

## Article

# Analysis of G-Quadruplex-Forming Sequences in Drought Stress-Responsive Genes, and Synthesis Genes of Phenolic Compounds in *Arabidopsis thaliana*

Petr Pečinka <sup>1</sup>, Natália Bohálová <sup>2</sup>, Adriana Volná <sup>3</sup>, Kristýna Kunderátová <sup>1</sup>, Václav Brázda <sup>2</sup>  
and Martin Bartas <sup>1,\*</sup>

<sup>1</sup> Department of Biology and Ecology, University of Ostrava, 70833 Ostrava, Czech Republic

<sup>2</sup> Department of Biophysical Chemistry and Molecular Oncology, Institute of Biophysics of the Czech Academy of Sciences, 61265 Brno, Czech Republic

<sup>3</sup> Department of Physics, University of Ostrava, 70833 Ostrava, Czech Republic

\* Correspondence: martin.bartas@osu.cz

**Abstract:** Sequences of nucleic acids with the potential to form four-stranded G-quadruplex structures are intensively studied mainly in the context of human diseases, pathogens, or extremophile organisms; nonetheless, the knowledge about their occurrence and putative role in plants is still limited. This work is focused on G-quadruplex-forming sites in two gene sets of interest: drought stress-responsive genes, and genes related to the production/biosynthesis of phenolic compounds in the model plant organism *Arabidopsis thaliana*. In addition, 20 housekeeping genes were analyzed as well, where the constitutive gene expression was expected (with no need for precise regulation depending on internal or external factors). The results have shown that none of the tested gene sets differed significantly in the content of G-quadruplex-forming sites, however, the highest frequency of G-quadruplex-forming sites was found in the 5'-UTR regions of phenolic compounds' biosynthesis genes, which indicates the possibility of their regulation at the mRNA level. In addition, mainly within the introns and 1000 bp flanks downstream gene regions, G-quadruplex-forming sites were highly underrepresented. Finally, cluster analysis allowed us to observe similarities between particular genes in terms of their PQS characteristics. We believe that the original approach used in this study may become useful for further and more comprehensive bioinformatic studies in the field of G-quadruplex genomics.

**Keywords:** G-quadruplex; PQS; *Arabidopsis thaliana*; drought stress; phenolic compounds



**Citation:** Pečinka, P.; Bohálová, N.; Volná, A.; Kunderátová, K.; Brázda, V.; Bartas, M. Analysis of G-Quadruplex-Forming Sequences in Drought Stress-Responsive Genes, and Synthesis Genes of Phenolic Compounds in *Arabidopsis thaliana*. *Life* **2023**, *13*, 199. <https://doi.org/10.3390/life13010199>

Academic Editors: Zhongjie Ding and Yanglin Ding

Received: 10 December 2022

Revised: 30 December 2022

Accepted: 8 January 2023

Published: 10 January 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

While the majority of the DNA in genomes is organized into a Watson–Crick B-DNA structure [1], the presence of various different structures with functional properties has been recently proved in vitro and in vivo [2]. Such structures are, thus, designated as Non-B-DNA structures or noncanonical nucleic acids. Examples of Non-B-DNA structures are left-handed DNA (Z-DNA), A-DNA (varying in the number of base pairs per turn, twist, and rise), i-motifs, R-loops, hairpins, and cruciforms [3]. It was documented that these structures are more common in living organisms than has been assumed. The presence of these noncanonical structures was confirmed even for plant species [4]. Nowadays, increased scientific effort has been dedicated to understanding the complex role of several noncanonical nucleic acids in plants under suboptimal conditions [5]. On the other hand, the literature dealing with plant G-quadruplexes (G4s) is still very limited, especially in the context of plant stress responses.

G4s are four-stranded DNA structures, which are widely distributed across all three domains of life (Bacteria, Archaea, and Eukaryota) [6–8]. The basic building block of G4s—a guanine tetrad—was first described in 1962 by Gellert, Lipsett, and Davies [9],

who experimented with guanylic acid and solved its crystal structure. A single guanine tetrad consists of four guanines which are oriented one to another by the angle of  $90^\circ$  and represent one stack of G4. Although it was assumed that mostly two-, three-, and four-stacked G4s are present in the genomes, it was shown, that even five- and six-stacked G4s exist and their formation was confirmed in vitro [10]. Such structure stability can be affected by the metal ions (especially by the monovalent potassium and sodium ions) concentration [11] and their presence within or outside the guanine tetrad (depending on the ion size). Although G4s are currently at the front of interest mainly in cancer research, they have an irreplaceable role in gene expression regulation in general. In plants, both DNA and RNA G4s are documented, and their formation can affect replication [12], transcription [13], and even translation [14] (when the G4 is formed for example in the 5'-UTR of the mRNA [15]). Although in most cases the G4s formation has a repressive effect (suppresses replication, transcription, or translation), recently it was documented that the G4s formation in the coding gene region can have a stimulative effect as well [16]. Regarding gene expression inhibition, G4s act either as a roadblock to RNA polymerases or as a physical obstacle to ribosome entry when located within 5' UTRs of mRNA [17].

An emerging number of studies dealing with the putative G-quadruplex sites (PQS) prediction in plant genomes appeared, but the number of the in vivo confirmed G4s by their direct detection remains very limited (approximately 1–5% from the predicted PQS—[16]). This fact indicates that the G4s formation can be affected by the outer environmental factors and thus can be much more plastic than was previously assumed. Similar knowledge limitation has to be faced when it comes to our understanding of the connection between G4s formation and stress responses. As the fact that G4s formation is affected by the ion concentrations can be concluded, especially drought, high salinity concentration, or other external factors affecting the concentration of ions available in the cytoplasm should be linked with the major changes in G4 occurrence.

