Available online at www.sciencedirect.com

ScienceDirect

Journal of Genetics and Genomics 41 (2014) 627–647

JGG

ORIGINAL RESEARCH

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

Carson M. Andorf [a,b], Mykhailo Kopylov [c], Drena Dobbs [d], Karen E. Koch [e], M. Elizabeth Stroupe [c,f], Carolyn J. Lawrence [d,g], Hank W. Bass [f,*]

[a] *USDA-ARS Corn Insects and Crop Genetics Research Unit, Iowa State University, Ames, IA 50011, USA*
[b] *Department of Computer Science, Iowa State University, Ames, IA 50011, USA*
[c] *Institute of Molecular Biophysics, Florida State University, Tallahassee, FL 32306-4370, USA*
[d] *Department of Genetics, Development and Cell Biology, Iowa State University, Ames, IA 50011, USA*
[e] *Plant Molecular and Cellular Biology Program, Horticultural Sciences Department, Genetics Institute, University of Florida, Gainesville, FL 32611, USA*
[f] *Department of Biological Science, Florida State University, Tallahassee, FL 32306-4295, USA*
[g] *Department of Agronomy, Iowa State University, Ames, IA 50011, USA*

Received 23 August 2014; revised 16 October 2014; accepted 24 October 2014
Available online 4 November 2014

ABSTRACT

The G-quadruplex (G4) elements comprise a class of nucleic acid structures formed by stacking of guanine base quartets in a quadruple helix. This G4 DNA can form within or across single-stranded DNA molecules and is mutually exclusive with duplex B-form DNA. The reversibility and structural diversity of G4s make them highly versatile genetic structures, as demonstrated by their roles in various functions including telomere metabolism, genome maintenance, immunoglobulin gene diversification, transcription, and translation. Sequence motifs capable of forming G4 DNA are typically located in telomere repeat DNA and other non-telomeric genomic loci. To investigate their potential roles in a large-genome model plant species, we computationally identified 149,988 non-telomeric G4 motifs in maize (*Zea mays* L., B73 AGPv2), 29% of which were in non-repetitive genomic regions. G4 motif hotspots exhibited non-random enrichment in genes at two locations on the antisense strand, one in the 5′ UTR and the other at the 5′ end of the first intron. Several genic G4 motifs were shown to adopt sequence-specific and potassium-dependent G4 DNA structures *in vitro*. The G4 motifs were prevalent in key regulatory genes associated with hypoxia (group VII ERFs), oxidative stress (DJ-1/GATase1), and energy status (AMPK/SnRK) pathways. They also showed statistical enrichment for genes in metabolic pathways that function in glycolysis, sugar degradation, inositol metabolism, and base excision repair. Collectively, the maize G4 motifs may represent conditional regulatory elements that can aid in energy status gene responses. Such a network of elements could provide a mechanistic basis for linking energy status signals to gene regulation in maize, a model genetic system and major world crop species for feed, food, and fuel.

KEYWORDS: Maize; G-quadruplex; G4; Hypoxia; Sucrose synthase

INTRODUCTION

The G-quadruplex (G4) elements are non-duplex four-stranded nucleic acid structures that can form in DNA or RNA (Simonsson, 2001; Burge et al., 2006; Qin and Hurley, 2008; Brooks et al., 2010; Huppert, 2010; Yang and Okamoto,

* Corresponding author. Tel: +1 850 644 9711, fax: +1 850 645 8447.
E-mail address: bass@bio.fsu.edu (H.W. Bass).

http://dx.doi.org/10.1016/j.jgg.2014.10.004

2010; Bochman et al., 2012). These G4s can form within one strand to produce an intra-molecular element made of planar stacks of guanine quartets in which each quadruplex has four strands and three loops. The ability to form G4 elements is known to be a characteristic of the G-rich strand of telomere repeat sequences for most eukaryotic species (Sen and Gilbert, 1988; Sundquist and Klug, 1989; Williamson et al., 1989). Roles for G4 elements in telomere metabolism, DNA replication, genome stability, and meiotic nuclear architecture are widely conserved across eukaryotic kingdoms (Blackburn et al., 2006), and telomere length homeostasis in animals is known to be affected by stress and aging (reviewed in Armanios and Blackburn, 2012; Blackburn and Epel, 2012). More recently, non-telomeric G4 motifs distributed throughout the genome at internal chromosomal sites have been shown to affect a vast array of genetic functions, including transcription, translation, replication, recombination, and DNA repair (Bochman et al., 2012; Maizels and Gray, 2013). The genomic distribution of G4 motifs in and around genes has prompted numerous investigations into their possible roles as *cis*-acting regulatory elements (Huppert and Balasubramanian, 2005; Todd et al., 2005; Burge et al., 2006; Eddy and Maizels, 2006, 2008; Maizels, 2006; Hershman et al., 2008; Du et al., 2009; Bochman et al., 2012). Finding G4 motif enrichment in the promoters of mammalian genes involved in cell-cycle control and cancer progression suggests a fundamental role for G4 in cell growth and development (Huppert, 2007; Qin and Hurley, 2008; Brooks and Hurley, 2009; Yang and Okamoto, 2010; Beckett et al., 2012; Juranek and Paeschke, 2012; Weng et al., 2012; Cahoon and Seifert, 2013; Yuan et al., 2013).

