRs) with reverse polarity, which separate a large and a small single-copy region [8, 9, 12], a feature that plays a significant role in evolutionary processes [13]. On the other hand, short IRs have been shown to be abundant in regulatory regions of cpDNA [14]. Additionally, the presence of IRs is closely linked to the high occurrence of G-quadruplex structures (G4s) within the genome [15, 16]. G4s are non-canonical DNA secondary structures consisting of four guanine bases connected by Hoogsteen hydrogen bonds to form stacked guanine tetrads, stabilized by the presence of monovalent cations [17]. These structures predominantly form in guanine-rich regions of the genome [18] and are associated with diverse regulatory functions, including modulation of

gene expression [19], maintenance of genome stability [20], and regulation of DNA replication and transcription [21].

G4 structures have been identified across a diverse range of organisms [22], including viruses [23–25], bacteria [26], archaea [27], fungi [28, 29] (including yeast [30]), and mammals [31], such as *Homo sapiens* [19, 32]. They have also been detected in mitochondrial DNA (mtDNA) across various taxonomic groups, where they are notably overrepresented in regions such as the 3′ untranslated regions (3′UTRs), displacement loops (D-loops), replication origins, and stem-loop structures, suggesting regulatory roles for quadruplexes in mtDNA processes [33]. Yet, even if one can easily imagine that G4s may also be present in cpDNA and participate to chloroplast biology, this hypothesis remains to be fully explored.

This study aims to explore the presence and impact of G4s within cpDNA. Understanding the distribution and functional significance of G4s in cpDNA may reveal new insights into the regulation of chloroplast genes and the broader implications for plant physiology and adaptation. Through this research, we aim to contribute to the growing body of knowledge on the complexity of chloroplast genomes and the potential regulatory mechanisms involving G4 structures.

Materials and methods

Process of analysis

The cpDNA sequences analysed in this study were obtained from the National Center for Biotechnology Information (NCBI). All IDs of cpDNA sequences were downloaded (11/7/2022) from the organelle genome database of the NCBI. After deleting duplicate organisms, 7,187 sequences were analyzed using the G4Hunter software (http://bioinformatics.ibp.cz) [34, 35] to identify putative quadruplex sequences (PQS). The default G4Hunter parameters were used (window size of 25 nucleotides and a threshold of 1.2), as these settings have been previously validated for identifying experimentally confirmed quadruplex structures [34]. Although the number of publicly available plastome sequences has increased since the download date, the most representative organisms across major plant lineages were already included at the time of retrieval. Much of the subsequent growth in available sequences reflects redundant entries from already sampled species, and the main conclusions of this study are therefore expected to remain robust with respect to dataset completeness. In the NCBI organelle genome database, cpDNAs are arranged into five groups (Plants, Protists, Other Protists, Green Algae and Other). We additionally divided downloaded sequences into 8 evolutionarily related groups including Angiosperms, Gymnosperms, Monilophytes, Lycophytes, Bryophytes, Green and Red Algae, and category labelled "Other". The "Other" group encompasses the highly diverse eukaryotic clade SAR (Stramenopiles, Alveolates, and Rhizaria), as well as organisms belonging to Discoba, Cryptophyceae, and Haptista. For more detailed statistics, including the subdivisions within the "Other" category, see Supplementary Material 01. This division allowed us to observe trends in PQS frequency across the groups. To assess conservation, a BLAST search of PQS found in the model organism *Arabidopsis thaliana* was conducted. Limitations were set to 5000 sequences, results only included plant species, and other *Arabidopsis* species were excluded from the analysis.

Analysis of PQSs around annotated NCBI features

To understand the potential roles of G4s in cpDNA, the locations of PQS within the chloroplast genome were identified. Genomic annotations were obtained from the NCBI database and overlaid with G4Hunter results to determine whether PQSs were located within or 100 bp upstream/downstream of these annotated regions. The repeat regions considered in this analysis correspond exactly to those annotated in the original NCBI database entries for each genome, encompassing features such as inverted repeats, tandem repeats, and other repeat-associated elements as deposited by the original submitters. No additional repeat classification or filtering was applied beyond what was provided in the source annotations. It should be noted that annotation quality may vary across publicly available chloroplast genomes and any associations involving annotated genomic features such as repeat regions, tRNA, or replication origins should be interpreted with this caveat in mind, as heterogeneous annotation represents an inherent limitation of database-scale comparative studies. To assess the PQS distribution patterns, results were visualized using GraphPad Prism to illustrate the relationship between PQSs and annotated genomic features.

Statistical analysis

Non-parametric Kruskal-Wallis tests with Dunn's multiple comparisons were employed to assess significance. Correlation was evaluated using the two-tailed Pearson's correlation coefficient. Data normality was assessed using the Shapiro-Wilk test. All figures and analyses were generated using GraphPad Prism (v 8.0.1) [36].

Comparison of PQSs in cpDNA and control

For each group, the mean GC content and sequence length were calculated. Ten control sequences per group were then generated using the online tool DNAGenIQ - Random DNA Sequence Generator (ProteinIQ, accessed 18/5/2026, https://proteiniq.io/app/random-dna), matching the group-specific average length and closely approximating the average GC content. In total, this

yielded 100 control sequences across 10 groups. The control sequences were analysed using G4Hunter under the same parameters applied to the cpDNA dataset (window size of 25 nucleotides and a threshold of 1.2). The resulting predictions were subsequently compared with those obtained for cpDNA and are provided in Supplementary Material 02 (Supplementary Table 01, Supplementary Fig. 01).

Results

Using G4Hunter, the number of PQS were quantified in 7187 cpDNAs across 8 divisions. Due to the high variability in sequence size between organisms, to normalise the data, the total number of PQS in addition to the frequency of PQS per kbp and number of PQS relative to the GC content were noted. The length of cpDNA sequences varies between 15.5 kbp for *Asarum minus* and 610 kbp for *Bulboplastis apyrenoidosa* with a median length of 155,328 bp. The highest GC content was found in *Trebouxiophyceae sp*. *MX-AZ01* (57.66%), the lowest in *Dixoniella grisea* (23.17%). The mean GC content was 37.05%. Using default values for G4Hunter (i.e., a window size of 25 nucleotides and threshold score of 1.2), we found nearly one million PQS among all cpDNAs, giving an average PQS frequency of 0.873 PQS per kbp. In comparison, the PQS frequency in human mitochondrial DNA is 6.6 PQS per kbp, highlighting that the PQS density in cpDNAs is considerably lower.

Across the groups, Monilophytes had the largest average number of PQS (293 PQS per sequence) and the highest average PQS frequency (2.136 PQS/kbp) followed by Lycophytes with 198 PQSs per sequence and 1.296 PQS/kbp. Within Lycophytes, the observed split between high-frequency and low-frequency clusters reflects underlying phylogenetic subgrouping. The high-frequency cluster (approximately 3 or more PQS/kbp) corresponds to *Selaginellales*, whereas *Isoetales* consistently show the lowest values within the group, with an average of 0.68 PQS/kbp. A comparable phylogenetic signal is evident in Gymnosperms, where the highest PQS frequencies are found in *Cycadales*, encompassing both *Zamiaceae* and *Cycadaceae*, followed by *Pinales*. The lowest frequencies within this group, falling below 0.5 PQS/kbp, are observed in *Cupressales*. These distribution patterns therefore appear to reflect genuine phylogenetic signal rather than technical artefact. The Red Algae scored lowest for average of PQS (15 PQSs per sequence) and average PQS frequency (0.080 PQS/kbp). Organisms within the Other group had on average 21 PQS per cpDNA (0.175 PQS/kbp). The Green Algae with 77 PQS per sequence had an average frequency of 0.448 PQS/kbp closely followed by Bryophytes with 55 PQS per sequence and 0.407 PQS/kbp. Among bryophytes, PQS frequency varied considerably across constituent classes, highlighting the limitation of treating this group as a single analytical unit. The highest PQS frequency within bryophytes was observed in *Haplomitriopsida* (*Marchantiophyta*), with an average of 2.284 PQS/kbp, though this value is based on a single sequence and should be interpreted with caution. Within *Marchantiophyta*, *Jungermanniopsida* showed a moderate frequency of 0.527 PQS/kbp, while *Marchantiopsida* had the lowest frequency within the phylum at 0.214 PQS/kbp. *Anthocerotophyta* classes showed uniformly low frequencies, with *Anthocerotopsida* and *Leiosporocerotopsida* averaging 0.346 and 0.289 PQS/kbp respectively. Notably, *Bryophyta* itself displayed considerable internal variation. *Takakiophytina* recorded the highest frequency within this phylum at 0.933 PQS/kbp, followed by *Sphagnophytina* at 0.706 PQS/kbp, while *Bryophytina* showed substantially lower values at 0.291 PQS/kbp, approaching the lowest frequencies observed across the entire bryophyte dataset. This internal heterogeneity within Bryophyta, spanning more than a threefold difference in PQS frequency between subgroups, underscores the importance of subclass-level resolution when interpreting PQS distribution patterns in this lineage.