In this study, we decided to search for the PQS in two sets of a well-defined group of genes (the drought stress-responsive genes, and the genes of phenolic compounds synthesis pathways) to find out whether the PQS occurrence significantly varies from the average in the whole *Arabidopsis thaliana* genome and also in comparison with the subset of constitutively expressed (housekeeping) genes. We hypothesize that distinct functional gene sets will have various PQS patterns. Drought stress-responsive genes and phenolic compounds synthesis genes were chosen because these two groups of (at the first sight nonrelated) genes are at the center of our research interest, but nowadays it is evident that drought also influences the synthesis of phenolic compounds [18–20]. Therefore, we were interested in whether the PQS patterns of phenolic compounds biosynthesis genes will be similar to the drought stress-responsive genes, or not. Nonetheless, the main aim of this article is to provide a novel methodological approach regarding how to analyze PQS patterns in distinct gene sets. To the best of our knowledge, this is the first study analyzing PQS from such a point of view (in the context of defined gene sets from a single organism).

## 2. Materials and Methods

### 2.1. Data Acquisition and PQS Prediction

FASTA files of the *Arabidopsis thaliana* drought stress-responsive genes (118 genes in total) obtained from the DroughtDB expert-curated database ([https://pgsb.helmholtz-muenchen.de/droughtdb/drought\\_db.html](https://pgsb.helmholtz-muenchen.de/droughtdb/drought_db.html), accessed on 24 July 2022) [21], phenolic compounds biosynthesis genes (30 genes in total), and housekeeping genes (20 genes in total), and FASTA files of their close surroundings 1000 bp upstream (+1000) and downstream (−1000) were acquired from the NCBI database (<https://www.ncbi.nlm.nih.gov/>, accessed on 30 July 2022). Obtained sequences were processed using our G4Hunter web server (<https://bioinformatics.ibp.cz/#/analyse/quadruplex>, accessed on 15 August 2022) [22] with the G4Score threshold of 1.2 and window size 25 nt. The annotations of the corresponding chromosome loci were downloaded from the NCBI database as well (where the gene of interest is localized), together with 1000 bp upstream and downstream regions.

Other genes and annotated regions overlapped with the genes of interest were manually inspected and removed. For purposes of annotation, only a single isoform per each gene was selected based on the criteria of whether it is a canonical isoform according to UniProt (<https://www.uniprot.org/>, accessed on 2 September 2022). In case there were more transcripts per one canonical isoform in NCBI, the longer transcript was always chosen. The positions of introns were specified together with 5'-UTRs and 3'-UTRs. Finally, the overlap of corrected/modified annotations with G4Hunter results was calculated.

## 2.2. Data Visualization and Statistical Analysis

A schematic figure of model gene features composition was made using Biorender (<https://biorender.com/>, accessed on 5 November 2022). Barplots of PQS counts and frequencies within particular NCBI features were made in Microsoft Excel 2021. Statistical analysis of PQS differences between gene sets was performed using the ANOVA test and plotted in GraphPad Prism 9. A heatmap comprising all analyzed genes with non-zero PQS occurrence either within the gene or its close surrounding ( $\pm 1000$  bp) was constructed using the Heatmapper tool [23] (<http://www.heatmapper.ca/expression/>, accessed on 11 November 2022). The single linkage clustering method and Kendall's Tau distance measurement method were used to obtain a meaningful number of resulting clusters. All numerical values were normalized by Row Z-score.

## 3. Results

### 3.1. Global Analysis of PQS Occurrence in the *Arabidopsis thaliana* Genome

In the whole *Arabidopsis thaliana* genome (TAIR 10, comprising five chromosomes, mitochondrial, and plastid DNA), a total of 41,557 PQS were found (Table 1). It may be seen that PQS frequencies per particular genomic chromosomes are very similar in the range of 0.32 to 0.36 PQS per 1000 bp (the mean PQS frequency in genomic DNA was 0.344 PQS per 1000 bp). The highest PQS frequency (1.34 PQS per 1000 bp) was found in mtDNA, which may be partly due to significantly higher GC content (44.79%), than in genomic DNA (around 36%). Plastid DNA contains 0.81 PQS per 1000 bp, which is more than twice as high as genomic DNA with similar GC content.

**Table 1.** PQS counts and frequencies in the whole *Arabidopsis thaliana* genome.

Chromosome	NCBI ID	Length	GC Content	PQS Count	Frequency per 1000 bp
chr1	NC_003070.9	30,427,671	35.68	10,989	0.36
chr2	NC_003071.7	19,698,289	35.86	6613	0.34
chr3	NC_003074.8	23,459,830	36.32	7893	0.34
chr4	NC_003075.7	18,585,056	36.20	6781	0.36
chr5	NC_003076.8	26,975,502	35.93	8664	0.32
mtDNA	NC_037304.1	367,808	44.79	492	1.34
cpDNA	NC_000932.1	154,478	36.29	125	0.81