A well-characterized example of G4 DNA influence on transcription is found in the promoter of the *c-MYC* proto-oncogene, where the hypersensitive site $NHEIII_1$ was shown to require G4 formation for functional regulation *in vitro* and *in vivo* (Siddiqui-Jain et al., 2002). Subsequent studies implicated G4 DNA elements in transcriptional regulation of other cancer genes, such as *KRAS* (Cogoi and Xodo, 2006), *c-myb* (Palumbo et al., 2008), *c-kit* (Fernando et al., 2006), *VEGF* (Guo et al., 2008), *PDGFR-β* (Qin et al., 2010), *HIF-1α* (De Armond et al., 2005), *bcl-2* (Dexheimer et al., 2006), *Rb* (Xu and Sugiyama, 2006), *RET* (Guo et al., 2007) and *hTERT* (Palumbo et al., 2009). Ongoing pharmacological (Chen and Yang, 2012), mutational (Pontier et al., 2009), and evolutionary research (Verma et al., 2009) into possible G4 DNA functions provides new opportunities for development of G4-targeted drugs to treat certain human diseases (Brooks and Hurley, 2009, 2010; Yang and Okamoto, 2010). Presence of G4 DNA is also involved in other diverse biological processes including DNA repair and helicase-associated genome modifications (Chinnapen and Sen, 2004; Huber et al., 2006; Clark et al., 2012). Importantly, advances in the ability to detect G4 DNA in fixed and living cells have borne out predictions of *in vivo* G4 DNA occurrence and its functional significance (Biffi et al., 2013; Lam et al., 2013; Henderson et al., 2014).

Genome-wide computational screens for G4 motifs have revealed their widespread occurrence in prokaryotic and eukaryotic species including *E. coli*, budding yeast, and humans (Todd et al., 2005; Rawal et al., 2006; Du et al., 2008; Halder et al., 2009; Capra et al., 2010; Eddy et al., 2011; Todd and Neidle, 2011). These studies enabled species-specific and comparative genomic analyses of potential G4 functions, revealing evidence for both conservation and divergence of certain G4 DNA functions (Huppert and Balasubramanian, 2005; Kikin et al., 2006; Verma et al., 2008; Yadav et al., 2008; Mani et al., 2009; Stegle et al., 2009; Wong et al., 2010; Clark et al., 2012; Menendez et al., 2012; Maizels and Gray, 2013). Only a few of these analyses include representatives of the plant kingdom (Mullen et al., 2010; Takahashi et al., 2012; Lexa et al., 2014).

Investigations in plant species offer wide eukaryotic comparisons as well as the potential for identification of genetic processes that directly impact traits of global importance in food, fuel, and biomass production. Initial research into plant G4 motifs focused on these elements in *Arabidopsis* transcriptomes (Mullen et al., 2010), the genomes of other plant species (Takahashi et al., 2012), and LTR (long-terminal repeat) retrotransposons (Lexa et al., 2014). A comparative survey of G4 motifs in several plant species revealed that plant G4 motifs are often located on the template strand near the transcription start site (TSS) of genes, whereas gene ontology analysis uncovered G4 associations with the terms chloroplast, nucleus, and histone (Mullen et al., 2010; Takahashi et al., 2012). Among the model plant species, maize has a rich history of genetic discoveries regarding transmission genetics, telomere functions, mobile DNA elements, epigenetics, and genetic diversity (reviewed in Bennetzen and Hake, 2009). The ten-chromosome maize genome ranges in size from 2.45 pg (~2.3 Gb) to about 3.35 pg (~3.2 Gb) among different inbred lines (Laurie and Bennett, 1985). The inbred B73 has been sequenced (B73 RefGen_v2) with a current assembly length of ~2.1 Gb (Schnable et al., 2009). To systematically examine the occurrence and conservation of G4 motifs in this model plant species, we used a bioinformatic screen to identify over 43,000 G4 motifs in the non-repetitive fraction of the genome. Here we analyze the distribution of maize G4 motifs, and discuss the possible roles of G4 DNA as *cis*-regulatory elements in genes responsive to energy status.

