



Article

The Characterization of G-Quadruplexes in Tobacco Genome and Their Function under Abiotic Stress

Kangkang Song^{1,†}, Bin Li^{1,†} , Haozhen Li¹, Rui Zhang², Xiaohua Zhang¹, Ruiwei Luan^{1,3}, Ying Liu¹ and Long Yang^{1,*}

¹ College of Plant Protection and Agricultural Big-Data Research Center, Shandong Agricultural University, Tai'an 271018, China; kangkangsong234@163.com (K.S.); libin2994429664@163.com (B.L.)

² State Key Laboratory of Plant Physiology and Biochemistry, College of Life Sciences, Zhejiang University, Hangzhou 310058, China

³ College of Agronomy, Shandong Agricultural University, Tai'an 271018, China

* Correspondence: lyang@sdau.edu.cn

† These authors contributed equally to this work.

Abstract: Tobacco is an ideal model plant in scientific research. G-quadruplex is a guanine-rich DNA structure, which regulates transcription and translation. In this study, the prevalence and potential function of G-quadruplexes in tobacco were systematically analyzed. In tobacco genomes, there were 2,924,271,002 G-quadruplexes in the nuclear genome, 430,597 in the mitochondrial genome, and 155,943 in the chloroplast genome. The density of the G-quadruplex in the organelle genome was higher than that in the nuclear genome. G-quadruplexes were abundant in the transcription regulatory region of the genome, and a difference in G-quadruplex density in two DNA strands was also observed. The promoter of 60.4% genes contained at least one G-quadruplex. Compared with up-regulated differentially expressed genes (DEGs), the G-quadruplex density in down-regulated DEGs was generally higher under drought stress and salt stress. The G-quadruplex formed by simple sequence repeat (SSR) and its flanking sequence in the promoter region of the *NtBBX* (*Nitab4.5_0002943g0010*) gene might enhance the drought tolerance of tobacco. This study lays a solid foundation for further research on G-quadruplex function in tobacco and other plants.

Keywords: tobacco; G-quadruplex; feature region; simple sequence repeat (SSR); abiotic stress



Citation: Song, K.; Li, B.; Li, H.; Zhang, R.; Zhang, X.; Luan, R.; Liu, Y.; Yang, L. 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>

Academic Editor: Ludmila Khrustaleva

Received: 29 February 2024

Revised: 9 April 2024

Accepted: 11 April 2024

Published: 14 April 2024



Copyright: © 2024 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

A G-quadruplex is a nucleic acid secondary structure formed by folding nucleotide sequences rich in guanine bases in DNA and RNA [1]. Four guanines are connected to each other by Hoogsteen hydrogen bonds to form a G-quartet, and two or more G-quartets are stacked to shape a G-quadruplex structure (Figure 1A) [2]. Monovalent cations, such as K⁺ and Na⁺, are often bound to the central position of the G-quadruplex to enhance its stability [3]. According to the number of DNA strands forming a G-quadruplex structure, G-quadruplexes are divided into unimolecular, bimolecular, and tetramolecular G-quadruplexes [4]. In addition, the different number of continuous guanines in conservative motifs will lead to different G-quartet layers, so G-quadruplex can be divided into different types, of which G2 (with two G-quartets) and G3 (with three G-quartets) G-quadruplexes are the most common [5]. The most emblematical G-quadruplex sequence pattern is G_{3+N₁₋₇}G_{3+N₁₋₇}G_{3+N₁₋₇}G₃₊. A large number of studies have shown that the G-quadruplex plays an important regulatory role in DNA replication, transcription, translation, and telomere structure maintenance (Figure 1B) [1,6–8].

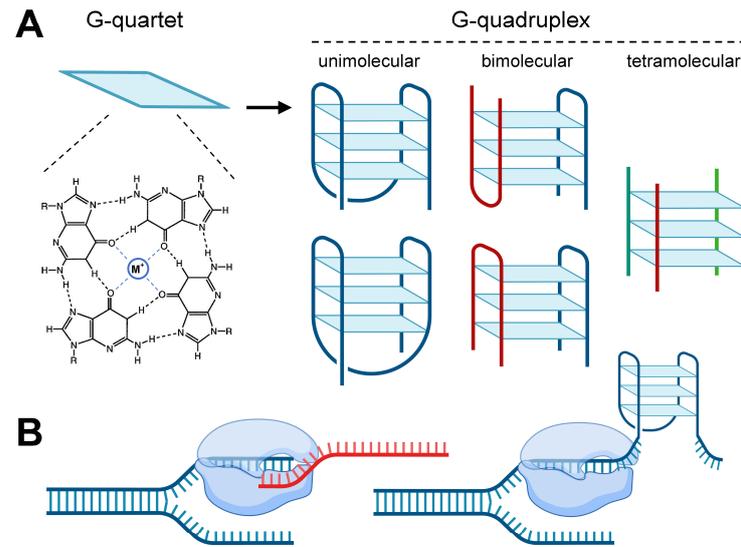


Figure 1. The structure and function of G-quadruplex. (A) G-quadruplex structure. Two or more G-quartets are stacked together to form the G-quadruplex structure, including unimolecular, bimolecular, and tetramolecular G-quadruplexes. (B) G-quadruplexes located on template strand generally inhibit transcription. The diagram was created with BioRender (<https://www.biorender.com/> (accessed on 8 April 2024)).

G-quadruplexes are widely prevalent in various plant genomes. In *Arabidopsis thaliana*, grape, rice, and *Populus tomentosa*, G-quadruplexes were usually located near transcription units or genes [9]. In *Arabidopsis thaliana* and rice, the content of G-quadruplexes at the 5'UTR/CDS junction was very rich [7]. Circular dichroism confirmed that a G-quadruplex structure can be formed in long terminal repeats in the maize LTR retrotransposon [10]. *Chlamydomonas reinhardtii* has a GC-rich genome (67%) and a high density of potential G-quadruplex formation sequences, which are located in the promoter region of DNA repair and photosynthetic genes [11]. A total of 23,685 G4 peaks were identified by chromatin immunoprecipitation of the BG4 antibody used to visualize the G-quadruplex structure and high-throughput sequencing in rice genome [12]. In *Spirodela polyrhiza*, there were strong G4 peaks in the promoters of the cytosolic nitrate reductase gene (*SpNR*) and nitrite reductase gene (*SpNiR*), which were located between -265 and -290 bp relative to the translation initiation site [13]. A stable G-quadruplex was revealed in the *RPB1* gene encoding the RNA polymerase II subunit by bioinformatics and circular dichroism [14]. G-quadruplexes were enriched in 5'UTR and 3'UTR of pea, suggesting the role of the G-quadruplex in post-transcriptional regulation [15]. In the barley genome, the G-quadruplex motif reached a peak near the 5'UTR, the first coding domain sequence and the first intron initiation site on the antisense strand [16]. The 5'UTR of ataxia telangiectasia mutated and rad3-related (ATR) mRNA in *Arabidopsis thaliana* was confirmed to have the G-quadruplex structure determined by biophysical and biochemical methods [17]. Using the nBMST computer program, the G-quadruplex structure was found to be enriched at the centromere of oat, and it may be formed in centromere duplication and the ENH3 nucleosome in vivo [18]. The prevalence and distribution of the G-quadruplex in the plant genome might vary with specific G-quadruplex types, and the G-quadruplex of G3 type was more abundant in intergenic regions, while the G-quadruplex of G2 type was established in gene regions [5].

The G-quadruplex plays a significant function in regulating the growth and development of plants. Notably, G-quadruplexes can influence gene expression via various mechanisms, both enhancing or suppressing it. In various plants, the conservative pattern of high-density G-quadruplexes at the promoter and 5'UTR positions indicated that G-quadruplexes have a vital role in regulating gene expression [19]. The G-quadruplex enriched in an intron may affect the transcription process [20]. The G-quadruplex in the 3'UTR region of *Arabidopsis* strongly enhanced the stability of mRNA and inhibited its

degradation [21]. In *Sapium sebiferum*, the G-quadruplex at the L-ascorbate peroxidase gene may fulfill a crucial function in the flowering process [22]. G-quadruplexes were also common in genes related to energy steady-state signals and in many genes related to target of rapamycin (TOR), adenosine monophosphate (AMP) kinase, and the oxidative stress signal pathway, indicating that the G-quadruplex had an important action in energy state regulation, signal transduction, and metabolic regulation [23]. The G-quadruplex in long terminal repeats in the maize LTR retrotransposon inhibits the expression of the reporter gene in yeast [8]. The formation of a plant RNA G-quadruplex inhibited the translation of *SMXL4/5* and restricted phloem differentiation [24]. Genetic and biochemical analysis showed that RNA G-quadruplex folding can regulate translation and plant growth [25]. In *Spirodela polyrhiza*, the complex regulation of the nitrogen assimilation gene involved the synergistic effect of multiple NRElike and GAATC/GATTC cis-elements and TATA-based enhancers, (GA/CT)_n repeats, and the G-quadruplex structure of promoters [13]. GO enrichment analysis has shown that an orthologous gene with a G-quadruplex in many dicotyledonous plant species is involved in important biological pathways, such as chromatin modification, the regulation of phosphorylation and intracellular signal transduction, auxin transport, seed development, and GTPase activity [16]. In monocotyledonous plant species, an orthologous gene with a G-quadruplex participates in biological processes, such as development, ion transport, transcriptional regulation, and protein folding [16].