### 3.2. Descriptive Analyses of PQS Found within Different Gene Sets and NCBI Features Overlay

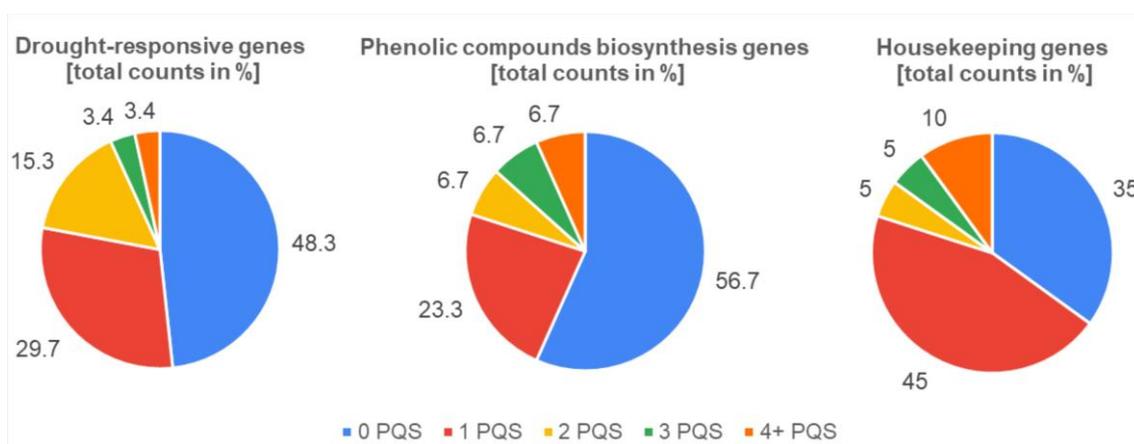
The following section summarizes PQS counts and frequencies in three analyzed gene sets (drought stress-responsive genes, phenolic compounds synthesis genes, and housekeeping genes). Even more detailed results are enclosed in Supplementary Table S1 (PQS counts and frequencies) and Supplementary Table S2 (NCBI features overlay).

In 118 drought stress-responsive genes analyzed, there were 102 PQS found (gene regions only), or 160 PQS when we are also considering 1000 bp flanks before and after genes. Overall, 35 genes contained exactly one PQS (29.7%). Some genes contained more than one PQS (26 in total, 22%), and some genes had none (a total of 57 genes, 48.3%). The highest number of PQS was found within the 3232 bp long gene HD1 (a total of six PQS).

This gene encodes for Histone deacetylase 1, which is an important chromatin-remodeling factor that contributes to transcriptional repression in eukaryotic organisms.

In 30 phenolic compounds synthesis genes analyzed, there were 27 PQS found (gene regions only), or 43 PQS (considering also 1000 bp flanks). Seven genes contained exactly one PQS (23.3%). Six genes contained more than one PQS (20%), and 17 genes had no PQS (56.7%). The highest number of PQS was found within the 1539 bp long gene TT5 (a total of 5 PQS). This gene encodes for the Chalcone-flavanone isomerase family protein which catalyzes the conversion of chalcone into flavanone.

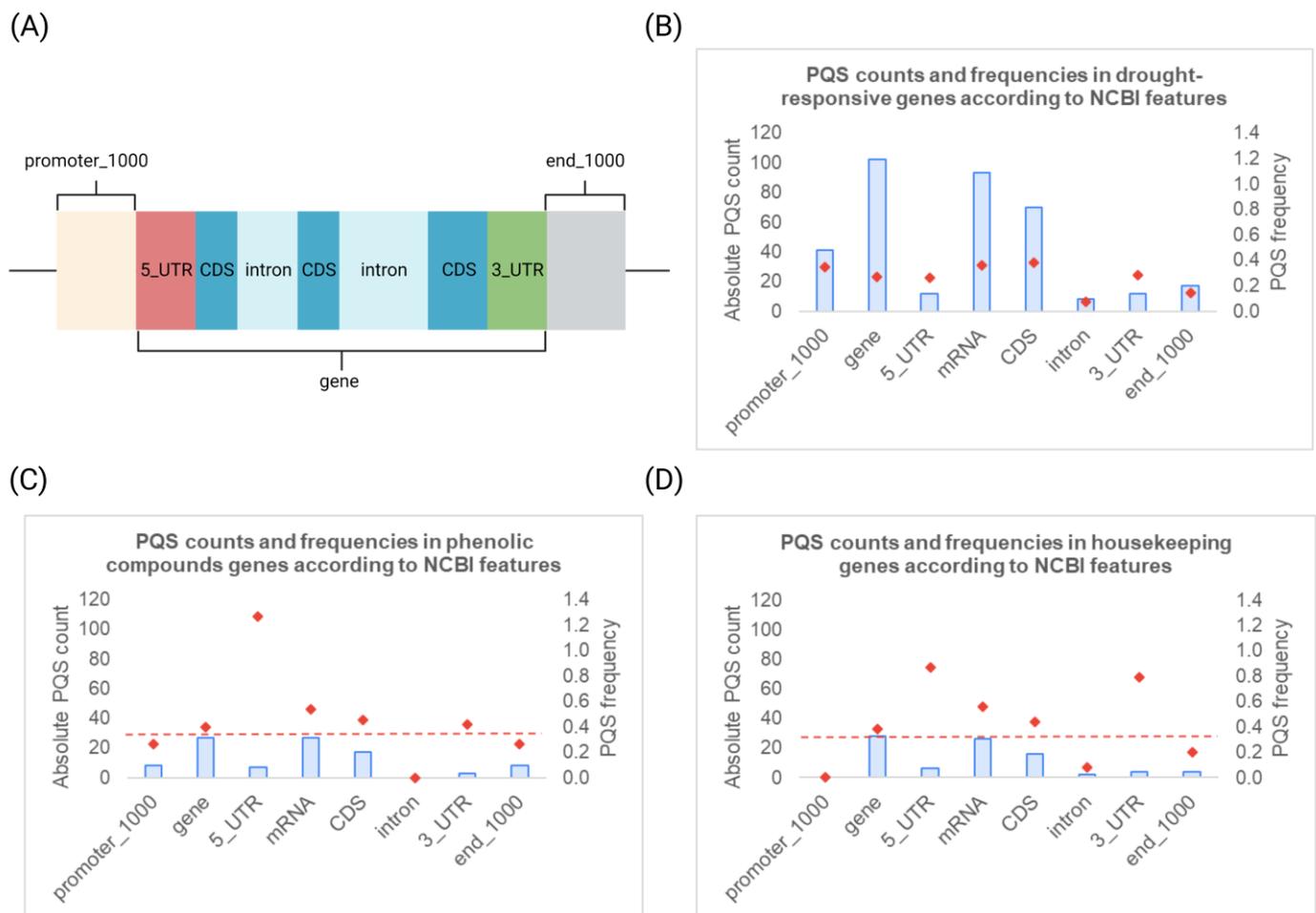
In the set of 20 housekeeping genes analyzed, there were 28 PQS found (gene regions only), or 36 PQS (considering also 1000 bp flanks). Nine genes contained exactly one PQS. Four genes contained more than one PQS (20%), and seven genes had no PQS (35%). The highest number of PQS was found within the 10,519 bp long gene HAF1 (a total of nine PQS). This gene encodes for the HAC13 protein (Histone acetyltransferase of the CBP family 13). The overall results of the percentual distribution of total PQS count in all three gene sets analyzed are depicted in Figure 1.