RESULTS

Genome-wide computational survey for G4 motifs in maize

We analyzed the entire maize genome using the conservative and commonly employed G4 motif consensus settings for the Quadparser software, $G_{3+}L_{1\text{-}7}$ (Huppert and Balasubramanian, 2005). This conservative definition of a G4 motif is defined by the presence of four runs of guanines (3 or more Gs per run) separated by intervening (loop) sequences from one to seven of any base, a motif referred to as $G_3L_{1\text{-}7}$ (Huppert and Balasubramanian, 2005). G4 motifs matching this consensus have been shown to be likely to form stable G4 structures *in vitro* (Hazel et al., 2004; Bugaut and Balasubramanian,

2008). Sequence searches using less stringent consensus sequences, such as longer loops or shorter runs of guanines result in identification of even longer lists of G4 motifs (Bourdoncle et al., 2006; Mullen et al., 2010). We identified and named nearly 150,000 G4 motifs, averaging 27 bases in size (summarized in Fig. 1). Each G4 sequence motif was given a unique identification (G4v2-1, G4v2-2, etc.) and defined by the chromosome number, bp coordinates within the B73 AGP_v2 sequence build, and strand (compiled in Table S1). The G4 motifs showed a global relative density of 74/Mb, but nearly twice that amount in the repeat-masked, non-repetitive portion of the maize genome (Fig. 1A). The RefGen_v2 build of the whole maize genome consists of 2.07 billion base pairs of which 1.2% is composed of gap sequence (a fixed number of N's that defines the type of gap between sequence contigs) and 83% is composed of repetitive elements. A large number of G4 motifs were detected in repeated sequences, including clustered patches of G4 motifs on the same strand (6910 G4 motif patches listed in Table S2). The analyses reported here are limited to the repeat-masked low copy-number sequence regions of the maize genome, as described in the Materials and Methods.

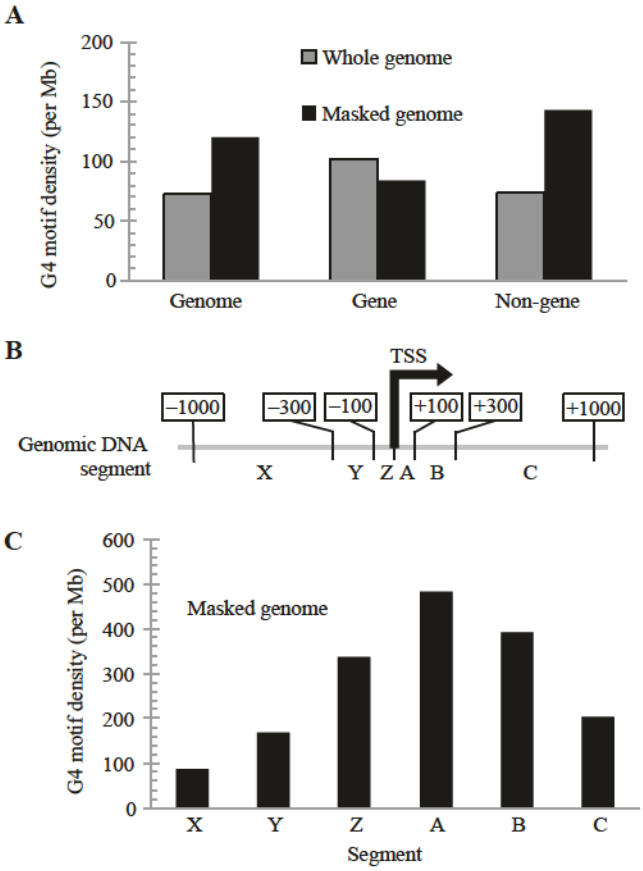


Fig. 1. G4 motif densities in specific genomic regions.

A: Histogram represents average number of G4 motifs per Mb for the B73 RefGen_v2 (Whole genome, gray), the B73 RefGen_v2 assembly masked to remove repetitive elements (Masked genome, black) for entire maize genome, including both gene space and non-gene space. **B:** Schematic of gene-associated segments analyzed. Segments X–Z represent binned regions upstream of the transcription start site (TSS) and segments A–C represent regions downstream from the TSS. For example, segment X represents sequences in the region between 1000 and 300 bp upstream of the TSS, and segment A represents sequences in the region between 1 and 100 bp downstream of the TSS. **C:** Histogram illustrating the G4 motif density in each gene segment defined in Fig. 1B. Values for each segment represent the average density computed overall genes in the Masked Genome (aligned relative to the TSS).