G-quadruplexes are also involved in the process of plant response to abiotic stress. A G-quadruplex exists in the nuclease supersensitive site in the promoter of the rice thermal response gene. At simulated physiological temperature and potassium concentration, representative G-quadruplexes can form stable G-quadruplex structures, which can block DNA polymerase. However, with the increase in temperature, some G-quadruplexes disappear, which implies that these G-quadruplexes can sense temperature changes through structural transformation [26]. Plants growing in a low-temperature climate contained more guanine and G-quadruplexes in their transcriptome. Cold conditions were likely to strongly promote the folding of the RNA G-quadruplex with a higher number of G-quartets and medium loop length. GO analysis of the genes with a higher RNA G-quadruplex folding fraction after low-temperature treatment showed the enrichment of specific transcription in biological functions, such as the response to abiotic stimuli, response to temperature stimuli, and response to cold [25]. There were also many G-quadruplexes in maize hypoxia response genes, and the expression pattern of maize hypoxia response genes carrying G-quadruplexes could be changed under the supply of sugar [19,27]. G-quadruplexes were also found to be enriched in differentially expressed genes of *Arabidopsis thaliana* under drought stress [28].

Tobacco is widely used in scientific research and is also an important cash crop. The G-quadruplex is an important DNA structure, which is widely involved in key life processes. However, unlike humans and animals, the research on the G-quadruplex in plants is very limited. With the release of high-quality genomes and the development of bioinformatics, the G-quadruplex has been systematically studied in *Arabidopsis thaliana* [9,28], rice [12,29], wheat [30], barley [16], and pea [15]. In this study, the G-quadruplex in tobacco was systematically analyzed by the bioinformatics method, using tobacco genome data and transcriptome data under abiotic stress. The G-quadruplex of feature sequences, the relationship between simple sequence repeats (SSRs) and the G-quadruplex, the G-quadruplex of differentially expressed genes (DEGs), and the G-quadruplex of the transcription factor gene family were carefully investigated. This study promotes the understanding of the prevalence and function of the G-quadruplex in tobacco and other plants.

2. Results

2.1. General Situation of G-Quadruplex Distribution in Tobacco Genome

The length of tobacco chromosomes ranged from 82,751,733 bp (Nt21) to 215,930,317 bp (Nt17). The GC content was between 38.6% (Nt19) and 39.5% (Nt01). The differences in the number of G-quadruplexes in chromosomes were large, with the smallest being 42,881

(Nt11) and the largest being 116,804 (Nt17) (Table 1). The density of G-quadruplexes in all nuclear chromosomes was basically the same, with an average density of 0.5/kbp. The density of G-quadruplexes in the mitochondrial chromosome was 1.6/kbp, and the density of G-quadruplexes in the chloroplast chromosome was 0.9/kbp.

Table 1. G-quadruplex profile of tobacco genome. The G-quadruplexes were identified by G4Hunter, with the window set to 25 and the threshold set to 1.2.

Chromosome	Length (bp)	GC Content (%)	G-Quadruplex Number	G-Quadruplex Density (per kbp)
Nt01	135,559,120	39.5	69,547	0.5
Nt02	109,624,155	38.7	60,928	0.6
Nt03	97,104,660	39.1	43,104	0.4
Nt04	136,037,944	38.8	75,885	0.6
Nt05	109,337,480	39.3	53,838	0.5
Nt06	136,518,381	39.3	70,455	0.5
Nt07	105,049,242	39.2	51,317	0.5
Nt08	108,393,918	39.4	52,757	0.5
Nt09	106,147,314	38.8	57,485	0.5
Nt10	116,194,611	39.1	55,884	0.5
Nt11	84,914,900	39.2	42,881	0.5
Nt12	127,111,110	38.8	71,171	0.6
Nt13	139,740,185	38.8	82,542	0.6
Nt14	115,579,444	38.8	69,329	0.6
Nt15	115,424,069	39	66,956	0.6
Nt16	99,759,613	39.2	45,931	0.5
Nt17	215,930,317	38.8	116,804	0.5
Nt18	113,077,399	39	51,195	0.5
Nt19	155,028,365	38.6	81,044	0.5
Nt20	105,109,812	39.2	52,074	0.5
Nt21	82,751,733	39	47,122	0.6
Nt22	163,185,734	39	85,914	0.5
Nt23	128,150,892	38.7	71,924	0.6
Nt24	118,540,604	38.7	66,729	0.6
mitochondrion	430,597	45	705	1.6
chloroplast	155,943	37.8	148	0.9

The potential G-quadruplex forming sequences in the nuclear genome amount to 2,924,271,002. The most abundant G-quadruplex sequence, “GGGGTGTGTACAGACTCCGAGGGG”, occurred 1302 times in the genome. The other four most abundant G-quadruplex sequences occurred approximately 600 times in the genome. These sequences had a roughly equal probability of occurrence on both the positive and negative strands (Table 2).

Table 2. The five most frequent G-quadruplex motif families in the tobacco nuclear genome.

Sequence	Number	Positive Strand	Negative Strand	Length (bp)	ABS_Score
GGGGTGTGTACAGACTCCGAGGGG	1302	635	667	26	1.423077
GGGGCCTCGGGTGTGTTCCGGATG	605	294	311	25	1.200000
GGGGTGTGTACAGACTCCGAGGGG	587	306	281	25	1.320000
CGGGGGTGTGACTTTTGTATCCGGGGT	599	304	295	28	1.357143
CTGGGGTGTACAGACTCCGAGGGGCT	575	283	292	28	1.214286

2.2. The Relationship between Tobacco G-Quadruplexes and Genome Characteristics

The prevalence of tobacco G-quadruplexes was potentially influenced by GC content, gene density, and SSR density (Figure 2). In some genome regions, such as 30–40 Mb of the Nt03 chromosome and 40–50 Mb of the Nt20 chromosome, GC density, gene density,

SSR density, and G-quadruplex density collectively exhibited higher levels compared to neighboring DNA regions. The effect of SSRs on G-quadruplex formation was investigated emphatically. A total of 109,865 SSRs were identified in the tobacco genome, among which p2 SSRs were largest, accounting for 51.95% of all SSRs, followed by p3 SSRs, accounting for 33.68% of all SSRs (Figure S1). Among p3 SSRs, the number of SSRs with repeating units of CCN, NGG, NCC, GGN, CNC, and GNG were 275, 400, 689, 340, 272, and 471 respectively, and these 2447 SSRs account for 6.61% of p3 SSRs and 2.23% of all SSRs. There are 34,553 p3 SSRs with AAN, NTT, NAA, TTN, TNT, ANA, and other types, accounting for 93.39% of p3 SSRs and 31.45% of all SSRs (Figure S1). A total of 5906 SSRs could form 1–15 G-quadruplexes, including 1692 p3 SSRs. A 132 bp SSR on the Nt07 chromosome, (AAG)5attttgg(ATA)5(aga)6gggtggata(agg)8at(gag)8a(agg)5, had the potential to form 15 G-quadruplexes (Table S1). SSRs in the tobacco genome could form a total of 7679 G-quadruplexes. These results implied that SSRs may indirectly affect the distribution of G-quadruplexes by impacting GC density and gene density (Figure 2).