**Figure 1.** Percentual distribution of total PQS counts in different gene sets. All analyzed genes were divided into five categories according to their total PQS counts (zero, one, two, three, four, and more PQS).

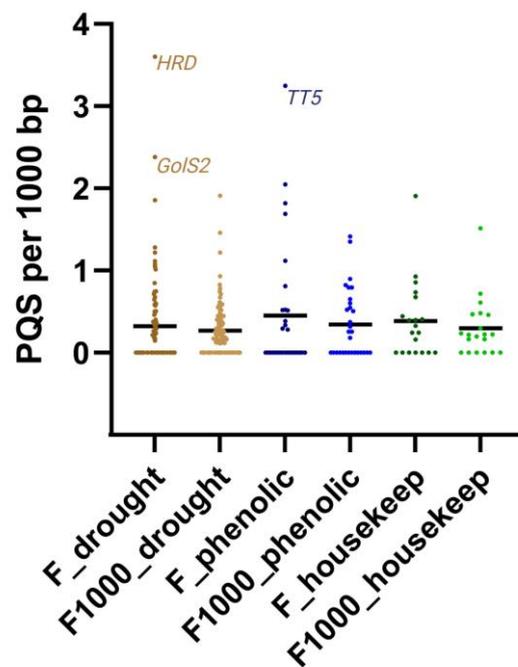
Surprisingly, the gene set of housekeeping genes contained the smallest proportion of genes without any PQS, in other words, 65% of housekeeping genes analyzed contained at least one PQS. Next, our analysis revealed that the occurrence of PQS varies notably between distinct genomic features. Because the length of analyzed genes significantly differs and could potentially skew the results, we have always also taken into account a PQS frequency per 1000 bp. Regarding particular gene features (schematically depicted in Figure 2A), the highest PQS frequency per 1000 bp was observed within 5'-UTR regions in phenolic compounds synthesis genes (1.27 PQS per 1000 bp), whereas the lowest PQS frequency per 1000 bp was observed within putative promoter regions of housekeeping genes (0 PQS per 1000 bp), and within intronic regions of phenolic compounds synthesis genes (0 PQS per 1000 bp), then by 0.07 PQS per 1000 bp in drought stress-responsive genes, and 0.08 PQS per 1000 bp in housekeeping genes (Figure 2B–D).

To statistically compare differences in PQS frequencies between particular gene sets we carried out a one-way ANOVA test (separately for gene regions only, and for gene regions with 1000 bp flanks). There were no statistically significant differences between gene sets. However, this may be caused by the small sample size of gene sets, as well as the fact that the data are extremely skewed (many zero “PQS per 1000 bp” values). The distribution of PQS frequencies within particular gene sets is shown in Figure 3.



**Figure 2.** Analyses of PQS frequencies within NCBI features. **(A)** Scheme of NCBI features used in this study to categorize DNA regions for PQS analysis. **(B)** PQS counts and relative frequencies in drought-responsive genes. **(C)** PQS counts and relative frequencies in phenolic compounds synthesis genes. **(D)** PQS counts and relative frequencies in selected housekeeping genes. Blue bars express the total PQS counts found in particular NCBI features, whereas red marks depict relative PQS frequencies (PQS count per 1000 bp). Red dashed horizontal lines indicate the mean PQS frequency (0.344 PQS per 1000 bp) in the *Arabidopsis thaliana* genome.

Considering particular genes, the highest PQS frequency (up to tenfold compared to the others) was found in the *HRD* gene from the drought-responsive gene set and in the *TT5* gene from the phenolic compounds biosynthesis gene set. The *HRD* gene encodes HRD Integrase-type DNA-binding superfamily protein, and its overexpression increases the density of the roots and improves water and salt stress tolerance in *Arabidopsis thaliana* [24]. These genes could therefore have their regulation dependent on the formation of G4s.