Over 43,000 G4 motifs remained after repetitive sequences were masked out of the analysis. Of these, nearly 12,000 G4 motifs were associated with gene models (see Table S3 for a list of all gene models with their respective G4 motifs). Among the 39,570 canonical transcripts from the high-quality filtered gene set (see Materials and Methods), 9572 have one or more G4 motifs. We examined the location of these motifs in genes that were segmented into various structural components and boundary areas (Fig. 1B). Using TSS-aligned, gene-average plots of G4 motifs, normalized per Mb of each segment, we found that the highest G4 motif density mapped to the first 100 bp downstream of the TSS (segment A in Fig. 1C). In this region, G4 motif density peaked at about 500 G4 motifs per Mb, nearly 10 times more abundant than that of the whole genome, and four times more abundant than that of the repeat-masked genome. The next highest G4 motif density was in segment "B", 100–300 bp downstream of the TSS. This segment includes varying amounts of gene structural elements including introns and coding sequences. The third highest G4 motif density was in segment "Z", in the first 100 bp just upstream of the TSS, where densities exceeded 350 G4 motifs per Mb of sequence. The G4 motif density in this TSS proximal region is approximately four times greater than that in the overall masked genome space, suggestive of a role for these sequences in regulating some aspect of gene expression.

G4 motif hotspots map to regulatory regions for gene expression

Having established that maize G4 motifs are enriched at genes relative to genomic DNA, we looked more closely at the locations of these elements in relation gene structural features. Takahashi et al. (2012) showed that the TSS region of the template strand was a hotspot for G4 motifs in four plant species: *Arabidopsis*, grape, rice, and poplar. If that location were generally conserved across other plant species, we would expect a similar TSS G4 motif hotspot in maize. We examined this possibility in detail along with two other gene structure boundaries, as summarized in Fig. 2. The segments of a schematic intron-containing gene (Fig. 2A) show the three boundaries around which averaged trend plots are diagrammed (Fig. 2B–D). The three boundaries were defined by the TSS (Fig. 2B, using the canonical transcript and orienting all genes in the same direction), the first exon-intron boundary ("Exon-Intron", Fig. 2C), and the start of the open reading frame on the mRNA ("AUG", Fig. 2D), as defined by the start codon of the canonical transcript model. To address the question of bulk G-richness as contributing to site-specific G4 motif abundance, we also simulated a set of 40,000 genes with randomized sequences. These simulations were produced by