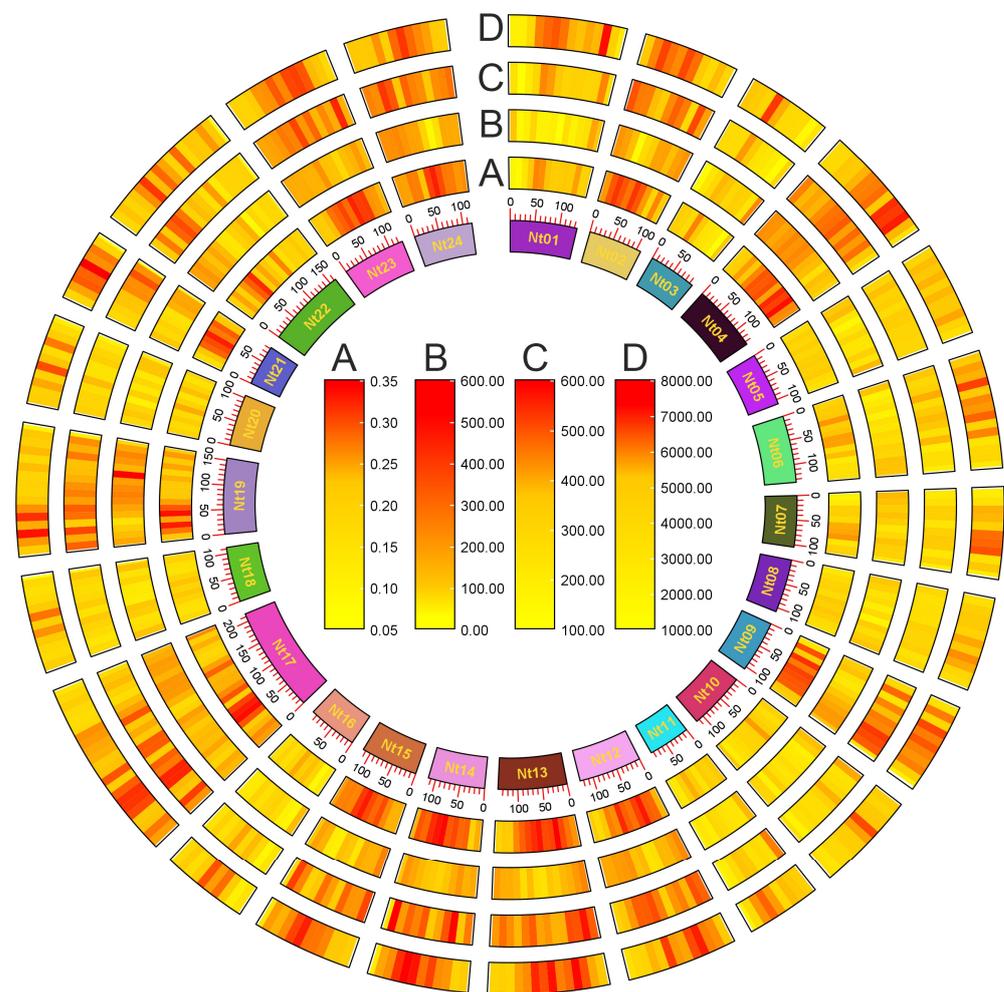


Figure 2. G-quadruplex landscape in tobacco genome. (A) GC content. (B) Gene density. (C) SSR density. (D) G-quadruplex density. The innermost circle represents the 24 chromosomes of tobacco. Chromosomes are divided into bins of 10 Mb each.

2.3. G-Quadruplex of Feature Regions in Tobacco Genome

The number of G-quadruplexes varied greatly across different genomic feature regions (Table S2). The number of G-quadruplexes in intergenic regions reached as high as 2,954,439, while only 484 were present in the 3'UTR region. Among the various regions of genes, introns harbored the highest number of G-quadruplexes, while the exons, CDS, and 5'UTR contained fewer, with the 3'UTR hosting the least. A large number of G-quadruplexes were

also found in promoters and TSS500 regions. Additionally, the number of G-quadruplexes varied between different DNA strands within the same genomic feature region. In certain regions such as the 5'UTR, there is a notable difference in the quantity of G-quadruplexes between the template and coding strands.

The density of G-quadruplexes in different genomic feature regions was different. In specific feature regions, the density of G-quadruplexes on template strands and coding strands was also different (Figure 3). On the DNA double strands, the G-quadruplex density in genomic feature regions was as follows, from highest to lowest: 5'UTR, TSS500, promoter500, promoter1000, CDS, exon, promoter1500, promoter2000, intron, gene, intergenic, and 3'UTR. In three sequence contexts (double strand, template strand, and coding strand), the G-quadruplex density was highest in the 5'UTR region and lowest in the 3'UTR region. For the promoter region, the G-quadruplex density follows the same pattern on both the double strand and coding strand, whereby shorter promoter sequences correspond to higher G-quadruplex density. Conversely, the G-quadruplex density in promoter sequences on the template strand exhibited an opposite trend, with shorter promoter sequences exhibiting lower G-quadruplex density.

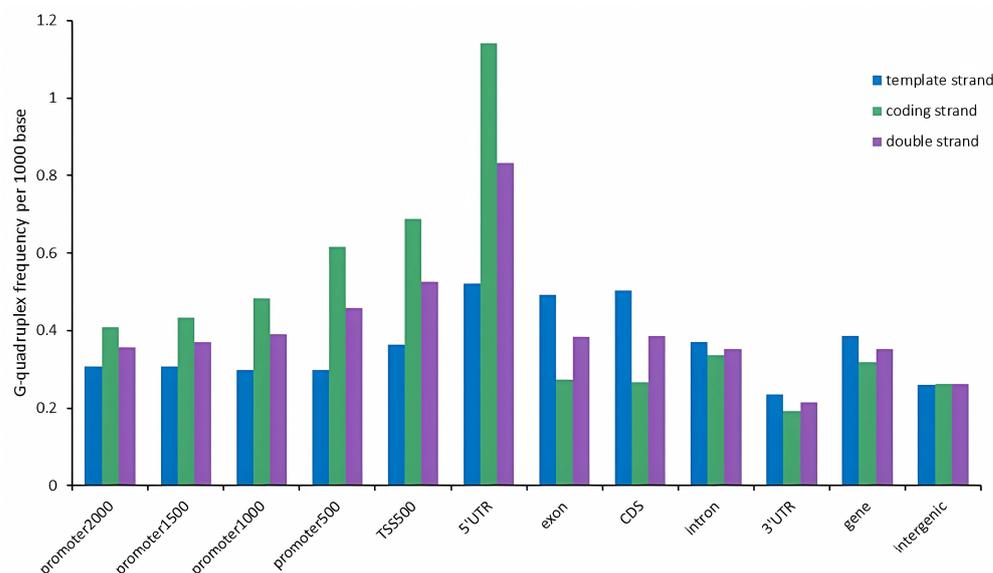


Figure 3. The density of G-quadruplexes in different feature regions of tobacco genome. The promoter2000, promoter1500, promoter1000, and promoter500 represent the sequences upstream of the gene by 2000 bp, 1500 bp, 1000 bp, and 500 bp, respectively. TSS500 represents the sequence consisting of 250 bp upstream and downstream of the transcription start site.

2.4. Functions of Genes with Highly Enriched G-Quadruplexes in Promoters

G-quadruplexes were enriched in different numbers in promoters of genomic genes, and there were also differences among the three sequence contexts (double strand, template strand, and coding strand) (Figure 4). In the template strand, G-quadruplexes were absent in the promoter regions of 23,401 genes, accounting for 65.88% of all genes in the genome. In the promoter regions of 11,806 genes, there were 1–5 G-quadruplexes, accounting for 33.24% of all genes. The promoter regions of 312 genes were capable of forming six or more G-quadruplexes, representing 0.88% of all genes.

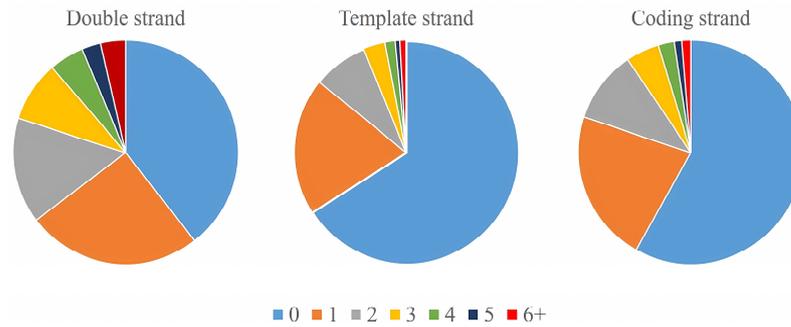


Figure 4. The proportion of genes with different G-quadruplex numbers in the promoter. The colors indicate genes containing various G-quadruplex numbers in the promoter. The value of 0 represents genes whose promoter missed G-quadruplexes. The values of 1, 2, 3, 4, and 5 represent genes whose promoter contains one, two, three, four, and five G-quadruplexes, respectively. The value of 6+ represents the genes whose promoter contains six or more G-quadruplexes.

The genes with more than 10 G-quadruplexes in the promoter template strand were selected for GO function enrichment analysis. The enrichment of BP (biological process), CC (cellular component), and MF (molecular function) ontologies was calculated and 21 enriched terms were found (Figure 5). The biological process category included RNA localization, protein import, nucleocytoplasmic transport, nuclear transport, the establishment of protein localization to organelles, RNA transport, RNA export from nuclei, protein localization to nucleus, protein import into nucleus, nucleobase-containing compound transport, nucleic acid transport, nuclear export, import into nucleus, the establishment of RNA localization, the cellular response to alcohol, and the cellular response to abscisic acid stimuli. The cellular component category included nuclear pores, nuclear envelopes, and mitochondrial inner membranes. The molecular function category included structural constituents of nuclear pores and mRNA 3'UTR binding. These aspects involve the important process of plant growth and development.