Table 1 G4Hunter analysis results for cpDNA

Group	Seq	Median	Short	Long	PQS	Mean f	Min f	Max f	GC%
All	7187	155,328	33,539	610,063	959,635	0.873	0.018	5.011	37.05
Green Algae	174	142,563	48,188	521,168	13,351	0.448	0.038	5.011	32.81
Red Algae	102	178,965	126,100	610,063	1556	0.080	0.018	0.385	29.07
Bryophytes	73	123,907	108,007	161,162	3757	0.407	0.171	2.284	31.23
Lycophytes	25	144,680	110,411	150,349	7328	2.136	0.672	4.553	44.35
Monilophytes	134	151,615	131,150	172,311	26,528	1.296	0.329	2.133	41.29
Gymnosperms	182	127,313	85,318	166,341	13,794	0.588	0.333	1.094	36.96
Angiosperms	6351	155,586	15,553	242,575	890,301	0.913	0.170	2.001	37.37
SAR	117	125,283	33,539	201,816	2539	0.178	0.057	1.356	31.66
Other	26	94,733	33,539	143,171	407	0.146	0.075	0.277	30.22

Seq (number of sequences), median, short and long correspond to the median, min. and max. lengths in the dataset, PQS (number of PQS), Mean f (mean of PQS per kb), Min f (lowest frequency), Max f (highest frequency), GC% (GC content)

Total PQS counts, sequence lengths, percentage of GC and PQS frequencies characteristics for each group are summarized in Table 1. The distribution of PQS/kbp within individual species in each of the groups can be seen in Fig. 1.

Higher G4Hunter scores (G4HS) correspond to motifs capable of forming very stable quadruplexes. As expected, the number of PQS tended to decrease with the G4Hunter threshold, and candidate sequences may be grouped into different categories depending on G4Hunter score. Across all groups, the majority of PQS identified in cpDNA sequences were classified within the moderate G4HS range of 1.2–1.4, representing 66.39% of all PQS. Following this, sequences falling within the 1.4–1.6 range comprised 20.54% of all PQS. Despite being less frequent, we still found a significant number of sequences with high G4HS (1.6–1.8 interval: 78,491 PQS; 1.8–2.0 interval: 31,244 PQS; 2.0–more: 16,922 PQS accounting for 8.1%, 3.2% and 1.8% of all PQS). A lower frequency of high scoring PQS has already been observed for all genomes analyzed so far, including helminths [37], bacteria [38], archaea [39], and viruses [40]. The distribution of G4HS scores was compared between evolutionally related groups (Fig. 2). All groups showed a significant drop in PQS frequencies when higher thresholds were selected, but to different extents. Groups including Gymnosperms, Monilophytes, Lycophytes, Bryophytes and Green Algae were less enriched with PQS in the G4HS interval 1.8–2.0 than 2.0–more.A summary of all PQS found in intervals of G4HS and PQS frequencies per kbp for all cpDNAs is shown in Table 2, and for all groups separately in Supplementary Material 02 (Supplementary Table 1). It should be noted that phylogenetic non-independence among sequences, particularly within the deeply structured Angiosperm group, and the highly uneven group sizes may limit the interpretation of among-group comparisons. Results from smaller groups should therefore be treated with caution and are presented primarily for completeness. It is also worth noting that treating Bryophytes as a single category may obscure biologically meaningful differences between mosses, liverworts, and hornworts, which differ substantially in cpDNA architecture, GC content, and IR organisation. Readers interested in bryophyte subdivisions or any other taxonomic subgroup can filter sequences by class, order, and family using the classification provided in Supplementary Material 01.

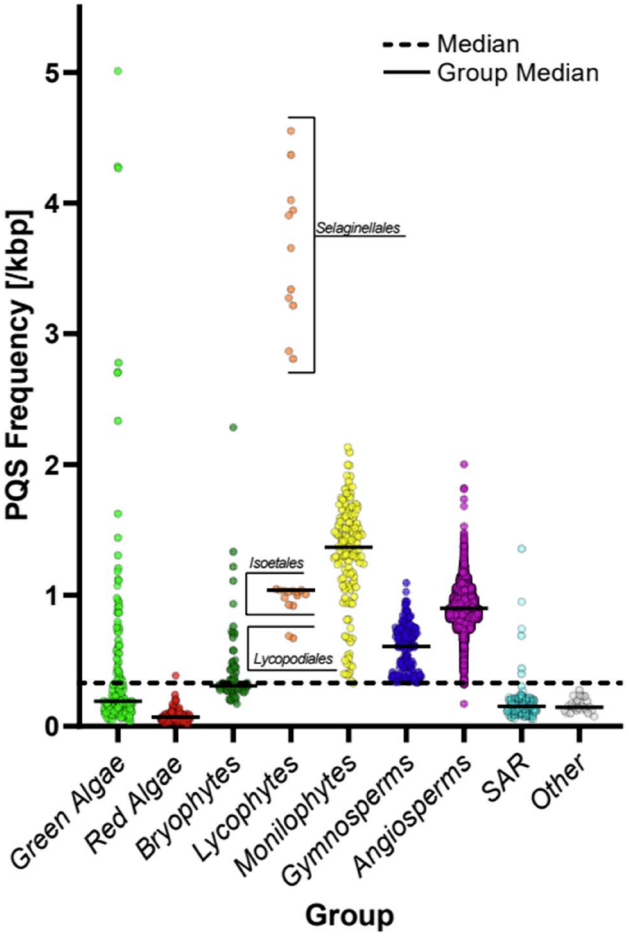


Fig. 1 The frequency of PQS per 1,000 nucleotides in the cpDNA of individual species within groups; Angiosperms ($n = 6,351$), Gymnosperms ($n = 182$), Monilophytes (= 134), Lycophytes (= 25), Bryophytes (= 73), Green Algae (= 174), Red Algae (= 102), SAR (= 117), Other (= 26); Horizontal solid lines represent median values; Frequencies across the groups were found to be significantly different to one another ($P < 0.0001$) as determined via a non-parametric Kruskal-Wallis test PQS were identified using G4Hunter with a threshold of 1.2

A steep decrease in highly stable PQS within the groups of Red Algae and Other was observed, likely due to the lower number of sequences in these groups (Fig. 2). A consistent decline in PQS was observed in the Lycophytes and Angiosperms and was particularly evident in the last two intervals. In contrast, Bryophytes and Gymnosperms showed relatively stable scores in the last two intervals. Notably, Monilophytes and Green Algae exhibited a remarkable number of highly stable PQS, which was not correlated with the length of the sequences nor with their GC content. PQS counts for each G4Hunter score threshold across the different groups are provided in Supplementary Material 02 (Supplementary Table 1).