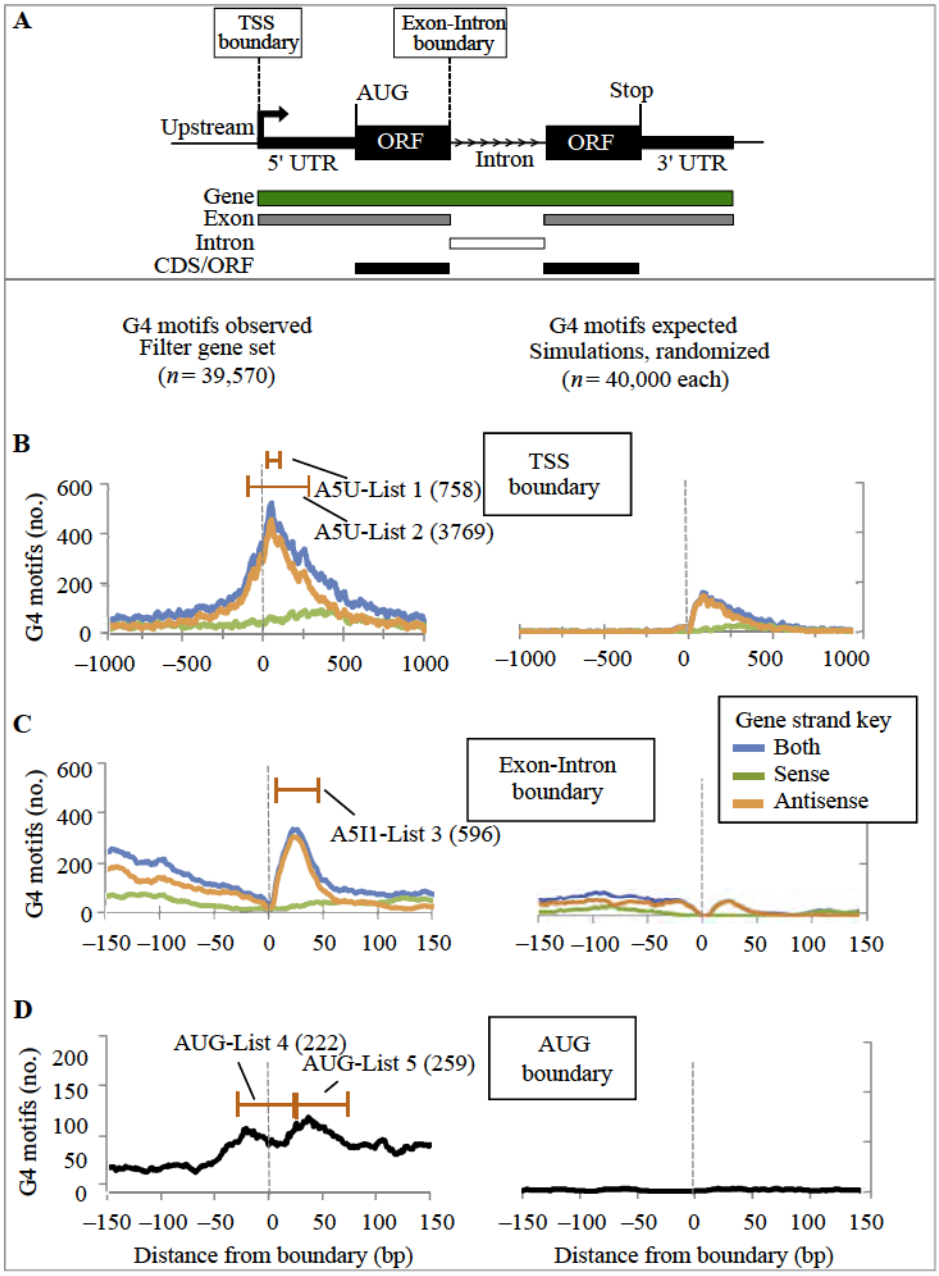


Fig. 2. Positional enrichment of G4 motifs around specific gene features.

Schematic of gene structure and boundary definitions for three boundaries analyzed by comparing observed expected G4 motif frequencies. **A:** The three boundaries are the transcription start site (TSS), the translation start site (AUG, start codon), and the first exon-intron boundary (Exon-Intron). The TSS arrow denotes the direction of transcription, and the segments of the schematic gene (5′ UTR, ORF, intron, 3′ UTR) are labeled. The abundance of G4 motifs at specific positions are plotted as the total number of G4 motifs that overlap each base for the observed (39,570 filtered gene set) or expected (40,000 simulations – see Materials and Methods) set of sequences. The G4 motif abundance was plotted for each strand-specific colors for the TSS (**B**) boundary and the 1st Exon-Intron (**C**) boundary, or for the coding/sense strand for the mRNA AUG (**D**) boundary. For each plot, the featured boundary is depicted by a vertical dashed line in the center of the plot, with base pair distances (*x* axis) indicated. Peaks selected to make gene lists (Lists 1–5) are delimited by brackets, defined in the results, totaled in parentheses, and recorded in Tables S3 and S4.

random sampling from nucleotide position- and strand-specific base frequency tables empirically derived from the filtered gene set (n = 39,570). Observation *versus* prediction (Fig. 2B–D, left *versus* right plots) comparison showed that G4 motifs were enriched in the antisense (template) strand at the 5′ UTR and 5′ end of intron 1. These enriched G4 motif hotspots were not explained by base composition alone, even when strand information was preserved. From these analyses, it appears that regional stretches of G content over 30% are sufficient to cause detectable G4 motifs in the randomized sequence simulations. Similarly, but on a smaller scale, we observed enrichment for G4 motifs on the mRNA (sense strand) around the ORF start codon boundary (Fig. 2D).

To further analyze groups of genes with G4 motifs in specific locations, we segmented the peaks in the G4 motif trend plots according to their location and strand to produce a total of

5 lists of genes for subsequent analysis (Fig. 2, Lists 1–5). These G4 motif hotspots are respectively denoted "A5U" for antisense, 5′ UTR, and "A5I1" for antisense, 5′ end of intron 1. Together, they account for the vast majority of genic G4 motifs, with over 4000 maize genes (10% of the maize genome) having at least one these two antisense G4 motif motifs. The G4 motifs in the predicted mRNA were denoted "AUG" for their proximity to the start codon and their presence in the mRNA transcript. We observed 481 maize genes with AUG G4 motifs by this criterion, and though fewer in number, these elements are interesting because of their potential to impact translational regulation. The "A5U-List 1" genes were a subset of the "A5U-List 2" genes, which collectively account for a remarkable 9.5% (3769 total) of all maize genes in the filtered gene set (39,570 total). In addition, 596 genes contained a G4 motif within the first 50 bases of intron 1 (A5I1-List 3). The genes with AUG-proximal G4 motifs also numbered in the hundreds −222 genes in AUG-List 4 and 259 genes in AUG-List 5. The identity of the genes for the 5 lists reveals that some genes contain more than one G4 motif (Tables S3 and S4). Together, this computational screen and positional frequency analysis have uncovered an extraordinary number of genes with potential to adopt G4, non-B-form DNA structures at locations implicated in the major gene regulatory processes of transcription, splicing, and translation.