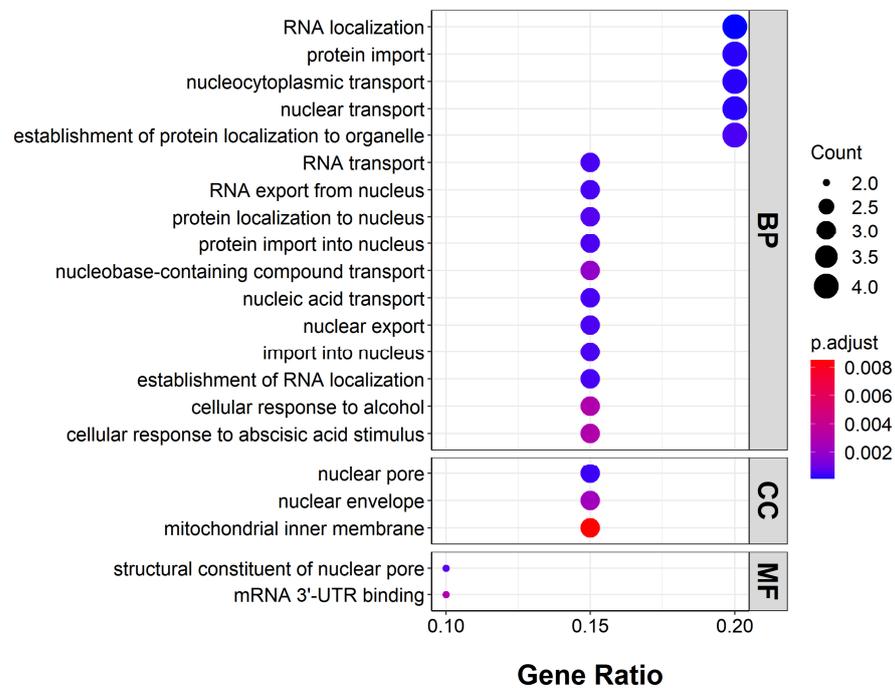


Figure 5. GO enrichment of genes rich in promoter G-quadruplex (>10) in template strand.

2.5. G-Quadruplex of DEGs under Abiotic Stress

Under drought stress, 7477 genes were differentially expressed (2718 up-regulated DEGs and 4759 down-regulated DEGs). Under salt stress, 2764 genes were differentially expressed (1198 up-regulated DEGs and 1566 down-regulated DEGs) (Table S3). Comparing DEGs and non-differentially expressed genes (nDEGs) under drought stress and salt stress, the number and density of G-quadruplexes in different feature regions showed distinct characteristics (Table S4, Figure 6). For the G-quadruplex density of the gene, intron, and 5'UTR regions, DEGs were lower than nDEGs. This pattern was especially obvious in the 5'UTR region under NaCl stress. For the G-quadruplex density of the CDS, 3'UTR, promoter2000, promoter1500, promoter1000, and TSS500, DEGs were higher than nDEGs under drought stress, but it exhibited the opposite trend under NaCl stress. Notably, for the G-quadruplex density of each feature region, DEGs of drought stress were higher than DEGs of salt stress.

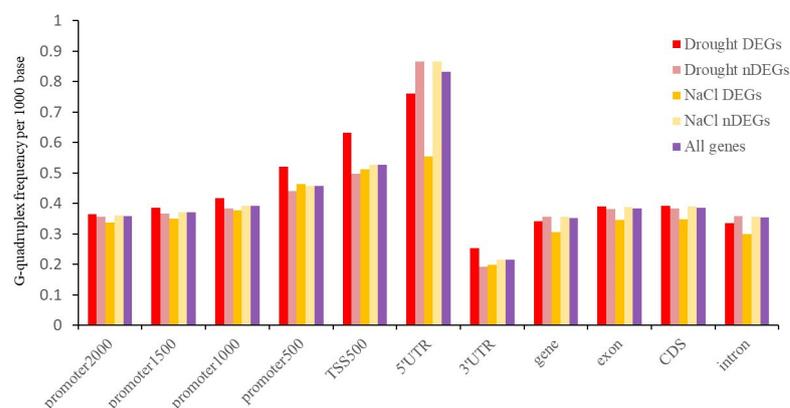


Figure 6. The density of G-quadruplex of DEGs and nDEGs under drought stress and salt stress.

G-quadruplexes of up-regulated DEGs and down-regulated DEGs were investigated in further detail. For up-regulated DEGs and down-regulated DEGs under drought stress and salt stress, the number and density of G-quadruplexes in different feature regions were different, and the template strand and coding strand in specific feature regions were also different (Tables S5 and S6 and Figure 7). Except for the template strand in the 3'UTR region, the density of G-quadruplexes of DEGs under drought stress was higher than that under salt stress in all other cases (Figure 7).

Whether up-regulated DEGs or down-regulated DEGs under drought stress, for the G-quadruplex density of the 5'UTR, TSS500, and promoter regions, the coding strand was higher than the template strand. In up-regulated DEGs under drought stress, for the G-quadruplex density of the 5'UTR region, the coding strand was 4.3-fold higher than the template strand. In down-regulated DEGs under drought stress, for the G-quadruplex density of the 5'UTR region, this value was 2.9-fold higher. In addition, whether up-regulated DEGs or down-regulated DEGs under drought stress, for the G-quadruplex density of the 3'UTR, gene, exon, CDS, and intron regions, the coding strand was lower than the template strand. In up-regulated DEGs under drought stress, for the G-quadruplex density of the CDS region, the coding strand in the CDS region was 2.2-fold lower than the template strand. In down-regulated DEGs under drought stress, for the G-quadruplex density of the CDS region, this value was 1.4-fold lower.

This strong model was also revealed under NaCl stress. In up-regulated DEGs under NaCl stress, for the G-quadruplex density of the 5'UTR region, the coding strand was 5.1-fold higher than the template strand. In down-regulated DEGs under NaCl stress, for the G-quadruplex density of the TSS500 region, this value was 3.5-fold higher. In addition, in up-regulated DEGs under NaCl stress, for the G-quadruplex density of the 3'UTR region, the coding strand was 2.4-folds lower than the template strand. In down-regulated DEGs under NaCl stress, for the G-quadruplex density of the 3'UTR region, this value was 2.6-fold lower.

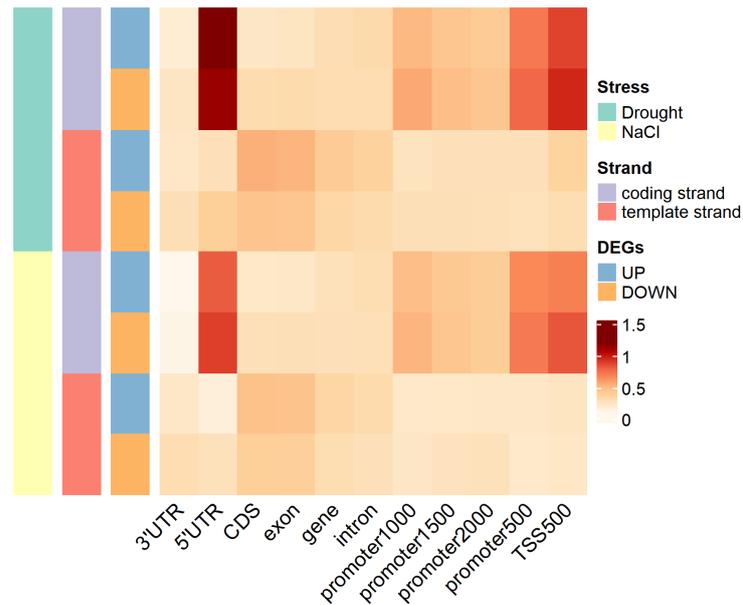


Figure 7. The density of G-quadruplexes in different feature regions of DEGs under drought stress and NaCl stress.

2.6. G-Quadruplexes in Transcription Factor Gene Family

The bZIP genes, NAC genes, BBX genes, and MADS-box genes were differentially expressed under drought stress, and most of these DEGs contained G-quadruplexes in their gene body and transcriptional regulatory regions (Figure 8). The promoter region of some genes comprised multiple G-quadruplexes, such as *Nitab4.5_0000246g0100* (bZIP, 4), *Nitab4.5_0000831g0030* (bZIP, 4), *Nitab4.5_0001811g0030* (NAC, 4), *Nitab4.5_0002943g0010* (BBX, 6), and *Nitab4.5_0000902g0340* (MADS-box, 5).

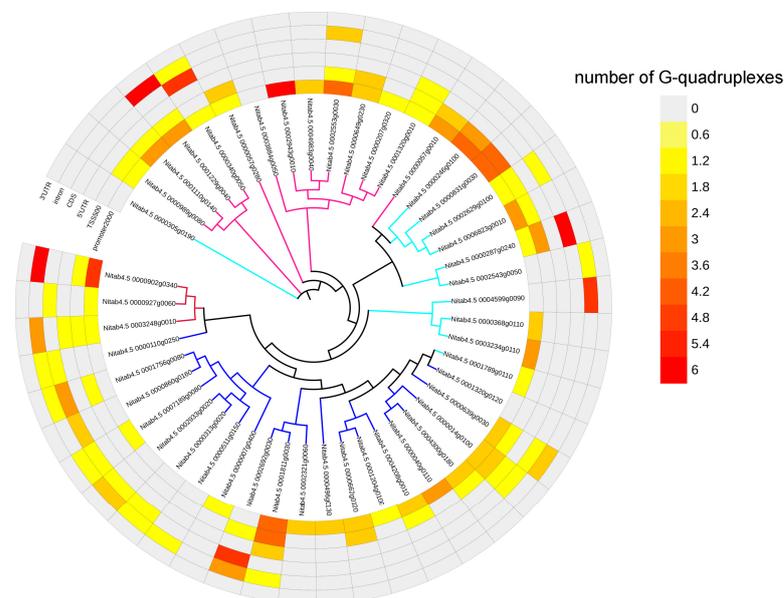


Figure 8. G-quadruplex in four transcription factor gene families responding to drought stress. The gene family types are represented by the color of the branches of the phylogenetic tree, with cyan as bZIP gene family members, blue as NAC gene family members, deep pink as BBX gene family members, and red as MADS-box gene family members. The color of the heat map represents the number of G-quadruplexes. The phylogenetic tree was constructed using maximum likelihood.