To assess whether the observed PQS frequencies exceeded what would be expected from nucleotide composition alone, control sequences matching the group-specific average length and GC content were generated and analysed under identical G4Hunter parameters. Across all groups, cpDNA sequences showed substantially higher PQS frequencies compared to their respective GC-matched controls, demonstrating that the abundance of putative G4 motifs in chloroplast genomes cannot be attributed to nucleotide composition alone. In

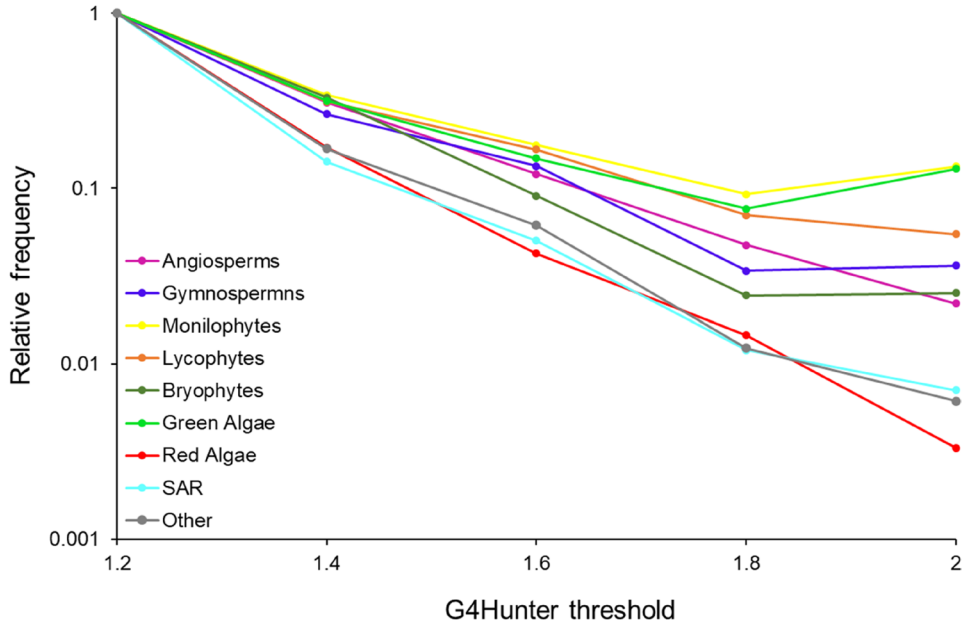


Fig. 2 Distribution of G4 prone motifs when analyzing genomes using a G4Hunter score ranging from 1.2 to 2.0 for the different groups (note the logarithmic axis on the Y-axis)

Table 2 Total number and frequencies of PQS found in cpDNA grouped by G4hunter score

G4Hunter threshold	Number of PQS	PQS frequency (PQS/kbp)
1.2–1.4	643,142	0.586
1.4–1.6	198,983	0.181
1.6–1.8	78,491	0.072
1.8–2.0	31,244	0.028
2.0 – more	16,922	0.015
Total	968,782	0.873

the 1.2–1.4 score interval, control sequences showed substantially lower PQS frequencies compared to cpDNA, with an overall decrease of 72% when considering the average across all sequences, and reductions as high as 84.5% and 83.6% observed for Green Algae and Bryophytes respectively. This difference became increasingly evident at higher score thresholds, where control sequences yielded negligible PQS counts while cpDNA retained a detectable number of high-scoring motifs. At score intervals above 1.8, control sequences of equivalent GC content yielded virtually no PQS predictions. The fact that cpDNA retained detectable high-scoring motifs at these thresholds suggests that their occurrence is not a random consequence of nucleotide composition but reflects genuine sequence-level organisation. A notable exception was observed in the Other category, where control sequences maintained comparable PQS frequencies to cpDNA across several score intervals and marginally exceeded them at some thresholds, likely reflecting the compositional and taxonomic heterogeneity of this group rather than genuine G4 enrichment in cpDNA. This pattern was otherwise consistent across the remaining nine groups examined, supporting the conclusion that PQS enrichment in cpDNA represents a non-random feature of plastid genome architecture. A full comparison of cpDNA and control PQS frequencies across all groups and score intervals can be found in Supplementary Material 02 (Supplementary Table 01, Supplementary Fig. 01).

Next, we assessed the frequency of PQS within these genomes relative to the GC content of the genomes. As expected, there was a positive correlation ($R = 0.6879$) between genome GC content and the frequency of PQS (Fig. 3). This correlation is relatively weak though, with a number of outliers. For example, we observed a greater frequency of PQS than expected in the genomes of several species. For instance, *Paradoxia multiseta* displayed the highest observed PQS frequency at 5.01 PQS/kbp, with a GC content of 50.6%, corresponding to 9.91 PQS per 1000 GC base pairs. Similarly elevated frequencies were detected in *Selaginella sanguinolenta* (4.55 PQS/kbp), *Coccomyxa sp. Obi* (4.28 PQS/kbp), *Oltmannsiellopsis viridis* (2.78 PQS/kbp), *Udotea flabellum* (2.70 PQS/kbp), and *Cephaleuros virescens* (2.33 PQS/kbp). These values indicate a substantial enrichment of PQS motifs in the chloroplast genomes of these organisms, independent of their overall GC content.

Conversely, several species exhibited notably low PQS frequencies relative to their GC content. Among the lowest was *Choricystis parasitica* (0.722 PQS/kbp) with a GC content of 45.5%, corresponding to only 1.59 PQS per 1000 GC base pairs. Similarly, *Mesotaenium*

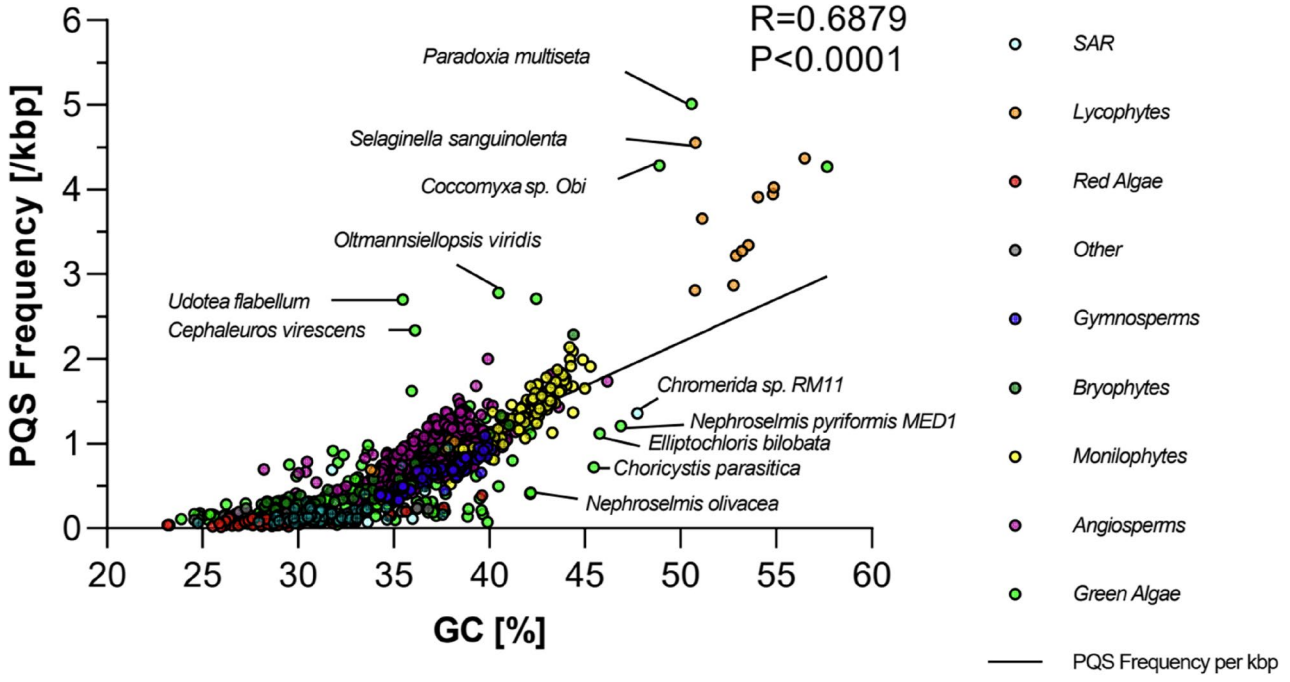


Fig. 3 The frequency of PQS relative to the GC content in cpDNA; The Pearson correlation coefficient was used to determine the association between PQS and GC content. The frequency of PQS is positively correlated with GC content; Correlation was calculated via two-tailed Pearson's correlation coefficient ($P < 0.0001$)

endlicherianum displayed a frequency of 0.422 PQS/kbp, equating to 1 PQS per 1000 GC base pairs. The chloroplast genome with the lowest overall PQS frequency was found in *Sonderella linearis* (0.018 PQS/kbp), with only three PQSs identified in the entire sequence. This was followed by *Acrosorium ciliolatum* and *Bryothamnion seaforthii*, both showing a frequency of 0.023 PQS/kbp and containing just four PQSs each across their cpDNA. *Sonderella linearis* (26.0% GC), *Acrosorium ciliolatum* (25.5% GC), and *Bryothamnion seaforthii* (29.5% GC) are all red algae with chloroplast genomes approximately 170 kbp in length. Despite having genome sizes comparable to other algae with higher PQS densities, their extremely low PQS counts suggest that genome length alone does not account for the reduced G4 potential. Rather, the combination of low GC content and possible lineage-specific constraints in red algae may contribute to this striking deficiency.