We next examined the evolutionary conservation of G4 motifs in homologs of maize genes. If the presence of these elements were non-functional or coincidental, then no evolutionary conservation would be expected. To test for the most parsimonious explanation, i.e., functional conservation, we examined conserved orthologs of maize G4 motif-containing genes and plotted the results as shown in Fig. 3. The positional frequencies of antisense G4 motifs for maize genes were classified according to the quality of gene model evidence within maize, and according to the degree of conservation between these genes and their syntenic counterparts in four relatives of maize. These relatives are, in order of increasing evolutionary divergence: sorghum, foxtail millet, brachy, and rice. This analysis shows that the "lowest quality" genes (i.e., those within the working gene set (WGS), but rejected from the higher quality filtered gene set (FGS); Fig. 3A, orange segment), show negligible likelihood of carrying an antisense/template G4 motif. These rejected genes include pseudogenes, gene fragments, or gene-like sequences for which mRNAs evidence is lacking or minimal. Within the FGS, we observe an increasing prevalence of antisense/template G4 motifs as we go from genes found only in maize (Fig. 3A, blue segment), to those found in maize and any two of the four pan-grass relatives (Fig. 3A, gold segment). The highest G4 motif densities occurred in the most highly-conserved genes, those with known syntenic orthologs (syntelogs) in all four of the other pan-grass species (Fig. 3A, blue segment), with G4 frequency peaks just downstream of the TSS (Fig. 3B). Taken together, this analysis supports the idea that the maize genic G4 motifs represent a large group of sequence motifs with evolutionarily constrained locations relative to gene structures in the grass family of plants.

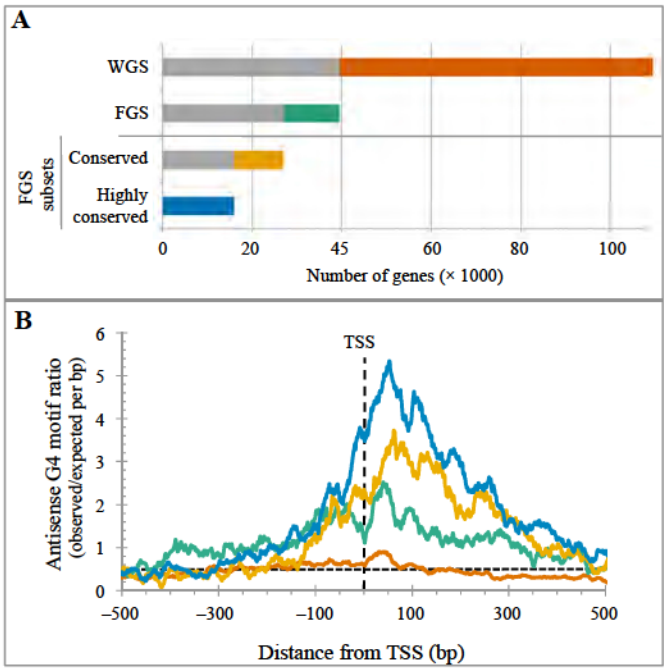


Fig. 3. Conservation of TSS-associated antisense (template strand) G4 motifs in maize and maize relatives.

A: Four horizontal, segmented bars represent the number of genes (*x* axis) in each of four nested subsets of maize annotated gene models. The first bar from the top shows the maize working gene set, WGS (n = 109,461 genes). The second bar from the top shows the filtered gene set, FGS, a high-confidence subset of the WGS. The third and fourth bars show subsets of the FGS that have syntenic orthologs (syntelogs) in other grass species. These subsets comprise the "conserved" (one or more species contain syntelogs to the maize gene; third horizontal bar) or "highly-conserved" (5 grass species contain syntelogs to the maize gene; fourth horizontal bar) gene sets. Each of the top three bars is segmented into two colors: the gray segment contains only genes that meet criteria to be included in the subsequent subset. The orange segment of WGS represents genes that were filtered out to designate a high-quality FGS. The green segment contains FGS-specific genes that have no syntelogs. Within the syntenic subsets, the gold segment contains FGS-specific maize genes that are represented by a syntelog in one to four of the five grass species interrogated. The blue segment, which encompasses the fourth bar, represents maize genes with syntelogs in all five grass species compared. **B:** Following the color coding convention defined in panel **A**, template/antisense G4 motif ratios (*y* axis) for the four annotated gene subsets are plotted against distance from TSS (*x* axis).

Metabolic pathway analysis associates maize G4 motifs with genes for responses to hypoxia, redox, and nutrient stresses

We next examined another prediction of the computational screen for maize G4 motif-containing genes, that potential for co-regulated roles could be revealed by analysis of functional groupings for genes containing G4 motifs. Functional classification of genes could reveal this trend, while also pointing to possible regulatory roles of these motifs. We explored the biological function of maize G4 motif-containing genes using the metabolic pathway databases MaizeCyc (Karp et al., 2010; Monaco et al., 2013) and MapMan (Thimm et al., 2004), looking specifically for non-random distribution of these genes among the many pathways depicted.