SSRs were involved in the formation of a G-quadruplex in the gene expression regulatory region of two transcription factors, namely *Nitab4.5_0002692g0030* and *Nitab4.5_0002943g0010*. *Nitab4.5_0002692g0030* belonged to the *NAC* gene family, which was up-regulated under drought stress. In the second intron region of this gene, an SSR with (GGA)₅ engaged the formation of a G-quadruplex. *Nitab4.5_0002943g0010* belonged to the *BBX* gene family, which was up-regulated under drought stress. In the promoter region of this gene, an SSR participated in the formation of a G-quadruplex, and the repeat unit of this SSR was GGA, with 13 repeats (Figure 9).

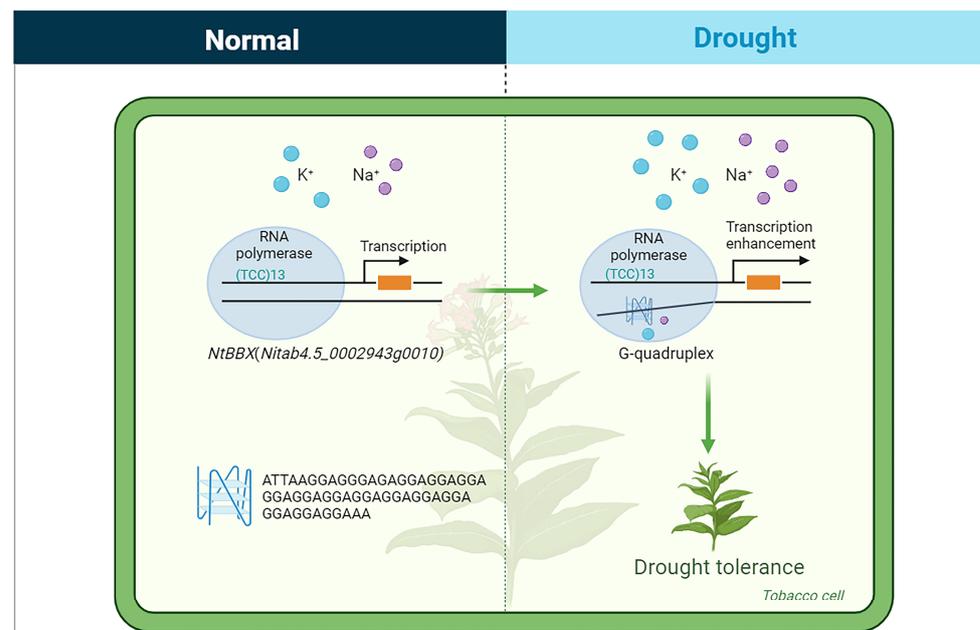


Figure 9. The G-quadruplex formed by SSR and its flanking sequence in the promoter region of the *NtBBX* (*Nitab4.5_0002943g0010*) gene might enhance drought tolerance in tobacco. The drawing was created with BioRender (<https://www.biorender.com/> (accessed on 15 December 2023)).

3. Discussion

3.1. Tobacco Organelle Genome Has Higher G-Quadruplex Density than Nuclear Genome

The density of G-quadruplexes in tobacco nuclear DNA was 0.5/kbp, while those in mitochondrial DNA and chloroplast DNA were 1.6/kbp and 0.9/kbp, respectively. A recent study on G-quadruplexes of pea confirmed the conclusion that the organelle genome had higher G-quadruplex density [15]. It is worth noting that the GC content of the tobacco mitochondrial genome is not significantly higher than that of the nuclear genome, and the GC content of the chloroplast genome is even lower than that of the nuclear genome. Therefore, GC content is not the only factor that causes the difference in G-quadruplex density between the organelle genome and the nuclear genome.

In the study of non-plant species, the mitochondrial genome G-quadruplex has been proved to play a direct role in mitochondrial genome replication, transcription, and respiratory function [31]. However, the function of these special advanced DNA structures in plant mitochondria is still unclear. The tendency of the G-quadruplex to be enriched in the genome of tobacco organelles suggests that the G-quadruplex plays a specific role in some physiological and biochemical processes. Outside the nucleus, there are important genomes in mitochondria and chloroplasts, and these extranuclear genes play a vital role in respiration, photosynthesis, and development [32,33]. At present, the editing schemes of the tobacco mitochondrial genome and chloroplast genome have been put forward [34,35]. In the future, gene editing technology can be used to change the base arrangement in the motif of the G-quadruplex in tobacco organelles to control the formation and stability of

the G-quadruplex, to change the expression level of related genes and ultimately change the yield and quality of tobacco.

3.2. G-Quadruplex Was Enriched in the Coding Strand of Upstream Regulatory Region

The density of the G-quadruplex in the tobacco promoter, TSS500, and 5'UTR regions is higher than that in other genomic feature regions, and the density of the G-quadruplex in the coding strand of these three feature regions is obviously higher than that of the template strand. The promoter, TSS500, and 5'UTR regions all belong to the upstream regulatory regions of the gene, and the enrichment of the G-quadruplex in these regions suggests that they may play an important role in regulating gene expression.

Whether these G-quadruplex structures in the upstream regulatory region of tobacco genes promote or inhibit gene expression needs further exploration. The G-quadruplex has long been considered an obstacle to gene expression. In 2012, the concept of the G-quadruplex as a direct transcription repressor was challenged [36]. Subsequently, more and more studies show that the G-quadruplex does not have a simple inhibitory effect on gene expression regulation, and many studies have found the relationship between the G-quadruplex and high transcription activity [37–39]. In rice, the G-quadruplex in the gene body was negatively correlated with gene expression, while when the G-quadruplex was located in the promoter, it was positively correlated with gene expression [12]. The G-quadruplex in the tobacco genome may also play a dual role in tobacco gene expression. The study on the location of the G-quadruplex in the tobacco genome and the effect of its quantity on gene expression activity will be helpful in the application of the G-quadruplex in improving tobacco characteristics.

3.3. SSR Provided Conditions for G-Quadruplex Formation

The SSR is widely distributed in the tobacco genome, and its high variability and relatively conservative flanking sequences provide conditions for the formation of G-quadruplexes. In tobacco, SSRs forming G-quadruplexes account for 5.38% of all SSRs, and the proportion is not high. This is determined by the formation of G-quadruplexes, which requires a high G content in the sequence and the specific algorithm logic of the G4Hunter program. Among all SSRs in the tobacco genome, p2 SSRs have the largest number, accounting for 51.95% of all SSRs, and the actual proportion is even larger than this value, because the compound SSR also contains dinucleotide repeats, which is similar to wheat [40] and rice [41]. However, all p2 SSR types have no potential to form a G-quadruplex. The number of p3 SSRs is also large, accounting for 33.68% of all SSRs. In p3 SSRs, only SSRs with repeating units of CCN, NGG, NCC, GGN, CNC, and GNG can form G-quadruplexes. However, in p3 SSRs of tobacco, the proportion of AAN, NTT, NAA, TTN, TNT, ANA, and other types of repetitive motifs is high, and p3 SSRs with G-quadruplex formation potential only account for 6.61% of all p3 SSRs. p2 SSRs and p3 SSRs account for a very high proportion in the genome, but few of them have the potential to form G-quadruplexes, which is the main reason for the small number of SSRs involved in the formation of G-quadruplexes.

In tobacco, although the SSR involved in the formation of G-quadruplexes only accounts for a small part of all SSRs in the genome, there are 5906 SSRs capable of forming G-quadruplexes, which is very considerable. The association between a large number of SSRs and G-quadruplexes in tobacco suggests that they may have the function of regulating gene expression. The appearance of these structures increases the instability of genes and may affect the process of gene expression, including DNA replication, repair, and transcription. In specific areas of the tea genome, a correlation between SSR density and the G-quadruplex, GC density, gene density, and CRISPR editing site density was found, and these areas were related to the secondary metabolism of tea [42]. In addition to SSRs, there are other types of tandem repeats and scattered repeats in the plant genome, including satellite DNA and transposons. The relationship between these repeats and the G-quadruplex in the tobacco genome needs further study.