Interestingly, species such as *Dixoniella grisea* SAG 72.90, *Bulboplastis apyrenoidosa* NIES-2742, and *Aphanochaete elegans*, all members of the Chlorophyta (green algae), exhibit chloroplast GC contents under 24%. Despite this, their PQS frequencies were at least double those of the red algae mentioned above, with total PQS counts of 20, 23, and 16, respectively. This contrast highlights the influence of phylogenetic background on PQS distribution and suggests that factors other than GC content, including genome architecture, selective pressures, and other evolutionary processes, may contribute to the retention or elimination of G-rich motifs in cpDNA across different algal groups.

G4s are often found in higher frequencies within promoter regions and untranslated regions (UTRs) of higher eukaryotes where they have key regulatory roles within processes such as transcription [41]. Therefore, we were interested in identifying the genomic locations of PQS within the cpDNA genomes. Here, we determined whether PQS fell within an annotated feature (inside) or within 100 bp before or after the annotated feature (before and after, respectively). To ensure robustness of interpretation, we focused only on features that occurred at least 30 times across the dataset, with the exception of the replication origin, which was included due to its functional relevance despite having fewer occurrences. We identified that PQS were most likely to be found before, within, and after the mobile elements, replication origins, repeat regions, and tRNA regions (Fig. 4).

In addition to the genome analyses above, we explored the chloroplast genomes for any conserved PQSs. We focused specifically on the genome of the model organism *A. thaliana* in this instance and highlighted the sequences within its chloroplast genome capable of potentially forming three-tetrad G4s. These sequences are more likely to form stable G4s. We found 15 of these sequences falling inside or within 100 bp of specific genes in the chloroplast genome (Table 3). For each of these PQSs, we performed a BLAST search to identify if they were conserved amongst other plants. *Arabidopsis*

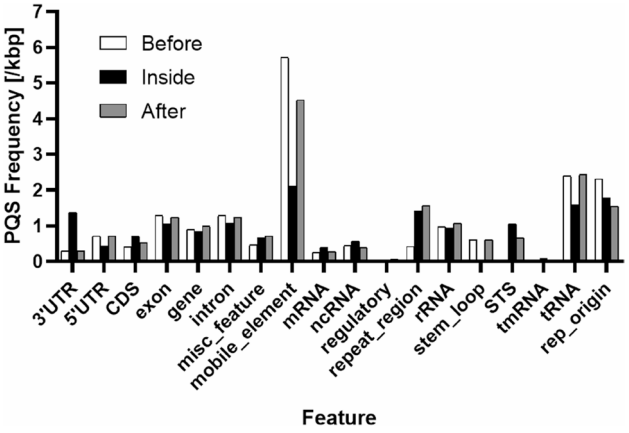


Fig. 4 The genomic locations of PQS found 100 bp before, within, and 100 bp after annotated features in cpDNA; Frequencies are normalized according to the PQS frequency; Annotations were taken from the NCBI database

species were not included in this search to obtain a wider perspective of their conservation amongst different genera and species, given that the BLAST search is limited to 5000 results. Three of the PQSs found within 100 base pairs upstream or downstream of *psaJ*, near *trnR*, and inside *atpB* in *A. thaliana* were found not to be conserved in any plant species (Table 3). In contrast, 6 of these G4-forming sequences (or slight variations of) were found to be highly conserved throughout hundreds of genera and thousands of species. PQSs with the greatest conservation amongst species were found within the *psbC*, *psaB*, *ycf2* and *trnR* genes (*trnR* contained 3 PQSs; see Table 3). The most conserved of these sequences, **CCCC**A**CCCC**GT**C**T**CC**A**C**TGGAT**C**TGTT**CCC**AG GAGTA**CCC** (a C-rich motif with a negative G4Hunter score of -1.35: its reverse complement would be prone to G4 formation), was located within 100 base pairs of *trnR*, which encodes the tRNA for arginine, and was found conserved in 977 different genera encompassing 3155 different species. The PQS in the genes *psbC* and *psaB*, which encode components of photosystems II and I, were also highly conserved in higher plants, but not in algae.

Complete sequence conservation alone does not distinguish selection of the G4 motif from broader sequence constraint. We therefore examined whether observed substitutions were compatible with retention of G4-forming potential. For *tnrR*, the motif was invariant across most sampled species, precluding substitution-bias analysis. However, a single mutation was found in the loop sequence of many species which likely has no impact on PQS formation. In *psaB*, substitutions were observed mainly in loop regions, such as those observed in *Canarium album* and *Vatica guangxiensis*. Indeed, mutations in the loop before the final C-tract were observed that may enhance formation of the structure (CCTCCC in *A. thaliana* but CCCCCC in other species such as *Hirtella racemosa*, *Acorus calamus*, *Capsicum annuum*, and *Solanum americanum*). The only genus that appeared to lose this PQS sequence was *Tragosporon* spp. where one of the cytosines in the C-tract was mutated to a thymine that would completely disrupt PQS formation. Similarly, for *psbC*, most mutations were in the flanking region prior to the first G-tract (TCTATTTT in *A. thaliana* versus TTTATTTT in *H. racemosa*, *A. calamus*, *C. annuum*, and *S. americanum*) or in the loop sequences preceding the final G-tract (CTCC in *A. thaliana* versus CTCCT in the other species). In some species, the G-tracts were mutated, but these mutations again appeared to conserve the PQS (GGGAGGGGGGG in *A. thaliana* versus GG GGGGGGGG in *H. racemosa* and *A. calamus*). These examples supported the idea that substitutions in these regions were compatible with G4 maintenance. However, G4 conservation seemed to be less consistent in *ycf2*, with mutations of the first guanine to an adenine in the second G-tract occurring frequently, potentially decreasing the stability of any G4 that may form. Taken together, there is strong evidence that evolution attempts to retain the PQS in most instances, but some caution must be taken when interpreting this conservation globally.

Discussion

G4 sequences in chloroplast genomes were first examined in 2012, when only a few plant species and a limited number of complete cpDNA sequences or tools for G4 analysis were available [42]. Since then, the rapid expansion of publicly accessible plastid genome data has enabled a much more comprehensive analysis across diverse plant lineages. Our findings reveal significant differences in the frequency of PQSs in cpDNA across various plant groups. The highest frequency of PQSs was observed in Monilophytes and Lycophytes followed by the Angiosperms, with the *Selaginella* genus showing particularly high frequencies. Interestingly, *Selaginella* species have notably small genomes with reduced gene numbers, which suggests that the remaining genomic regions, such as intergenic and noncoding areas, may harbour G4-prone sequences. Additionally, *Selaginella* plastids have shorter cpDNA sequences (average of 133,590 bp) compared to other plants (average 155,328 bp) and feature a significantly higher GC content of 53%, compared to the typical 37% found in other Lycophytes. *Selaginella* species represent a critical evolutionary stage in vascular plant development, showing a transition from simpler to more complex photosynthetic systems. The high frequency of PQSs in their cpDNA likely plays a role in regulating the expression of photosynthetic genes, helping these plants adapt to varying environmental conditions [43]. As an early-diverging lineage of vascular plants, *Selaginella* retains ancestral features, and the abundance of PQSs could reflect an ancient regulatory

Table 3 G4-forming sequences with their G4Hunter score and location within the Arabidopsis thaliana genome and their conservation amongst different plant genera and species as found by BLAST