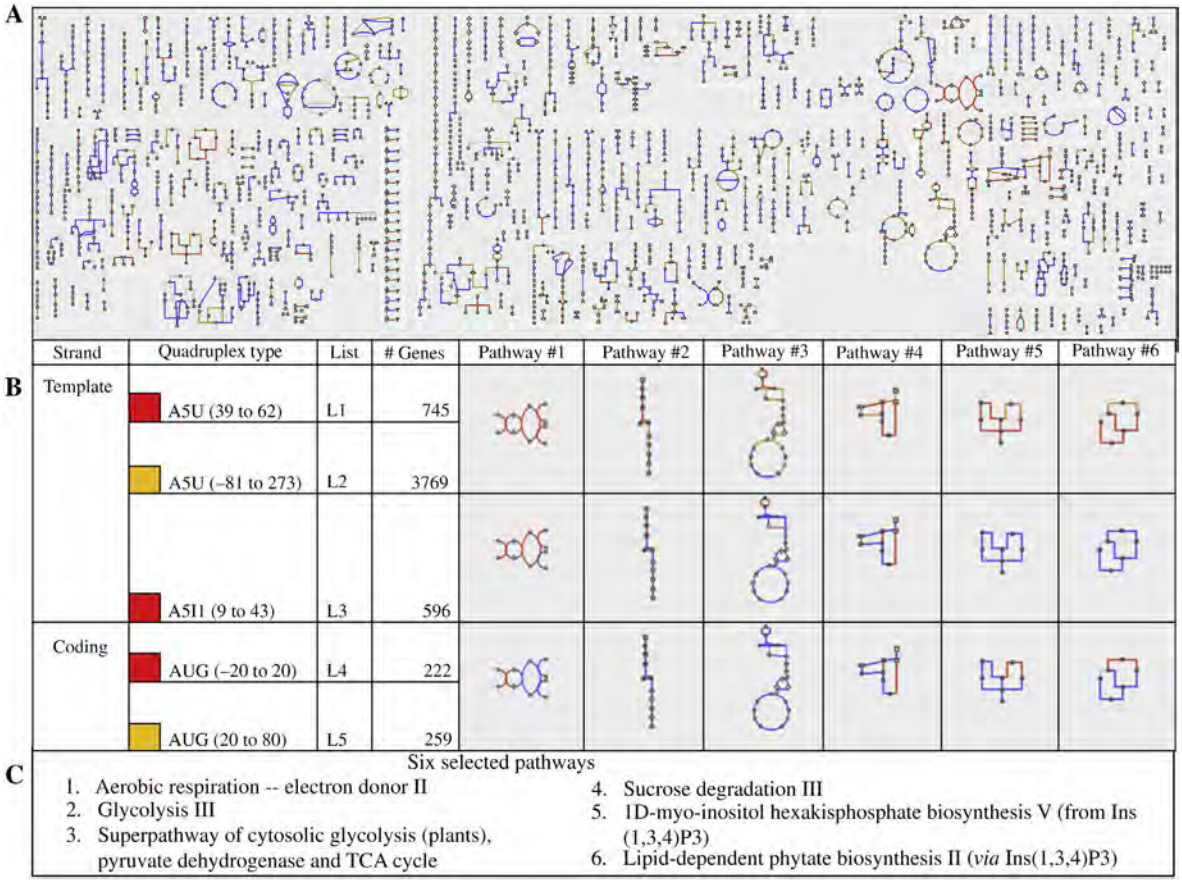


Fig. 4. Metabolic pathways highlighted with G4 motif gene lists.

A: Schematic view of all pathways of *Zea mays* ssp. *mays* metabolism (version 2.02 of MaizeCyc). Nodes represent metabolites and lines represent reactions, highlighted as follows: RED: At least one metabolite has been assigned to a gene model in G4 motif List 1 (see Fig. 2C and Tables S3 and S4); ORANGE: At least one metabolite has been assigned to a gene model in G4 motif List 2; BLUE: Gene model(s) has been assigned to a metabolite, but none of the gene models are in List 1 or 2; GRAY: No maize gene model has been assigned to the metabolite reaction. **B:** Table showing the quadruplex types associated with the five gene lists (Lists 1–5), along with highlighted diagrams for six selected pathways (see panel **C**), color-coded as described above. **C:** Names of the 6 selected pathways used in panel **B**, as defined in MaizeCyc.