3.4. G-Quadruplex May Be Involved in the Response of Abiotic Stress in Tobacco

Abiotic stress has a significant impact on plant growth and development. The transcription factor is the main regulator of abiotic stress and an excellent candidate gene for crop improvement [43]. The *bZIP*, *NAC*, *BBX*, and *MADS-box* genes, as important transcription factors, have been found to perform crucial functions in tobacco growth and abiotic stress tolerance [44–47]. There are a large number of G-quadruplexes in the feature region of tobacco DEGs under salt stress and drought stress, among which four transcription factor family genes, *bZIP*, *NAC*, *BBX*, and *MADS-box*, are expressed in different degrees under drought stress. During abiotic stress such as drought, the cytoplasmic concentration of K^+ and Na^+ cations increased [48]. It is known that higher levels of K^+ and Na^+ can promote the formation of a G-quadruplex [3], and the ability of the G-quadruplex motif to form a G-quadruplex structure may change under drought stress and salt stress, thus affecting the transcription activity of genes to varying degrees. Therefore, the dynamic changes of G-quadruplex stability in the DEG promoter under drought stress and salt stress may be the potential mechanism for tobacco to cope with these two abiotic stresses. Studies in rice have proved that the G-quadruplex on the promoter can promote gene expression [12]. The G-quadruplex in the upstream regulatory region of the tobacco gene may also promote gene expression. A possible model of the *NtBBX* transcription factor (*Nitab4.5_0002943g0010*) gene in drought tolerance was proposed. The *NtBBX* transcription factor (*Nitab4.5_0002943g0010*) is an up-regulated differentially expressed gene under drought stress, and its promoter coding strand has a p3 SSR, which has the ability to form a G-quadruplex with its flanking sequence. It is known that K^+ and Na^+ can increase the stability of a G-quadruplex, and the increase in intracellular K^+ and Na^+ concentration under drought stress can promote the formation of a G-quadruplex, and then promote the expression of the *NtBBX* transcription factor to enhance the drought stress tolerance of tobacco (Figure 9). It should be emphasized that this hypothesis only preliminarily revealed that the G-quadruplex induced by SSR regulated the expression of the transcription factor. This potential relationship provides an important direction for future research, and the applicability of this hypothesis to all genomic genes needs to be explored from multiple perspectives in future work. Under certain conditions such as abiotic stress, whether the DNA G-quadruplexes in root, shoot, and flower parts are different is also an interesting and important direction for future research. For the key G-quadruplex that regulates plant development and environmental adaptation, the nucleotide that constitutes the G-quadruplex can be mutated by gene editing technology to control the formation and stability of the G-quadruplex, to enhance the resistance of cash crops such as tobacco to abiotic stress.

4. Materials and Methods

4.1. In Silico Identification and Characterization of G-Quadruplexes in Tobacco Genome

The tobacco nuclear genome was obtained from the Sol Genomics Network (https://solgenomics.net/ftp/genomes/Nicotiana_tabacum/edwards_et_al_2017/ (accessed on 18 August 2022)) [49]. The organelle genomes (mitochondrial genome, BA000042.1; chloroplast genome, Z00044.2) were acquired from the National Center for Biotechnology Information [50]. The nucleotide sequence of each chromosome was submitted to G4Hunter to identify the potential G-quadruplexes, with the window set to 25 and the threshold set to 1.2 [51]. The genome coordinates of each G-quadruplex were obtained. The G4Hunter data were further cleaned. The number and density of G-quadruplexes of each chromosome and detailed information of all G-quadruplexes, as well as the sequence length and GC content, were summarized. The SSRs in the tobacco genome were identified by MISA v2.1, which was set as default parameters (1-20 2-6 3-5 4-5 5-5 6-5, interruptions: 100, GFF: true) [52]. The GC density, gene density, SSR density, and G-quadruplex density of each chromosome were calculated and visualized using the Advanced Circos function module of Tbtools v1.108 [53].

According to the genome annotation file, the genome coordinates of feature regions, including the exon, CDS, intron, 5'UTR, 3'UTR, gene, intergenic, promoter2000, promoter1500,

promoter1000, promoter500, and TSS500, were extracted and calculated. The promoter2000, promoter1500, promoter1000, and promoter500 were defined as the sequences upstream of the gene by 2000 bp, 1500 bp, 1000 bp, and 500 bp, respectively. TSS500 was defined as the sequence consisting of 250 bp upstream and downstream of the transcription start site. Based on the genome coordinates of the G-quadruplex and the genome coordinates of the above feature regions, the number of G-quadruplexes in each feature region was calculated, and the further density of G-quadruplexes in each feature region was calculated. These computations were all conducted using a customized R script.

4.2. Number Level of G-Quadruplexes in Promoter and Functional Enrichment Analysis

All genes of the tobacco genome were divided into seven classes based on the number of G-quadruplexes in the promoter region. The first class was the genes whose promoter missed G-quadruplexes. The second to sixth classes were genes with one, two, three, four, and five G-quadruplexes in their promoters, respectively. The seventh class was the genes whose promoter contained six or more G-quadruplexes.

Genes with more than 10 G-quadruplexes in the promoter template strand were selected as a set for Gene ontology (GO) enrichment analysis to explore which biological functions these G-quadruplex-enriched genes were related to. The functional annotation of these genes was performed through eggNOG-mapper (<http://eggno-mapper.embl.de/> (accessed on 19 October 2022)) by submitting the corresponding protein sequences [54]. The annotation result was sorted using the eggNOG-mapper Helper module of TBtools v1.108. GO enrichment analysis and visualization were performed through the clusterProfiler 4.0 R package [55].

4.3. G-Quadruplex Analysis of DEGs under Abiotic Stresses

RNA-seq data were obtained from the Sequence Read Archive (SRA) database (<https://www.ncbi.nlm.nih.gov/sra/> (accessed on 21 December 2022)) [56], including drought stress (SRP301492) and NaCl stress (SRP193166). Trimmomatic v0.39 [57] was employed to remove adapters and cut off the first 12 bases of reads. The genome index was established and reads were mapped using Hisat2 v2.2.1 [58]. The sam files were converted to bam files by Samtools v1.7 [59]. The FPKM values were calculated by Stringtie v2.1.7 [60]. The counts values were obtained by the prepDE.py3 program provided by Stringtie. The DEGs were analyzed by using DESeq2 [61], where the screening standard was $|\log_2\text{FoldChange}| \geq 1$ and $\text{padj} \leq 0.05$. The gene IDs of DEGs, nDEGs, up-regulated DEGs, and down-regulated DEGs were gained. Then, the genome coordinates of all feature regions of these genes were obtained from the results in Section 4.1. Based on the genome coordinates of the G-quadruplex and genome coordinates of these genes, the number of G-quadruplexes in these genes was calculated, and the further density of G-quadruplexes in each feature region was calculated.

4.4. G-Quadruplex Analysis of Transcription Factor Genes Responding to Drought Stress

Annotated *bZIP*, *NAC*, *BBX*, and *MADS-box* family genes were extracted, and these genes were intersected with DEGs under drought stress to obtain the DEGs of these four transcription factor family genes under drought stress using the intersect R function. Based on the genome coordinates of the G-quadruplex and genome coordinates of these transcription factor genes, the number of G-quadruplexes in each transcription factor gene was calculated. The ML phylogenetic tree was constructed using the family protein sequences through the One Step Build a ML Tree function module of TBtools v1.108. The number of G-quadruplexes in differentially expressed family genes was displayed by iTOL (<https://itol.embl.de/> (accessed on 24 March 2023)) [62].

5. Conclusions

In this study, a large number of G-quadruplexes were revealed to exist in the tobacco genome, and the G quadruplex density in the organelle genome was greater than that in the

nuclear genome. G-quadruplexes were abundant in the regulatory region related to gene expression, and there are differences in G-quadruplex density between the coding strand and template strand. For the G-quadruplex density of DEGs under drought stress and salt stress, the down-regulated DEGs are generally higher than the up-regulated DEGs. The G-quadruplex formed by SSRs and the flanking sequences of the transcription factor promoter region might enhance the drought tolerance of tobacco. This study greatly promotes the understanding of the distribution and function of G-quadruplexes in tobacco and other plants and provides plentiful available genetic resources for future research.

Supplementary Materials: The following supporting information can be downloaded at <https://www.mdpi.com/article/10.3390/ijms25084331/s1>.

Author Contributions: Conceptualization, L.Y.; methodology, L.Y.; software, K.S. and B.L.; validation, K.S., B.L., H.L., and R.Z.; formal analysis, K.S., B.L., X.Z., and R.L.; investigation, K.S., B.L., H.L., R.Z., and Y.L.; resources, K.S. and X.Z.; data curation, L.Y.; writing—original draft preparation, K.S. and B.L.; writing—review and editing, K.S., B.L., and L.Y.; visualization, K.S. and B.L.; supervision, L.Y.; project administration, L.Y.; funding acquisition, K.S. and L.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Foundation of Shandong Province Modern Agricultural Technology System (SDAIT-25-01).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data are contained within the article or Supplementary Materials. All R scripts have been submitted to GitHub (<https://github.com/KangkangSong123/G4GenomeMAP/tree/master>).