Closest gene	Sequence position (bp)	Gene product	Sequence	G4Hunter score	Genera	Species
rpoB	23,382	Chloroplast DNA-dependent RNA polymerase β subunit	**CCC**T**CC**AGAG**CCC**AAA**CC**T**CC**ATTT**C**G**CC**TA**CCC**G**C**TG**CCCCCC**	-1.57	274	921
psbC	34,285	CP43 Photosystem II	**GGG**CTCC**GGG**A**GGGGGGG**	2.278	685	3001
psaB	38,646	Photosystem I	**CCC**TTATG**CCC**A**C**G**CCCC**AAT**C**GT**CC**T**CCC**	-1.533	838	2944
atpB	54,577	ATPase subunit β	**CCC**ATTTATT**C**AT**CCC**TTTAG**C**AA**CCCCCCC**	-1.407	0	0
accD	58,048	Carboxytransferase β subunit of the acetyl-CoA carboxylase complex	**GGG**TTAT**GG**ATTTTC**GG**TTTAT**GGGGGG**TA**G**TAT**GGG**	1.37	42	105
psaJ	67,164	CP43 Photosystem II	**GGGGGGGGGGGG**AAA**GG**TCATAAAAAC**G**A**GGG**	1.939	0	0
psbB	72,614	CP47 Photosystem II	**GGGGGG**ACTATAAC**G**AATCC**GGG**TCTTT**G**A**G**TTAC**G**AA**GGG**		212	524
psbB	73,334	CP47 Photosystem II	**GGGGGG**TTATTCC**G**A**G**C**GGG**TTCAAT**GG**ACAAT**GGG**		94	500
trnR	100,857	tRNA-Arg	**GGG**ATT**G**AC**G**T**G**A**GGGGG**TA**GGGG**TA**G**CTATATTTCT**GGG**	1.375	95	489
trnR	108,265	tRNA-Arg	**GGG**TACTCCT**GGG**AACA**G**ATCCA**G**T**GG**A**G**AC**GGGG**T**GGGG**	1.32	979	3126
trnR	108,390	tRNA-Arg	**GGG**AA**GG**ACCTTTCCCTCC**GGGGGGG**TA**GG**AAAATCAT**G**CTC**GGG**	1.37	0	0
trnR	130,244	tRNA-Arg	**GGG**AAA**GG**TCCTTCCCTTTT**GGG**CC**GG**TT**G**T**GGG**C**G**A**GG**A**GGG**	1.28	805	2936
trnR	130,383	tRNA-Arg	**CCCC**A**CCCC**GT**C**T**CC**A**C**TGGAT**C**TGTT**CCC**AGGAGTA**CCC**	-1.32	977	3155
arthct112	137,791	tRNA-Val	**CCC**AGAAATATAG**C**TA**CCCC**TA**CCCCC**T**C**AGT**C**AAT**CCC**	-1.385	97	490
ycf2	147,257	Probable ATPase	**GGGG**ATA**GGG**CTAA**G**C**GG**A**G**C**G**AAAA**GGG**TTTTCCAT**G**A**G**AC**GGG**		948	3098

strategy for cpDNA that has since diminished in other plant groups. Furthermore, many *Selaginella* species are known for their desiccation tolerance and ability to survive in extreme habitats [44]. G4 structures may regulate stress-responsive genes in the chloroplasts, contributing to their remarkable adaptability [45]. The high frequency of G4s could help maintain cpDNA integrity under oxidative stress, a common condition in chloroplasts due to elevated levels of reactive oxygen species (ROS). Moreover, *Selaginella* produces unique secondary metabolites, such as biflavonoids, which are synthesized with contributions from chloroplast-encoded genes and are researched for their anti-tumor activity [46]. G4s may regulate the expression of genes involved in these biosynthetic pathways, further supporting the plant's adaptability.

In contrast, Red Algae, which use chlorophyll-a and phycobiliproteins for photosynthesis [47], exhibited very low PQS frequencies. The streamlined genome of Red Algae, adapted to low-light aquatic environments [48], likely reduces the need for G4 structures to regulate pigment-related gene expression. Red Algae have evolved unique light-harvesting mechanisms that may not require the complexity provided by G4s. Their simpler chloroplast genomes, which lack extensive regulatory regions seen in other plants, may rely on alternative secondary structures, such as hairpins, to perform roles typically fulfilled by G4s.

The phylogenetic patterns observed in PQS frequency raise the question of what biological machinery might interact with these structures in the chloroplast context. Unlike the nuclear genome, the chloroplast transcriptional and replication apparatus is largely of prokaryotic origin, yet the organelle imports hundreds of nucleus-encoded proteins that have co-evolved with the plastid genome [49]. G4 structures in cpDNA could therefore interact with both plastid-encoded and nuclear-encoded components. In bacteria, G4 motifs have been shown to stall DNA polymerase progression and influence transcription elongation, and analogous effects may occur in plastids given the conserved prokaryotic-type RNA polymerase present in chloroplasts [26, 49]. The plastid-encoded RNA polymerase and the nucleus-encoded RNA polymerase differ substantially in their promoter recognition and transcriptional dynamics [50], and it is plausible that G4 structures located upstream of plastid genes differentially influence their activity, potentially contributing to lineage-specific differences in gene expression regulation. Supporting this, G4 structures have been shown to modulate transcription factor binding in nuclear contexts [51], and several nucleus-encoded, plastid-targeted proteins involved in transcription and RNA processing, including members of the pentatricopeptide repeat (PPR) protein family, are known to recognise specific nucleic acid secondary structures [52–54]. Whether PPR proteins or other imported RNA-binding factors interact with G4 motifs in cpDNA remains to be directly demonstrated, but the structural specificity of these proteins makes such interaction mechanistically plausible. Furthermore, the observed enrichment of PQSs in mobile genetic elements and repeat regions suggests a possible role in modulating recombination and genome stability, processes in which G4-resolving helicases such as RecQ-family proteins are implicated in both prokaryotes and eukaryotes [55–57]. The lineage-specific variation in PQS frequency documented here, particularly the elevated frequencies in *Selaginellales* [58] and *Cycadales* [59] relative to phylogenetically adjacent groups, is consistent with the hypothesis that PQS-mediated regulation has been differentially retained or elaborated across plant evolution, possibly in connection with lineage-specific adaptations in plastid gene expression, genome architecture, or stress response strategies [60, 61]. Feature analysis revealed an especially high frequency of PQSs in mobile genetic elements, which can move within or between genomes [62]. Mobile elements are crucial for gene evolution, including gene duplication, horizontal gene transfer, and mutations that can alter protein functions [62–64]. These genetic rearrangements enhance genomic diversity and can provide a selective advantage by facilitating the acquisition of beneficial traits, such as antibiotic resistance [65] or stress-resilience genes [66]. The high PQS density in mobile elements may help regulate these elements by modulating gene expression related to their mobility and interactions with host organisms, further supporting their evolutionary importance. It should be noted, however, that only 42 entries spanning 55,332 bp were annotated as mobile elements across all analysed cpDNA sequences, using exclusively the exact NCBI term "mobile_element" without aggregation of related terms. Given the heterogeneity of mobile element annotation across database entries, this limited count likely reflects annotation inconsistencies rather than a comprehensive survey, and the observed PQS enrichment in this feature class should therefore be interpreted with caution.

Previous studies have suggested that RNA G4s may act as molecular signatures involved in plant environmental adaptation [45]. Additionally, it has been shown that G4 structures can influence DNA replication in plants by inhibiting DNA polymerase movement in vitro [67]. In our earlier work on *Pisum sativum*, we extended the analysis beyond cpDNA to include the nuclear and mitochondrial genomes. That study revealed non-random PQS distribution in all three compartments, with experimental confirmation of G4 formation in organellar DNA and evidence linking PQSs to regulatory regions and transposable elements. These results demonstrated that

G4 structures are functionally relevant across the entire plant genome, not only in cpDNA [68].

Our current findings align with these earlier discoveries, as the presence of G4 structures in chloroplast genomes may represent an evolutionary strategy for gene regulation. G4s provide flexibility in gene expression, particularly in response to environmental stress. Furthermore, G4s may play a role in genome repair and maintenance, helping to stabilize cpDNA in the face of oxidative damage. Supporting this, recent work in *Nicotiana tabacum* revealed a higher G4 density in organellar genomes compared to nuclear DNA, enrichment of G4s in regulatory regions, and a potential role in enhancing drought tolerance through structures formed in transcription factor promoter regions [69]. Plant cpDNA has a stable, repeat-rich structure with large IRs, supporting compactness and gene retention, unlike the variable fungal mitochondrial genomes [15]. Although much has been learned about chloroplast genome evolution in green plants, several questions still need to be addressed. One of them concerns the loss of the rRNA operon-encoding IR, a structural feature conserved in many algae [70, 71] and almost all land plants [61, 72]. The large IR regions may also contribute disproportionately to PQS density in certain plant lineages. Because IRs are duplicated regions enriched in highly conserved genes, including ribosomal RNA operons [61, 73], PQSs located within these regions are effectively represented twice within the plastome, potentially increasing the observed number of predicted PQSs. Moreover, lineage-specific expansion or contraction of IR boundaries may influence both the distribution and frequency of PQSs among chloroplast genomes. Given the established role of IRs in plastome stability and homologous recombination, G4 structures within these regions may additionally contribute to the repair mechanisms, and genome maintenance through the formation of stable secondary DNA structure [74–76].