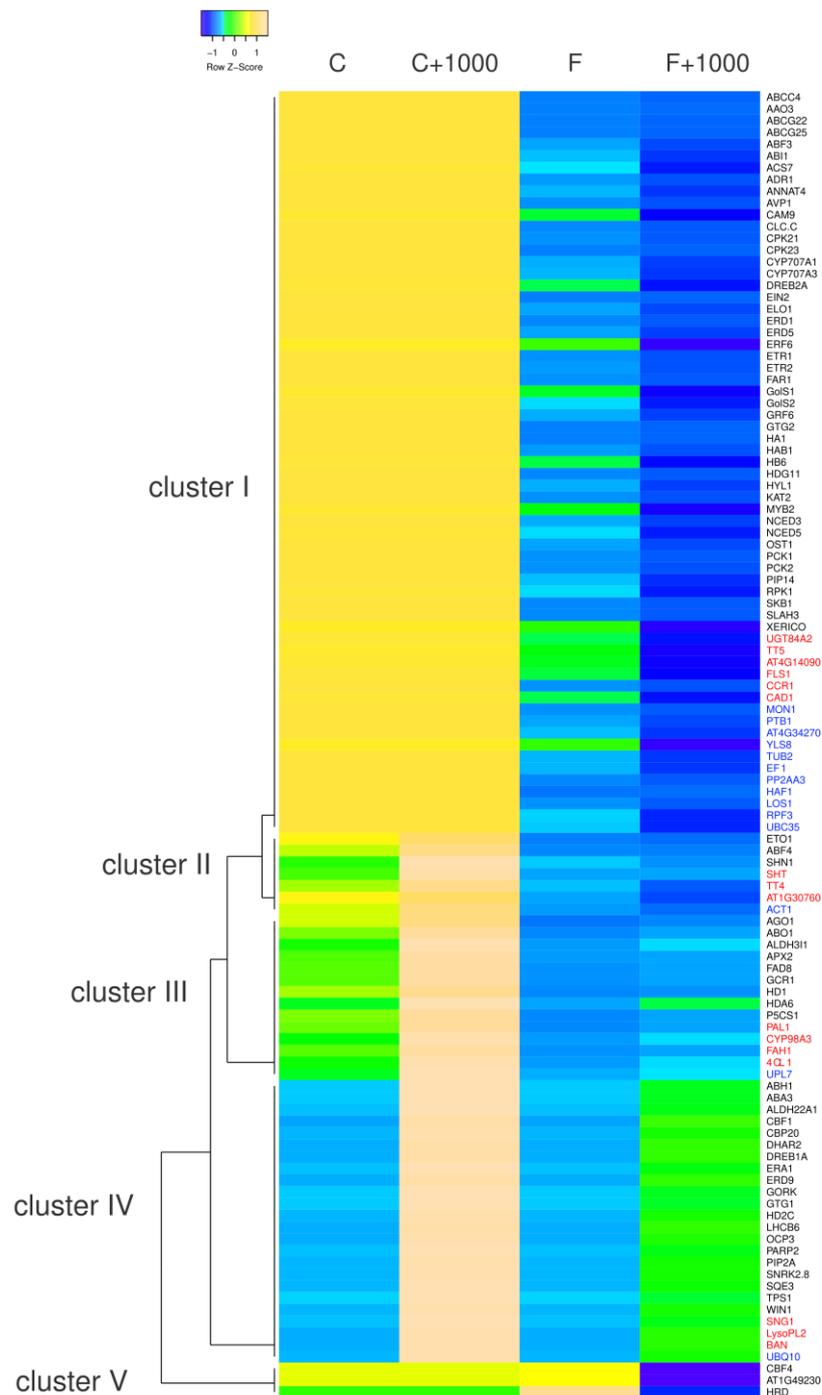


**Figure 3.** Gene sets comparison. Distributions of all “PQS per 1000 bp” values in analyzed gene sets are shown. F stands for PQS frequencies within gene regions only, and F1000 stands for PQS frequencies in genes with 1000 bp flanks on both sides. Black horizontal lines indicate mean PQS frequency values in particular gene sets, and three outliers/genes with the highest PQS frequencies are depicted.

### 3.3. Clustering of PQS Patterns in Analyzed Genes

To inspect PQS patterns in analyzed genes containing at least one PQS, we carried out a cluster analysis based on these four parameters: total PQS counts in the whole gene regions; total PQS counts in gene regions with 1000 bp flanks on both ends; PQS frequencies per 1000 bp in the whole genes; and PQS frequencies in gene regions with 1000 bp flanks on both ends. As a result, five main clusters were obtained (Figure 4). Cluster I (the biggest one) contained 46 genes, cluster II contained 7 genes, cluster III contained 14 genes, cluster IV contained 24 genes, and cluster V (the smallest one) contained 3 genes (out of 111 genes containing at least 1 PQS either in gene regions or within 1000 bp flanks). Noticeably, in cluster V there were only drought-responsive genes with high PQS frequencies in gene regions, but not in 1000 bp flanks, where PQS frequencies were low. On the other hand, cluster IV gathered genes with relatively high PQS frequencies when also considering 1000 bp flanks, but low PQS frequencies in gene regions themselves. Clusters III and II were quite similar and had low PQS frequencies both within gene regions as well as 1000 bp flanks. The largest cluster I contained genes with high total PQS counts, variable PQS frequencies in gene regions, and low PQS frequencies in 1000 bp flanks. Finally, it can be seen that each of the analyzed gene sets contained both “PQS-rich” and “PQS-poor” genes.