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. 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*, 459–474. [[CrossRef](#)]
2. Moon, J.; Han, J.H.; Kim, D.Y.; Jung, M.J.; Kim, S.K. Effects of deficient of the Hoogsteen base-pairs on the G-quadruplex stabilization and binding mode of a cationic porphyrin. *Biochem. Biophys. Rep.* **2015**, *2*, 29–35. [[CrossRef](#)] [[PubMed](#)]
3. Bhattacharyya, D.; Mirihana Arachchilage, G.; Basu, S. Metal Cations in G-Quadruplex Folding and Stability. *Front. Chem.* **2016**, *4*, 38. [[CrossRef](#)] [[PubMed](#)]
4. Farag, M.; Mouawad, L. Comprehensive analysis of intramolecular G-quadruplex structures: Furthering the understanding of their formalism. *Nucleic Acids Res.* **2024**. [[CrossRef](#)] [[PubMed](#)]
5. Yadav, V.; Hemansi; Kim, N.; Tuteja, N.; Yadav, P. G Quadruplex in Plants: A Ubiquitous Regulatory Element and Its Biological Relevance. *Front. Plant Sci.* **2017**, *8*, 269762. [[CrossRef](#)] [[PubMed](#)]
6. Sparks, M.A.; Singh, S.P.; Burgers, P.M.; Galletto, R. Complementary roles of Pif1 helicase and single stranded DNA binding proteins in stimulating DNA replication through G-quadruplexes. *Nucleic Acids Res.* **2019**, *47*, 8595–8605. [[CrossRef](#)] [[PubMed](#)]
7. Kopec, P.M.; Karlowski, W.M. Sequence Dynamics of Pre-mRNA G-Quadruplexes in Plants. *Front. Plant Sci.* **2019**, *10*, 812. [[CrossRef](#)] [[PubMed](#)]
8. Wu, W.Q.; Zhang, M.L.; Song, C.P. A comprehensive evaluation of a typical plant telomeric G-quadruplex (G4) DNA reveals the dynamics of G4 formation, rearrangement, and unfolding. *J. Biol. Chem.* **2020**, *295*, 5461–5469. [[CrossRef](#)]
9. Mullen, M.A.; Olson, K.J.; Dallaire, P.; Major, F.; Assmann, S.M.; Bevilacqua, P.C. RNA G-Quadruplexes in the model plant species *Arabidopsis thaliana*: Prevalence and possible functional roles. *Nucleic Acids Res.* **2010**, *38*, 8149–8163. [[CrossRef](#)]
10. Tokan, V.; Puterova, J.; Lexa, M.; Kejnovsky, E. Quadruplex DNA in long terminal repeats in maize LTR retrotransposons inhibits the expression of a reporter gene in yeast. *BMC Genom.* **2018**, *19*, 184. [[CrossRef](#)]
11. Vinyard, W.A.; Fleming, A.M.; Ma, J.; Burrows, C.J. Characterization of G-Quadruplexes in *Chlamydomonas reinhardtii* and the Effects of Polyamine and Magnesium Cations on Structure and Stability. *Biochemistry* **2018**, *57*, 6551–6561. [[CrossRef](#)] [[PubMed](#)]
12. Feng, Y.; Tao, S.; Zhang, P.; Sperti, F.R.; Liu, G.; Cheng, X.; Zhang, T.; Yu, H.; Wang, X.E.; Chen, C.; et al. Epigenomic features of DNA G-quadruplexes and their roles in regulating rice gene transcription. *Plant Physiol.* **2022**, *188*, 1632–1648. [[CrossRef](#)] [[PubMed](#)]
13. Zhou, Y.; Kishchenko, O.; Stepanenko, A.; Chen, G.; Wang, W.; Zhou, J.; Pan, C.; Borisjuk, N. The Dynamics of NO₃⁻ and NH₄⁺ Uptake in Duckweed Are Coordinated with the Expression of Major Nitrogen Assimilation Genes. *Plants* **2021**, *11*, 11. [[CrossRef](#)] [[PubMed](#)]

14. Volna, A.; Bartas, M.; Karlicky, V.; Nezval, J.; Kundratova, K.; Pecinka, P.; Spunda, V.; Cerven, 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](#)] [[PubMed](#)]
15. Dobrovolna, M.; Bohalova, N.; Peska, V.; Wang, J.; Luo, Y.; Bartas, M.; Volna, A.; Mergny, J.L.; Brazda, 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](#)] [[PubMed](#)]
16. Cagirici, H.B.; Budak, H.; Sen, T.Z. Genome-wide discovery of G-quadruplexes in barley. *Sci. Rep.* **2021**, *11*, 7876. [[CrossRef](#)] [[PubMed](#)]
17. 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](#)] [[PubMed](#)]
18. Liu, Q.; Yi, C.; Zhang, Z.; Su, H.; Liu, C.; Huang, Y.; Li, W.; Hu, X.; Liu, C.; Birchler, J.A.; et al. Non-B-form DNA tends to form in centromeric regions and has undergone changes in polyploid oat subgenomes. *Proc. Natl. Acad. Sci. USA* **2023**, *120*, e2211683120. [[CrossRef](#)] [[PubMed](#)]
19. 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. [[CrossRef](#)]
20. Doluca, O. G-Quadruplex enrichment analysis reveals their role as intronic regulatory elements in plants. *Turk. J. Bot.* **2019**, *43*, 151–166. [[CrossRef](#)]
21. Yang, X.; Yu, H.; Duncan, S.; Zhang, Y.; Cheema, J.; Liu, H.; Benjamin Miller, J.; Zhang, J.; Kwok, C.K.; Zhang, H.; et al. RNA G-quadruplex structure contributes to cold adaptation in plants. *Nat. Commun.* **2022**, *13*, 6224. [[CrossRef](#)] [[PubMed](#)]
22. Zhang, T.; Yang, M.; Wu, Y.; Jin, S.; Hou, J.; Mao, Y.; Liu, W.; Shen, Y.; Wu, L. Flower Bud Transcriptome Analysis of *Sapium sebiferum* (Linn.) Roxb. and Primary Investigation of Drought Induced Flowering: Pathway Construction and G-Quadruplex Prediction Based on Transcriptome. *PLoS ONE* **2015**, *10*, e0118479. [[CrossRef](#)]
23. Andorf, C.M.; Kopylov, M.; Dobbs, D.; Koch, K.E.; Stroupe, M.E.; Lawrence, C.J.; Bass, H.W. G-Quadruplex (G4) Motifs in the Maize (*Zea mays* L.) Genome Are Enriched at Specific Locations in Thousands of Genes Coupled to Energy Status, Hypoxia, Low Sugar, and Nutrient Deprivation. *J. Genet. Genom.* **2014**, *41*, 627–647. [[CrossRef](#)] [[PubMed](#)]
24. 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](#)]
25. Yang, X.; Cheema, J.; Zhang, Y.; Deng, H.; Duncan, S.; Umar, M.I.; Zhao, J.; Liu, Q.; Cao, X.; Kwok, C.K.; et al. RNA G-quadruplex structures exist and function in vivo in plants. *Genome Biol.* **2020**, *21*, 226. [[CrossRef](#)] [[PubMed](#)]
26. Chang, T.; Li, G.; Ding, Z.; Li, W.; Zhu, P.; Lei, W.; Shangguan, D. Potential G-quadruplexes within the Promoter Nuclease Hypersensitive Sites of the Heat-Responsive Genes in Rice. *Chembiochem* **2022**, *23*, e202200405. [[CrossRef](#)]
27. Sanclemente, M.A.; Ma, F.; Liu, P.; Della Porta, A.; Singh, J.; Wu, S.; Colquhoun, T.; Johnson, T.; Guan, J.C.; Koch, K.E. Sugar modulation of anaerobic-response networks in maize root tips. *Plant Physiol.* **2021**, *185*, 295–317. [[CrossRef](#)]
28. Pečinka, P.; Bohálová, N.; Volná, A.; Kundrátová, 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. [[CrossRef](#)]
29. Huang, R.; Feng, Y.; Gao, Z.; Ahmed, A.; Zhang, W. The Epigenomic Features and Potential Functions of PEG- and PDS-Favorable DNA G-Quadruplexes in Rice. *Int. J. Mol. Sci.* **2024**, *25*, 634. [[CrossRef](#)]
30. Cagirici, H.B.; Sen, T.Z. Genome-Wide Discovery of G-Quadruplexes in Wheat: Distribution and Putative Functional Roles. *G3 Genes | Genomes | Genet.* **2020**, *10*, 2021–2032. [[CrossRef](#)]
31. Falabella, M.; Kolesar, J.E.; Wallace, C.; de Jesus, D.; Sun, L.; Taguchi, Y.V.; Wang, C.; Wang, T.; Xiang, I.M.; Alder, J.K.; et al. G-quadruplex dynamics contribute to regulation of mitochondrial gene expression. *Sci. Rep.* **2019**, *9*, 5605. [[CrossRef](#)] [[PubMed](#)]
32. Chen, Z.; Zhao, N.; Li, S.; Grover, C.E.; Nie, H.; Wendel, J.F.; Hua, J. Plant Mitochondrial Genome Evolution and Cytoplasmic Male Sterility. *Crit. Rev. Plant Sci.* **2017**, *36*, 55–69. [[CrossRef](#)]
33. Daniell, H.; Jin, S.; Zhu, X.G.; Gitzendanner, M.A.; Soltis, D.E.; Soltis, P.S. Green giant—a tiny chloroplast genome with mighty power to produce high-value proteins: History and phylogeny. *Plant Biotechnol. J.* **2021**, *19*, 430–447. [[CrossRef](#)] [[PubMed](#)]
34. Forner, J.; Kleinschmidt, D.; Meyer, E.H.; Fischer, A.; Morbitzer, R.; Lahaye, T.; Schottler, M.A.; Bock, R. Targeted introduction of heritable point mutations into the plant mitochondrial genome. *Nat. Plants* **2022**, *8*, 245–256. [[CrossRef](#)] [[PubMed](#)]
35. Martin Avila, E.; Gisby, M.F.; Day, A. Seamless editing of the chloroplast genome in plants. *BMC Plant Biol.* **2016**, *16*, 168. [[CrossRef](#)] [[PubMed](#)]
36. Rodriguez, R.; Miller, K.M.; Forment, J.V.; Bradshaw, C.R.; Nikan, M.; Britton, S.; Oelschlaegel, T.; Xhemalce, B.; Balasubramanian, S.; Jackson, S.P. Small-molecule-induced DNA damage identifies alternative DNA structures in human genes. *Nat. Chem. Biol.* **2012**, *8*, 301–310. [[CrossRef](#)] [[PubMed](#)]
37. Hansel-Hertsch, R.; Beraldi, D.; Lensing, S.V.; Marsico, G.; Zyner, K.; Parry, A.; Di Antonio, M.; Pike, J.; Kimura, H.; Narita, M.; et al. G-quadruplex structures mark human regulatory chromatin. *Nat. Genet.* **2016**, *48*, 1267–1272. [[CrossRef](#)]
38. Hansel-Hertsch, R.; Simeone, A.; Shea, A.; Hui, W.W.I.; Zyner, K.G.; Marsico, G.; Rueda, O.M.; Bruna, A.; Martin, A.; Zhang, X.; et al. Landscape of G-quadruplex DNA structural regions in breast cancer. *Nat. Genet.* **2020**, *52*, 878–883. [[CrossRef](#)] [[PubMed](#)]
39. Lago, S.; Nadai, M.; Cernilogar, F.M.; Kazerani, M.; Dominiguez Moreno, H.; Schotta, G.; Richter, S.N. Promoter G-quadruplexes and transcription factors cooperate to shape the cell type-specific transcriptome. *Nat. Commun.* **2021**, *12*, 3885. [[CrossRef](#)]