The relationship between GC content and PQS frequency is inherently bidirectional and therefore difficult to interpret unidirectionally. Higher GC content increases the probability of G-runs forming PQS-competent sequences, while PQS-rich regions simultaneously contribute to elevated GC content [34, 77]. Notably, GC content accounts for approximately 70% of the variance in PQS frequency across the dataset, meaning that roughly 30% of the variance remains unexplained by nucleotide composition alone. We argue that this residual fraction, while smaller, is biologically meaningful and likely reflects lineage-specific evolutionary pressures rather than compositional bias. This interpretation is consistent with the phylogenetically structured distribution patterns observed across the eight evolutionary groups examined in this study, where order- and family-level differences in PQS frequency persist even among lineages with comparable GC content.

In conclusion, our results emphasize the possible roles of G4 structures across different plant groups, suggesting that their presence in genomes is not merely a consequence of GC content. Rather, G4s may play a key role in the adaptive potential and evolutionary trajectory of plants. Further research is necessary to unravel the specific molecular mechanisms by which G4 structures influence gene expression and contribute to plant evolution, particularly in response to environmental stressors. It should be noted that the G4Hunter threshold of 1.2 applied here represents a permissive setting, and a proportion of predicted PQSs may not fold into stable structures under physiological conditions. In this context, however, the potassium-rich environment of the chloroplast stroma [78] may be favourable for G4 folding in cpDNA, lending biological plausibility to the motifs identified. As direct experimental validation lies beyond the scope of the present comparative bioinformatic study, the functional roles proposed here should be regarded as informed hypotheses rather than established mechanisms. PQS motifs conserved across broad taxonomic sampling and associated with functionally significant regions represent priority targets for future biophysical characterisation. Elucidating their specific roles in plastid gene expression and plant adaptation remains an important direction for experimental work.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12870-026-09148-8.

Supplementary Material 1. G4Hunter results.

Supplementary Material 2. Summary of G4Hunter results across all scores for the different groups and controls.

Acknowledgements

This work has been conducted in the sustainability period of the project SYMBIT No. CZ.02.1.01/0.0/0.0/15_003/0000477 as its follow-up activity.

Authors' contributions

Conceptualization, Stefan Bidula, Jean-Louis Mergny, Václav Brázda; Funding acquisition, Václav Brázda, Jean-Louis Mergny; Supervision, Stefan Bidula, Jean-Louis Mergny, Václav Brázda; Methodology, Michaela Dobrovolná, Lionel Guittat, Jean-Louis Mergny,, Václav Brázda; Investigation, Michaela Dobrovolná, Stefan Bidula; Formal analysis, Michaela Dobrovolná, Vratislav Peška; Visualization, Michaela Dobrovolná; Writing Original Draft, Michaela Dobrovolná; Writing Review & Editing, Michaela Dobrovolná, Georgie Middleton, Stefan Bidula, Jean-Louis Mergny, Václav Brázda. All authors read and approved the final manuscript.

Funding

This work was supported by the Ministry of Education, Youth and Sports of the Czech Republic under the INTER-EXCELLENCE II program, project number LUAKR25078, and by the European Regional Development Fund through the INTERREG Austria–Czech Republic programme, project number ATCZ00052. GM was supported by an Academy of Medical Sciences Springboard Award (SBF008\1046). SB was supported by the Biotechnology and Biological Sciences Research Council New Investigator Award (BB/Y005058/1).

Data availability
All data supporting the findings of this study are included within the article and its supplementary materials.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]Institute of Biophysics of the Czech Academy of Sciences, Brno, Czech Republic
[2]Faculty of Chemistry, Brno University of Technology, Purkyňova 118, Brno 612 00, Czech Republic
[3]School of Chemistry, Pharmacy and Pharmacology, University of East Anglia, Norwich, UK
[4]Centre for Microbial Interactions, Norwich Research Park, Norwich, UK
[5]Laboratoire d'Optique et Biosciences, Ecole Polytechnique, CNRS, INSERM, Institut Polytechnique de Paris, Palaiseau 91120, France