### Heatmap of PQS patterns in analyzed genes and potential regulatory regions



**Figure 4.** Heatmap comprising all genes with non-zero PQS occurrence either within the gene or its close surrounding ( $\pm 1000$  bp) analyzed in this study. The column designated C is based on absolute PQS counts within gene regions, and column F is based on PQS frequencies per 1000 bp within gene regions. Columns C+1000 and F+1000 are focused on PQS counts and frequencies within the close surrounding of genes ( $\pm 1000$  bp). Five resulting clusters of genes are designated on the left side. The single linkage clustering method and Kendall’s Tau distance measurement method were applied, and all numerical values were normalized by Row Z-score. Drought-responsive gene names are in black, phenolic compounds biosynthesis gene names are in red, and housekeeping gene names are in blue.

#### 4. Discussion

Our previously published approach for PQS prediction showed that they are non-randomly distributed in the *Pisum sativum* genome [25], where the frequency of PQSs for nuclear DNA was 0.42 PQSs per 1000 bp in nuclear DNA, 0.53 PQS per 1000 bp in plastid DNA, and 1.58 PQS per 1000 bp in mitochondrial DNA. A similar trend is observable also in current data from *Arabidopsis thaliana*, where the frequency of 0.34 PQS per 1000 bp was found in nuclear DNA, a higher frequency of 0.81 in plastid DNA, and the highest frequency of 1.34 in mitochondrial DNA. This result indicates the importance of PQS in circular and especially mitochondrial genomes. While in algae and fungi the frequency of PQS in mtDNA is lower than in plants, including *Arabidopsis thaliana*, in vertebrates the frequency of PQS in mtDNA is even significantly higher [26]. Although such large-scale predictions are computationally extensive, it would be beneficial to also inspect other higher plants in the future to see if this trend (frequency of PQS in nuclear DNA < frequency of PQS in plastid DNA < frequency of PQS in mitochondrial DNA) is always the same. These data should then be correlated with GC content as well because generally higher GC content may favor more PQS, although recently it was shown that genomes with a very similar GC content may differ significantly in their PQS frequencies [27]. Considering nuclear and plastid genomes which have very similar GC content in *Arabidopsis thaliana*, an explanation for different PQS frequencies may be found in significantly different gene densities. In the nuclear genome, there are 31.8 genes per 100,000 bp, whereas in the plastid genome, more than twice the genes (83.5 per 100,000 bp) are present (unpublished data, computation is based on the information from the NCBI database, [https://www.ncbi.nlm.nih.gov/data-hub/genome/GCF\\_000001735.4/](https://www.ncbi.nlm.nih.gov/data-hub/genome/GCF_000001735.4/), accessed on 28 December 2022). Higher PQS frequencies could therefore positively correlate with higher gene density in the plastid genome because each gene is a potential source of PQS-rich regulatory features (e.g., 5'-UTRs). In addition, this article has shown that the PQS frequencies within the gene regions were generally higher than the average PQS frequency of the whole *Arabidopsis thaliana* genome (comprising both genic and intergenic loci).

Our recent analysis in *Arabidopsis thaliana* further revealed that both the total number and frequencies of PQS within introns were very low (at least in three gene sets analyzed in this study), as only 10 PQS were found in these locations in total (8 in drought-responsive genes, 2 in housekeeping genes, and 0 in phenolic compounds synthesis genes). On the other hand, PQS frequencies within CDS were much higher. This fact further strengthens the hypothesis about the regulatory potential of not only DNA G4s, but also RNA G4s which may be formed in mRNA prior to or during the translation (PQS-poor introns are excised during the mRNA processing by the spliceosome complex). It is worth noting that the mean and median lengths of introns are only 168 bp, or 100 bp, respectively, in *Arabidopsis thaliana* (for comparison, in maize, the mean intron length is 516 bp and the median is 146 bp; and in humans, the mean intron length is 3356 bp and the median is 1023 bp) [28]. With such small intron sizes in *Arabidopsis thaliana*, PQS (usually 25 bp long) would span a substantial part of them, and maybe this is one of the reasons why they are depleted there—to avoid potential problems with RNA splicing.

A similar trend as in the case of introns was observed in the regions 1000 bp downstream of genes (end\_1000)—both absolute and relative PQS frequencies were very low. Such an observation would make sense, as the regulation behind the gene region would not be necessary compared to the need for precise regulation in the promoter sequence of the same gene. Nonetheless, in further studies, a whole genome of (not only) *Arabidopsis thaliana* could be analyzed by individual genes/gene sets and their features to gain an even more comprehensive picture of PQS distribution within functional units/pathways of genome and transcriptome.

One of the most interesting things found in this study is the fact that in all three gene sets analyzed, there were always slightly higher PQS frequencies in CDS than in the whole *Arabidopsis thaliana* genome (mean PQS frequency). Traditionally it has been assumed that the frequency of PQS in noncoding regions, such as promoters, telomeres [29],

and untranslated regions [30], is much higher than within coding regions. Nonetheless, additional data are needed to possibly generalize this observation for the plants (and other eukaryotic genomes).