40. Zhang, Y.; Fan, C.; Chen, Y.; Wang, R.R.; Zhang, X.; Han, F.; Hu, Z. Genome evolution during bread wheat formation unveiled by the distribution dynamics of SSR sequences on chromosomes using FISH. *BMC Genom.* **2021**, *22*, 55. [[CrossRef](#)]
41. Kaur, S.; Panesar, P.S.; Bera, M.B.; Kaur, V. Simple sequence repeat markers in genetic divergence and marker-assisted selection of rice cultivars: A review. *Crit. Rev. Food Sci. Nutr.* **2015**, *55*, 41–49. [[CrossRef](#)] [[PubMed](#)]
42. Li, H.; Song, K.; Li, B.; Zhang, X.; Wang, D.; Dong, S.; Yang, L. CRISPR/Cas9 Editing Sites Identification and Multi-Elements Association Analysis in *Camellia sinensis*. *Int. J. Mol. Sci.* **2023**, *24*, 15317. [[CrossRef](#)]
43. Hoang, X.L.T.; Nhi, D.N.H.; Thu, N.B.A.; Thao, N.P.; Tran, L.-S.P. Transcription Factors and Their Roles in Signal Transduction in Plants under Abiotic Stresses. *Curr. Genom.* **2017**, *18*, 483–497. [[CrossRef](#)] [[PubMed](#)]
44. Duan, L.; Mo, Z.; Fan, Y.; Li, K.; Yang, M.; Li, D.; Ke, Y.; Zhang, Q.; Wang, F.; Fan, Y.; et al. Genome-wide identification and expression analysis of the bZIP transcription factor family genes in response to abiotic stress in *Nicotiana tabacum* L. *BMC Genom.* **2022**, *23*, 318. [[CrossRef](#)]
45. Nuruzzaman, M.; Sharoni, A.M.; Kikuchi, S. Roles of NAC transcription factors in the regulation of biotic and abiotic stress responses in plants. *Front. Microbiol.* **2013**, *4*, 248. [[CrossRef](#)]
46. Song, K.; Li, B.; Wu, H.; Sha, Y.; Qin, L.; Chen, X.; Liu, Y.; Tang, H.; Yang, L. The Function of BBX Gene Family under Multiple Stresses in *Nicotiana tabacum*. *Genes* **2022**, *13*, 1841. [[CrossRef](#)]
47. Bai, G.; Yang, D.H.; Cao, P.; Yao, H.; Zhang, Y.; Chen, X.; Xiao, B.; Li, F.; Wang, Z.Y.; Yang, J.; et al. Genome-Wide Identification, Gene Structure and Expression Analysis of the MADS-Box Gene Family Indicate Their Function in the Development of Tobacco (*Nicotiana tabacum* L.). *Int. J. Mol. Sci.* **2019**, *20*, 5043. [[CrossRef](#)] [[PubMed](#)]
48. Zhang, H.; Zhu, J.; Gong, Z.; Zhu, J.-K. Abiotic stress responses in plants. *Nat. Rev. Genet.* **2021**, *23*, 104–119. [[CrossRef](#)]
49. Fernandez-Pozo, N.; Menda, N.; Edwards, J.D.; Saha, S.; Teclé, I.Y.; Strickler, S.R.; Bombarely, A.; Fisher-York, T.; Pujar, A.; Foerster, H.; et al. The Sol Genomics Network (SGN)—From genotype to phenotype to breeding. *Nucleic Acids Res.* **2015**, *43*, D1036–D1041. [[CrossRef](#)]
50. Sayers, E.W.; Bolton, E.E.; Brister, J.R.; Canese, K.; Chan, J.; Comeau, D.C.; Connor, R.; Funk, K.; Kelly, C.; Kim, S.; et al. Database resources of the national center for biotechnology information. *Nucleic Acids Res.* **2022**, *50*, D20–D26. [[CrossRef](#)]
51. Bedrat, A.; Lacroix, L.; Mergny, J.L. Re-evaluation of G-quadruplex propensity with G4Hunter. *Nucleic Acids Res.* **2016**, *44*, 1746–1759. [[CrossRef](#)] [[PubMed](#)]
52. Beier, S.; Thiel, T.; Munch, T.; Scholz, U.; Mascher, M. MISA-web: A web server for microsatellite prediction. *Bioinformatics* **2017**, *33*, 2583–2585. [[CrossRef](#)] [[PubMed](#)]
53. Chen, C.; Chen, H.; Zhang, Y.; Thomas, H.R.; Frank, M.H.; He, Y.; Xia, R. TBtools: An Integrative Toolkit Developed for Interactive Analyses of Big Biological Data. *Mol. Plant* **2020**, *13*, 1194–1202. [[CrossRef](#)] [[PubMed](#)]
54. Cantalapiedra, C.P.; Hernandez-Plaza, A.; Letunic, I.; Bork, P.; Huerta-Cepas, J. eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. *Mol. Biol. Evol.* **2021**, *38*, 5825–5829. [[CrossRef](#)] [[PubMed](#)]
55. Wu, T.; Hu, E.; Xu, S.; Chen, M.; Guo, P.; Dai, Z.; Feng, T.; Zhou, L.; Tang, W.; Zhan, L.; et al. clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innovation* **2021**, *2*, 100141. [[CrossRef](#)] [[PubMed](#)]
56. Katz, K.; Shutov, O.; Lapoint, R.; Kimelman, M.; Brister, J.R.; O’Sullivan, C. The Sequence Read Archive: A decade more of explosive growth. *Nucleic Acids Res.* **2022**, *50*, D387–D390. [[CrossRef](#)]
57. Bolger, A.M.; Lohse, M.; Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **2014**, *30*, 2114–2120. [[CrossRef](#)] [[PubMed](#)]
58. Kim, D.; Paggi, J.M.; Park, C.; Bennett, C.; Salzberg, S.L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* **2019**, *37*, 907–915. [[CrossRef](#)] [[PubMed](#)]
59. Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R.; Genome Project Data Processing, S. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **2009**, *25*, 2078–2079. [[CrossRef](#)]
60. Pertea, M.; Pertea, G.M.; Antonescu, C.M.; Chang, T.C.; Mendell, J.T.; Salzberg, S.L. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* **2015**, *33*, 290–295. [[CrossRef](#)]
61. Love, M.I.; Huber, W.; Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **2014**, *15*, 550. [[CrossRef](#)] [[PubMed](#)]
62. Letunic, I.; Bork, P. Interactive Tree Of Life (iTOL) v5: An online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* **2021**, *49*, W293–W296. [[CrossRef](#)] [[PubMed](#)]

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.