Received: 29 March 2026 / Accepted: 27 May 2026
Published online: 04 June 2026

References

1. Sabater B. Evolution and Function of the Chloroplast. Current Investigations and Perspectives. Int J Mol Sci. 2018;19:3095. https://doi.org/10.3390/ijms19103095.
2. Nomura T, Nakayama N, Murata T, Akazawa T. Biosynthesis of Starch in Chloroplasts 1. Plant Physiol. 1967;42:327–32.
3. Sato N, Ikemura E, Uemura M, Awai K. Genomic and biochemical analyses of lipid biosynthesis in Cyanophora paradoxa: limited role of the chloroplast in fatty acid synthesis. J Exp Bot. 2024;erae420. https://doi.org/10.1093/jxb/erae420.
4. Benz P, Soll J, Bölter B. 17 - The Role of the Tic Machinery in Chloroplast Protein Import. In: Dalbey RE, Koehler CM, Tamanoi F, editors. The Enzymes. Volume 25. Academic; 2007. pp. 439–62. https://doi.org/10.1016/S1874-6047(07)25017-6.
5. Inaba T, Ito-Inaba Y. Versatile Roles of Plastids in Plant Growth and Development. Plant Cell Physiol. 2010;51:1847–53. https://doi.org/10.1093/pcp/pcq147.
6. Dobrogojski J, Adamiec M, Luciński R. The chloroplast genome: a review. Acta Physiol Plant. 2020;42:98. https://doi.org/10.1007/s11738-020-03089-x.
7. Morley SA, Nielsen BL. Chloroplast DNA Copy Number Changes during Plant Development in Organelle DNA Polymerase Mutants. Front Plant Sci. 2016;7:57. https://doi.org/10.3389/fpls.2016.00057.
8. Finkeldey R, Gailing O. Chloroplasts. In: Maloy S, Hughes K, editors. Brenners Encycl. Genet. Second ed. San Diego: Academic; 2013. pp. 525–7. https://doi.org/10.1016/B978-0-12-374984-0.00231-X.
9. Antolin MF, Black WC. Genes, Description of. In: Levin SA, editor. Encycl. Biodivers. Second Ed. Waltham: Academic; 2001. pp. 654–61. https://doi.org/10.1016/B978-0-12-384719-5.00063-0.
10. Kaur P, Kaila T, Dhkal M, Gaikwad K. Chapter 26 - Chloroplast genome and plant–virus interaction. In: Sharma P, Yadav D, Gaur RK, editors. Bioinforma. Agric. Academic; 2022. pp. 419–36. https://doi.org/10.1016/B978-0-323-89778-5.00025-8.
11. Fosket DE. 3 - The Size and Complexity of Plant Genomes. In: Fosket DE, editor. Plant Growth Dev. Boston: Academic; 1994. pp. 79–152. https://doi.org/10.1016/B978-0-12-262430-8.50007-3.
12. Jansen RK, Raubeson LA, Boore JL, dePamphilis CW, Chumley TW, Haberle RC, et al. Methods for obtaining and analyzing whole chloroplast genome sequences. Methods Enzymol. 2005;395:348–84. https://doi.org/10.1016/S0076-6879(05)95020-9.
13. Palmer JD. Contrasting modes and tempos of genome evolution in land plant organelles. Trends Genet. 1990;6:115–20. https://doi.org/10.1016/0168-9525(90)90125-P.
14. Brázda V, Lýsek J, Bartas M, Fojta M. Complex Analyses of Short Inverted Repeats in All Sequenced Chloroplast DNAs. BioMed Res Int. 2018;2018:1097018. https://doi.org/10.1155/2018/1097018.
15. Christinaki AC, Theelen B, Zania A, Coutinho SDA, Cabañes JF, Boekhout T, et al. Co-evolution of large inverted repeats and G-quadruplex DNA in fungal mitochondria may facilitate mitogenome stability: the case of Malassezia. Sci Rep. 2023;13:6308. https://doi.org/10.1038/s41598-023-33486-4.
16. Čutová M, Manta J, Porubiaková O, Kaura P, Šťastný J, Jagelská EB, et al. Divergent distributions of inverted repeats and G-quadruplex forming sequences in Saccharomyces cerevisiae. Genomics. 2020;112:1897–901. https://doi.org/10.1016/j.ygeno.2019.11.002.
17. Burge S, Parkinson GN, Hazel P, Todd AK, Neidle S. Quadruplex DNA: sequence, topology and structure. Nucleic Acids Res. 2006;34:5402–15. https://doi.org/10.1093/nar/gkl655.
18. Rhodes D, Lipps HJ. G-quadruplexes and their regulatory roles in biology. Nucleic Acids Res. 2015;43:8627–37. https://doi.org/10.1093/nar/gkv862.
19. Hänsel-Hertsch R, Di Antonio M, Balasubramanian S. DNA G-quadruplexes in the human genome: detection, functions and therapeutic potential. Nat Rev Mol Cell Biol. 2017;18:279–84. https://doi.org/10.1038/nrm.2017.3.
20. Cimino-Reale G, Zaffaroni N, Folini M. Emerging Role of G-quadruplex DNA as Target in Anticancer Therapy. Curr Pharm Des. 2016;22:6612–24. https://doi.org/10.2174/1381612822666160831101031.
21. Varshney D, Spiegel J, Zyner K, Tannahill D, Balasubramanian S. The regulation and functions of DNA and RNA G-quadruplexes. Nat Rev Mol Cell Biol. 2020;21:459–74. https://doi.org/10.1038/s41580-020-0236-x.
22. Wu F, Niu K, Cui Y, Li C, Lyu M, Ren Y, et al. Genome-wide analysis of DNA G-quadruplex motifs across 37 species provides insights into G4 evolution. Commun Biol. 2021;4:1–11. https://doi.org/10.1038/s42003-020-01643-4.
23. Brázda V, Valková N, Dobrovolná M, Mergny J-L. Abundance of G-Quadruplex Forming Sequences in the Hepatitis Delta Virus Genomes. ACS Omega. 2024;9:4096–101. https://doi.org/10.1021/acsomega.3c09288.
24. Métifiot M, Amrane S, Litvak S, Andreola M-L. G-quadruplexes in viruses: function and potential therapeutic applications. Nucleic Acids Res. 2014;42:12352–66. https://doi.org/10.1093/nar/gku999.
25. Lavezzo E, Berselli M, Frasson I, Perrone R, Palù G, Brazzale AR et al. G-quadruplex forming sequences in the genome of all known human viruses: a comprehensive guide 2018:344127. https://doi.org/10.1101/344127.
26. Cueny RR, McMillan SD, Keck JL. G-quadruplexes in bacteria: insights into the regulatory roles and interacting proteins of non-canonical nucleic acid structures. Crit Rev Biochem Mol Biol. 2022;57:539–61. https://doi.org/10.1080/10409238.2023.2181310.
27. Aktary Z, Sorg K, Cucchiarini A, Vesco G, Noury D, Zhang R et al. Archaeal G-Quadruplexes: A Novel Model for Understanding Unusual DNA/RNA Structures Across the Tree of Life 2025:2024.01.16.575881. https://doi.org/10.1101/2024.01.16.575881.
28. Warner EF, Bohálová N, Brázda V, Waller ZAE, Bidula S. Analysis of putative quadruplex-forming sequences in fungal genomes: novel antifungal targets? Microb Genomics. 2021;7:000570. https://doi.org/10.1099/mgen.0.000570.
29. Middleton G, Mahamud FO, Storer ISR, Williams-Gunn A, Wostear F, Abdolrasouli A, et al. Evidence that G-quadruplexes form in pathogenic fungi and represent promising antifungal targets. EMBO Mol Med. 2025;17:3636–56. https://doi.org/10.1038/s44321-025-00340-1.
30. Capra JA, Paeschke K, Singh M, Zakian VA, G-Quadruplex DNA. Sequences Are Evolutionarily Conserved and Associated with Distinct Genomic Features in Saccharomyces cerevisiae. PLOS Comput Biol. 2010;6:e1000861. https://doi.org/10.1371/journal.pcbi.1000861.
31. Prorok P, Artufel M, Aze A, Coulombe P, Peiffer I, Lacroix L, et al. Involvement of G-quadruplex regions in mammalian replication origin activity. Nat Commun. 2019;10:3274. https://doi.org/10.1038/s41467-019-11104-0.
32. Huppert JL, Balasubramanian S. G-quadruplexes in promoters throughout the human genome. Nucleic Acids Res. 2007;35:406–13. https://doi.org/10.1093/nar/gkl1057.
33. Bohálová N, Dobrovolná M, Brázda V, Bidula S. Conservation and over-representation of G-quadruplex sequences in regulatory regions of mitochondrial DNA across distinct taxonomic sub-groups. Biochimie. 2022;194:28–34. https://doi.org/10.1016/j.biochi.2021.12.006.
34. Bedrat A, Lacroix L, Mergny J-L. Re-evaluation of G-quadruplex propensity with G4Hunter. Nucleic Acids Res. 2016;44:1746–59. https://doi.org/10.1093/nar/gkw006.
35. Brázda V, Kolomazník J, Lýsek J, Bartas M, Fojta M, Šťastný J, et al. G4Hunter web application: a web server for G-quadruplex prediction. Bioinformatics. 2019;35:3493–5. https://doi.org/10.1093/bioinformatics/btz087.