In plants, there is still very little evidence about the mechanistic function of particular PQS and corresponding G4. One of the first studies dealing with such problematics was published by Cho et al. [14], focusing on particular PQS in 5'-UTR regions of *SMXL4* and *SMXL5* genes in *Arabidopsis thaliana*. These genes are responsible for phloem differentiation. It was observed that the formation of G4s within 5'-UTR of mRNA molecule (induced/stabilized by G4-binding protein JULGI) causes suppression of *SMXL4/5* translation and restricts phloem differentiation [14]. Another work focused on a single PQS was published by Kwok et al. [31], where they identified PQS and corresponding G4 within the 5'-UTR of the *ATR* gene in *Arabidopsis thaliana*. Transient reporter gene assay further revealed that this G4 inhibits translation, but not the transcription of the *ATR* gene in living plants [31]. G4s also have potential in applied/agronomic science, as it is believed that their formation and stability (and thereby the expression of genes containing PQS) could be affected by environmental conditions, e.g., macronutrient availability (mainly K<sup>+</sup> and Na<sup>+</sup> ions) and/or drought [30], temperature [32], irradiation [33], etc. From this point of view, it would be interesting to know which genes and pathways contain the highest amount/frequencies of PQS. All in all, we believe that this study and used approach will inspire scientists in the G4 field to gradually answer more and more remaining questions about the integral map of PQS within (not only) plant genomes.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/life13010199/s1>, Table S1: PQS counts and frequencies in analyzed gene sets; Table S2: PQS counts and frequencies in NCBI features.

**Author Contributions:** Conceptualization, P.P. and M.B.; Data curation, N.B., K.K. and M.B.; Formal analysis, M.B.; Funding acquisition, P.P. and M.B.; Investigation, A.V. and K.K.; Methodology, P.P., N.B. and M.B.; Project administration, M.B.; Resources, K.K. and M.B.; Software, V.B.; Supervision, M.B.; Validation, M.B.; Visualization, A.V. and M.B.; Writing—original draft, A.V. and M.B.; Writing—review & editing, P.P., V.B. and M.B. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the University of Ostrava (SGS10/PřF/2022 to P.P. and K.K., and SGS11/PřF/2022 to A.V). P.P. and M.B. were supported by the National Agency for Agricultural Research (NAZV) of the Czech Republic grant no. QK1810391 “Utilization of genomic and transcriptomic approaches to create genetic resources and breeding materials of poppy with specific traits”.

**Institutional Review Board Statement:** Not applicable.

**Data Availability Statement:** The data presented in this study are available in this article and supplementary materials.

**Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

## References

1. Watson, J.D.; Crick, F.H. Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature* **1953**, *171*, 737–738. [[CrossRef](#)]
2. Poggi, L.; Richard, G.-F. Alternative DNA Structures In Vivo: Molecular Evidence and Remaining Questions. *Microbiol. Mol. Biol. Rev.* **2020**, *85*, e00110-20. [[CrossRef](#)]
3. Ghosh, A.; Bansal, M. A Glossary of DNA Structures from A to Z. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2003**, *59*, 620–626. [[CrossRef](#)]
4. Griffin, B.D.; Bass, H.W. Plant G-Quadruplex (G4) Motifs in DNA and RNA.; Abundant, Intriguing Sequences of Unknown Function. *Plant Sci.* **2018**, *269*, 143–147. [[CrossRef](#)]
5. Kim, J.-M.; Sasaki, T.; Ueda, M.; Sako, K.; Seki, M. Chromatin Changes in Response to Drought, Salinity, Heat, and Cold Stresses in Plants. *Front. Plant Sci.* **2015**, *6*, 114. [[CrossRef](#)]
6. Bartas, M.; Čutová, M.; Brázda, V.; Kaura, P.; Šťastný, J.; Kolomazník, J.; Coufal, J.; Goswami, P.; Červeň, J.; Pečinka, P. The Presence and Localization of G-Quadruplex Forming Sequences in the Domain of Bacteria. *Molecules* **2019**, *24*, 1711. [[CrossRef](#)]

7. Brázda, V.; Luo, Y.; Bartas, M.; Kaura, P.; Porubiaková, O.; Št'astný, J.; Pečinka, P.; Verga, D.; Da Cunha, V.; Takahashi, T.S. G-Quadruplexes in the Archaea Domain. *Biomolecules* **2020**, *10*, 1349. [[CrossRef](#)]
8. Bedrat, A.; Lacroix, L.; Mergny, J.-L. Re-Evaluation of G-Quadruplex Propensity with G4Hunter. *Nucleic Acids Res.* **2016**, *44*, 1746–1759. [[CrossRef](#)]
9. Gellert, M.; Lipsett, M.N.; Davies, D.R. Helix Formation by Guanylic Acid. *Proc. Natl. Acad. Sci. USA* **1962**, *48*, 2013. [[CrossRef](#)]
10. Bartas, M.; Brázda, V.; Karlický, V.; Červeň, J.; Pečinka, P. Bioinformatics Analyses and in Vitro Evidence for Five and Six Stacked G-Quadruplex Forming Sequences. *Biochimie* **2018**, *150*, 70–75. [[CrossRef](#)]
11. Li, X.; Sánchez-Ferrer, A.; Bagnani, M.; Adamcik, J.; Azzari, P.; Hao, J.; Song, A.; Liu, H.; Mezzenga, R. Metal Ions Confinement Defines the Architecture of G-Quartet, G-Quadruplex Fibrils and Their Assembly into Nematic Tactoids. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 9832–9839. [[CrossRef](#)]
12. Havlová, K.; Fajkus, J. G4 Structures in Control of Replication and Transcription of RRNA Genes. *Front. Plant Sci.* **2020**, *11*, 593692. [[CrossRef](#)]
13. Kim, N. The Interplay between G-Quadruplex and Transcription. *Curr. Med. Chem.* **2019**, *26*, 2898–2917. [[CrossRef](#)]
14. Cho, H.; Cho, H.S.; Nam, H.; Jo, H.; Yoon, J.; Park, C.; Dang, T.V.T.; Kim, E.; Jeong, J.; Park, S.; et al. Translational Control of Phloem Development by RNA G-Quadruplex–JULGI Determines Plant Sink Strength. *Nat. Plants* **2018**, *4*, 376–390. [[CrossRef](#)]
15. Bugaut, A.; Balasubramanian, S. 5'-UTR RNA G-Quadruplexes: Translation Regulation and Targeting. *Nucleic Acids Res.* **2012**, *40*, 4727–4741. [[CrossRef](#)]
16. Feng, Y.; Tao, S.; Zhang, P.; Sperti, F.R.; Liu, G.; Cheng, X.; Zhang, T.; Yu, H.; Wang, X.; Chen, C. Epigenomic Features of DNA G-Quadruplexes and Their Roles in Regulating Rice Gene Transcription. *Plant Physiol.* **2022**, *188*, 1632–1648. [[CrossRef](#)]
17. Lejault, P.; Mitteau, J.; Sperti, F.R.; Monchaud, D. How to Untie G-Quadruplex Knots and Why? *Cell Chem. Biol.* **2021**, *28*, 436–455. [[CrossRef](#)]
18. Król, A.; Amarowicz, R.; Weidner, S. Changes in the Composition of Phenolic Compounds and Antioxidant Properties of Grapevine Roots and Leaves (*Vitis Vinifera* L.) under Continuous of Long-Term Drought Stress. *Acta Physiol. Plant* **2014**, *36*, 1491–1499. [[CrossRef](#)]
19. Varela, M.C.; Arslan, I.; Reginato, M.A.; Cenzano, A.M.; Luna, M.V. Phenolic Compounds as Indicators of Drought Resistance in Shrubs from Patagonian Shrublands (Argentina). *Plant Physiol. Biochem.* **2016**, *104*, 81–91. [[CrossRef](#)]
20. Sarker, U.; Oba, S. Drought Stress Enhances Nutritional and Bioactive Compounds, Phenolic Acids and Antioxidant Capacity of Amaranthus Leafy Vegetable. *BMC Plant Biol.* **2018**, *18*, 258. [[CrossRef](#)]
21. Alter, S.; Bader, K.C.; Spannagl, M.; Wang, Y.; Bauer, E.; Schön, C.-C.; Mayer, K.F.X. DroughtDB: An Expert-Curated Compilation of Plant Drought Stress Genes and Their Homologs in Nine Species. *Database* **2015**, *2015*, bav046. [[CrossRef](#)]
22. Brázda, V.; Kolomazník, J.; Lýsek, J.; Bartas, M.; Fojta, M.; Št'astný, J.; Mergny, J.-L. G4Hunter Web Application: A Web Server for G-Quadruplex Prediction. *Bioinformatics* **2019**, *35*, 3493–3495. [[CrossRef](#)]
23. Babicki, S.; Arndt, D.; Marcu, A.; Liang, Y.; Grant, J.R.; Maciejewski, A.; Wishart, D.S. Heatmapper: Web-Enabled Heat Mapping for All. *Nucleic Acids Res.* **2016**, *44*, W147–W153. [[CrossRef](#)]
24. Pasha, A.; Subramaniam, S.; Cleary, A.; Chen, X.; Berardini, T.; Farmer, A.; Town, C.; Provart, N. Araport Lives: An Updated Framework for Arabidopsis Bioinformatics. *Plant Cell* **2020**, *32*, 2683–2686. [[CrossRef](#)]
25. Dobrovolná, M.; Bohálová, N.; Peška, V.; Wang, J.; Luo, Y.; Bartas, M.; Volná, A.; Mergny, J.-L.; Brázda, V. 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. [[CrossRef](#)]
26. 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. [[CrossRef](#)]
27. Vannutelli, A.; Perreault, J.-P.; Ouangraoua, A. G-Quadruplex Occurrence and Conservation: More than Just a Question of Guanine–Cytosine Content. *NAR Genom. Bioinform.* **2022**, *4*, lqac010. [[CrossRef](#)]
28. Chang, N.; Sun, Q.; Hu, J.; An, C.; Gao, H. Large Introns of 5 to 10 Kilo Base Pairs Can Be Spliced out in Arabidopsis. *Genes* **2017**, *8*, 200. [[CrossRef](#)]
29. Bryan, T.M. G-Quadruplexes at Telomeres: Friend or Foe? *Molecules* **2020**, *25*, 3686. [[CrossRef](#)]
30. Yadav, V.; Kim, N.; Tuteja, N.; Yadav, P. G Quadruplex in Plants: A Ubiquitous Regulatory Element and Its Biological Relevance. *Front. Plant Sci.* **2017**, *8*, 1163. [[CrossRef](#)]
31. Kwok, C.K.; Ding, Y.; Shahid, S.; Assmann, S.M.; Bevilacqua, P.C. A Stable RNA G-Quadruplex within the 5'-UTR of Arabidopsis Thaliana ATR mRNA Inhibits Translation. *Biochem. J.* **2015**, *467*, 91–102. [[CrossRef](#)]
32. Liu, H.; Chu, Z.; Yang, X. A Key Molecular Regulator, RNA G-Quadruplex and Its Function in Plants. *Front Plant Sci.* **2022**, *13*, 926953. [[CrossRef](#)]
33. Volná, A.; Bartas, M.; Karlický, V.; Nezval, J.; Kundrátová, K.; Pečinka, P.; Špunda, V.; Červeň, J. G-Quadruplex in Gene Encoding Large Subunit of Plant RNA Polymerase II: A Billion-Year-Old Story. *Int. J. Mol. Sci.* **2021**, *22*, 7381. [[CrossRef](#)]

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.