36. Swift ML, GraphPad Prism DA, Graphing S. J Chem Inf Comput Sci. 1997;37:411–2. https://doi.org/10.1021/ci960402j.
37. Cantara A, Luo Y, Dobrovolná M, Bohalova N, Fojta M, Verga D, et al. G-quadruplexes in helminth parasites. Nucleic Acids Res. 2022;50:2719–35. https://doi.org/10.1093/nar/gkac129.
38. Porubiaková O, Havlík J, Indu null, Šedý M, Přepechalová V, Bartas M, et al. Variability of Inverted Repeats in All Available Genomes of Bacteria. Microbiol Spectr. 2023;11:e0164823. https://doi.org/10.1128/spectrum.01648-23.
39. Brázda V, Luo Y, Bartas M, Kaura P, Porubiaková O, Šťastný J, et al. G-Quadruplexes in the Archaea Domain. Biomolecules. 2020;10:1349. https://doi.org/10.3390/biom10091349.
40. Bohálová N, Cantara A, Bartas M, Kaura P, Šťastný J, Pečinka P, et al. Analyses of viral genomes for G-quadruplex forming sequences reveal their correlation with the type of infection. Biochimie. 2021;186:13–27. https://doi.org/10.1016/j.biochi.2021.03.017.
41. Esnault C, Magat T, El Zine A, Garcia-Oliver E, Cucchiarini A, Bouchouika S, et al. G4access identifies G-quadruplexes and their associations with open chromatin and imprinting control regions. Nat Genet. 2023;55:1359–69. https://doi.org/10.1038/s41588-023-01437-4.
42. Takahashi H, Nakagawa A, Kojima S, Takahashi A, Cha B-Y, Woo J-T, et al. Discovery of novel rules for G-quadruplex-forming sequences in plants by using bioinformatics methods. J Biosci Bioeng. 2012;114:570–5. https://doi.org/10.1016/j.jbiosc.2012.05.017.
43. Shim H, Lee HJ, Lee J, Lee H-O, Kim J-H, Yang T-J, et al. Plastid Genomes of the Early Vascular Plant Genus Selaginella Have Unusual Direct Repeat Structures and Drastically Reduced Gene Numbers. Int J Mol Sci. 2021;22:641. https://doi.org/10.3390/ijms22020641.
44. Ariyarathne MA, Wone BWM. Overexpression of the *Selaginella lepidophylla* bHLH transcription factor enhances water-use efficiency, growth, and development in *Arabidopsis*. Plant Sci. 2022;315:111129. https://doi.org/10.1016/j.plantsci.2021.111129.
45. Yang X, Yu H, Duncan S, Zhang Y, Cheema J, Liu H, et al. RNA G-quadruplex structure contributes to cold adaptation in plants. Nat Commun. 2022;13:6224. https://doi.org/10.1038/s41467-022-34040-y.
46. Ren M, Li S, Gao Q, Qiao L, Cao Q, Yang Z, et al. Advances in the Anti-Tumor Activity of Biflavonoids in Selaginella. Int J Mol Sci. 2023;24:7731. https://doi.org/10.3390/ijms24097731.
47. Coronado-Parra T, Roldán M, Aboal M. Autofluorescence Imaging to Evaluate Red Algae Physiology. J Vis Exp JoVE. 2023. https://doi.org/10.3791/64533.
48. Voerman SE, Ruseckas A, Turnbull GA, Samuel IDW, Burdett HL. Red algae acclimate to low light by modifying phycobilisome composition to maintain efficient light harvesting. BMC Biol. 2022;20:291. https://doi.org/10.1186/s12915-022-01480-3.
49. Barkan A. Expression of Plastid Genes: Organelle-Specific Elaborations on a Prokaryotic Scaffold1. Plant Physiol. 2011;155:1520–32. https://doi.org/10.1104/pp.110.171231.
50. Börner T, Aleynikova AY, Zubo YO, Kusnetsov VV. Chloroplast RNA polymerases: Role in chloroplast biogenesis. Biochim Biophys Acta. 2015;1847:761–9. https://doi.org/10.1016/j.bbabio.2015.02.004.
51. Spiegel J, Cuesta SM, Adhikari S, Hänsel-Hertsch R, Tannahill D, Balasubramanian S. G-quadruplexes are transcription factor binding hubs in human chromatin. Genome Biol. 2021;22:117. https://doi.org/10.1186/s13059-021-02324-z.
52. Barkan A, Small I. Pentatricopeptide repeat proteins in plants. Annu Rev Plant Biol. 2014;65:415–42. https://doi.org/10.1146/annurev-arplant-050213-040159.
53. Schmitz-Linneweber C, Small I. Pentatricopeptide repeat proteins: a socket set for organelle gene expression. Trends Plant Sci. 2008;13:663–70. https://doi.org/10.1016/j.tplants.2008.10.001.
54. Kamba PF, Dickson DA, White NA, Ekstrom JL, Koslowsky DJ, Hoogstraten CG. The 27 kDa Trypanosoma brucei Pentatricopeptide Repeat Protein is a G-tract Specific RNA Binding Protein. Sci Rep. 2018;8:16989. https://doi.org/10.1038/s41598-018-34377-9.
55. Everitt RG, Didelot X, Batty EM, Miller RR, Knox K, Young BC, et al. Mobile elements drive recombination hotspots in the core genome of Staphylococcus aureus. Nat Commun. 2014;5:3956. https://doi.org/10.1038/ncomms4956.
56. Kim N. G4-interacting proteins endangering genomic stability at G4 DNA-forming sites. Biochem Soc Trans. 2023;51:403–13. https://doi.org/10.1042/BST20221018.
57. Mendoza O, Bourdoncle A, Boulé J-B, Brosh RM, Mergny J-L. G-quadruplexes and helicases. Nucleic Acids Res. 2016;44:1989–2006. https://doi.org/10.1093/nar/gkw079.
58. Zhang M-H, Xiang Q-P, Zhang X-C. Plastid phylogenomic analyses of the *Selaginella sanguinolenta* group (Selaginellaceae) reveal conflict signatures resulting from sequence types, outlier genes, and pervasive RNA editing. Mol Phylogenet Evol. 2022;173:107507. https://doi.org/10.1016/j.ympev.2022.107507.
59. Wu C-S, Chaw S-M. Evolutionary Stasis in Cycad Plastomes and the First Case of Plastome GC-Biased Gene Conversion. Genome Biol Evol. 2015;7:2000–9. https://doi.org/10.1093/gbe/evv125.
60. Daniell H, Lin C-S, Yu M, Chang W-J. Chloroplast genomes: diversity, evolution, and applications in genetic engineering. Genome Biol. 2016;17:134. https://doi.org/10.1186/s13059-016-1004-2.
61. Wicke S, Schneeweiss GM, dePamphilis CW, Müller KF, Quandt D. The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. Plant Mol Biol. 2011;76:273–97. https://doi.org/10.1007/s11103-011-9762-4.
62. Xie Z. The Methods and Tools for Mobile Genetic Element Detection and their Application to Systems Medicine. In: Wolkenhauer O, editor. Syst. Med. Oxford: Academic; 2021. pp. 203–7. https://doi.org/10.1016/B978-0-12-801238-3.11537-5.
63. Haudiquet M, de Sousa JM, Touchon M, Rocha EPC. Selfish, promiscuous and sometimes useful: how mobile genetic elements drive horizontal gene transfer in microbial populations. Philos Trans R Soc B Biol Sci. 2022;377:20210234. https://doi.org/10.1098/rstb.2021.0234.
64. Kojima S, Koyama S, Ka M, Saito Y, Parrish EH, Endo M, et al. Mobile element variation contributes to population-specific genome diversification, gene regulation and disease risk. Nat Genet. 2023;55:939–51. https://doi.org/10.1038/s41588-023-01390-2.
65. Partridge SR, Kwong SM, Firth N, Jensen SO. Mobile Genetic Elements Associated with Antimicrobial Resistance. Clin Microbiol Rev. 2018;31:e00088–17. https://doi.org/10.1128/CMR.00088-17.
66. Hull DM, Harrell E, Harden L, Thakur S. Multidrug resistance and virulence genes carried by mobile genomic elements in Salmonella enterica isolated from live food animals, processed, and retail meat in North Carolina, 2018–2019. Int J Food Microbiol. 2022;378:109821. https://doi.org/10.1016/j.ijfoodmicro.2022.109821.
67. Garg R, Aggarwal J, Thakkar B. Genome-wide discovery of G-quadruplex forming sequences and their functional relevance in plants. Sci Rep. 2016;6:28211. https://doi.org/10.1038/srep28211.
68. Dobrovolná M, Bohálová N, Peška V, Wang J, Luo Y, Bartas M, et al. The Newly Sequenced Genome of Pisum sativum Is Replete with Potential G-Quadruplex-Forming Sequences-Implications for Evolution and Biological Regulation. Int J Mol Sci. 2022;23:8482. https://doi.org/10.3390/ijms23158482.
69. Song K, Li B, Li H, Zhang R, Zhang X, Luan R, et al. The Characterization of G-Quadruplexes in Tobacco Genome and Their Function under Abiotic Stress. Int J Mol Sci. 2024;25:4331. https://doi.org/10.3390/ijms25084331.
70. Lang BF, Nedelcu AM. Plastid Genomes of Algae. In: Bock R, Knoop V, editors. Genomics Chloroplasts Mitochondria. Dordrecht: Springer Netherlands; 2012. pp. 59–87. https://doi.org/10.1007/978-94-007-2920-9_3.
71. Nickelsen J, Bohne A-V. Plastid Transformation in Algae. In: Bock R, Knoop V, editors. Genomics Chloroplasts Mitochondria. Dordrecht: Springer Netherlands; 2012. pp. 379–92. https://doi.org/10.1007/978-94-007-2920-9_16.
72. Jansen RK, Ruhlman TA. Plastid Genomes of Seed Plants. In: Bock R, Knoop V, editors. Genomics Chloroplasts Mitochondria. Dordrecht: Springer Netherlands; 2012. pp. 103–26. https://doi.org/10.1007/978-94-007-2920-9_5.
73. Turmel M, Otis C, Lemieux C. Divergent copies of the large inverted repeat in the chloroplast genomes of ulvophycean green algae. Sci Rep. 2017;7:994. https://doi.org/10.1038/s41598-017-01144-1.
74. Maréchal A, Parent J-S, Véronneau-Lafortune F, Joyeux A, Lang BF, Brisson N. Whirly proteins maintain plastid genome stability in Arabidopsis. Proc Natl Acad Sci U S A. 2009;106:14693–8. https://doi.org/10.1073/pnas.0901710106.
75. Rowan BA, Oldenburg DJ, Bendich AJ. RecA maintains the integrity of chloroplast DNA molecules in Arabidopsis. J Exp Bot. 2010;61:2575–88. https://doi.org/10.1093/jxb/erq088.
76. Maréchal A, Brisson N. Recombination and the maintenance of plant organelle genome stability. New Phytol. 2010;186:299–317. https://doi.org/10.1111/j.1469-8137.2010.03195.x.
77. Huppert JL, Balasubramanian S. Prevalence of quadruplexes in the human genome. Nucleic Acids Res. 2005;33:2908–16. https://doi.org/10.1093/nar/gki609.
78. Sardans J, Peñuelas J. Potassium Control of Plant Functions: Ecological and Agricultural Implications. Plants. 2021;10:419. https://doi.org/10.3390/plants10020419.

Publisher's note

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