We used the MaizeCyc gene expression omics tool to color-code the presence or absence of G4 motif-containing genes as shown in Fig. 4. According to the positional hotspot peaks of G4 motif locations, we developed three sets of gene lists: A5U, A5I1, and AUG (defined in Fig. 2, Lists 1–5). We showed the MaizeCyc display for the gene list from one of these groups, A5U, using orange and red lines to indicate enzymatic steps to which one of our listed genes has been mapped (Fig. 4A). Several pathways were identified by this method and are enlarged for each of the three categories, A5U, A5I1, and AUG (Fig. 4B). These pathways include aerobic respiration, Glycolysis, parts of the TCA cycle, and two inositol phosphate-related pathways (Fig. 4B and C, pathways 5 and 6). The html files of omics overview pathway, like that shown for the A5U (A5U-List 1, A5U-List 2, Fig. 4A), were produced for these and the other two gene lists (A5I1-List 3, AUG-List 4, and AUG-List 5) and made available as web-browser files for mouse-over identification of the pathways with G4 motif-containing genes (URLs in Table S5).

Among the notable G4 motif-containing genes observed using the MaizeCyc omics viewer were numerous genes also reported to modulate responses to hypoxia in diverse plant species (Bailey-Serres et al., 2012a). Select examples from specific biological processes or pathways are listed in Table 1, along with the gene model ID, chromosome location, and class of G4 motif contained therein. For instance, in maize, the sucrose synthase gene encoded by *Shrunken1* contains an A5U G4 motif, and this gene is also induced by hypoxia (Sachs et al., 1980) and was expressed in the low-oxygen region of the endosperm of developing maize seed (Zeng et al., 1998). Using Fisher's exact test for non-random associations between antisense G4 motifs in genes or gene locations and these metabolic pathways, we found evidence of either enrichment or depletion, as summarized in Table 2. For instance, the G4 motifs of A5U-List 1 were enriched 5-fold ($p = 1.8E\text{-}4$) for the sucrose degradation III/"Pathway 4", a pathway to which *Shrunken1* is assigned.

Unexpectedly, two recurring sets of genes involved in *myo*-inositol and phytic acid biosynthesis that have not previously been associated with plant response to hypoxia were observed in our G4 motif screen (Fig. 4B and C, pathways 5 and 6, and Table 1). Detection of these pathways raises questions of

Table 1
Maize G4-motif-containing nuclear genes with key roles in energy metabolism, hypoxia, and nutrient signaling pathways

Functional class	Maize locus name[a]	Protein product[b] (PFAM domain)	Genic G4 motif class[c]	Gene model ID[d]	Genetic bin[e]
Sugar metabolism	*sh1, shrunken1*	Sucrose synthase (PF00862, PF00534)	A5U(G4v2-133636)	GRMZM2G089713	9.01
	amyb1, beta amylase1	Beta-amylase (PF01373)	S3U(G4v2-334)	GRMZM2G450125	1.01
	**amyb2, beta amylase2*	beta-amylase (PF01373)	2-S3U#(G4v2-82729, G4v2-82730)	GRMZM2G025833	5.03
Sugar transport	*sut1, sucrose transporter1*	Sucrose transporter (cd06174)[f]	A5U(G4v2-533)	GRMZM2G034302	1.04
Inositol modification	*ipp2k, inositol polyphosphate 2-kinase*	Inositol-pentakisphosphate 2-kinase (PF06090)	A5U(G4v2-143979)	GRMZM2G067299	10.07
	**lpa3, low phytic acid3*	Inositol kinase; ribokinase (PF00294)	2-A5U, AUG(G4v2-9798, G4v2-9799, G4v2-20711)	GRMZM2G361593	1.10
Energy, glycolysis	**hex4, hexokinase4*	Hexokinase (PF00349, PF03727)	A5U, A5I1, AUG(G4v2-53046, G4v2-53047, G4v2-44784)	GRMZM2G068913	3.05
	eno1, enolase1	Enolase (PF03952, PF00113, PF07476)	A5U, A5I1, AUG(G4v2-134002)	GRMZM2G064302	9.02
	pdlk1, pyruvate dehydrogenase (lipoamide) kinase1	Mitochondrial pyruvate dehydrogenase kinase (PF10436, PF02518)	A5U(G4v2-13214)	GRMZM2G107196	1.04
	**pyk3, pyruvate kinase3*	Cytosolic pyruvate kinase (PF00224, PF02887)	A5U(G4v2-30561)	GRMZM2G150098	2.01
	pdh2, pyruvate dehydrogensase2	Pyruvate dehydrogenase (PF02779, PF02780)	A5U, AI(G4v2-17986)	GRMZM2G043198	1.06
Energy, NADPH production	*pdg2, 6-phosphogluconate dehydrogenase2*	NADPH producing dehydrogenase of the oxidative PPP (PF03446, PF00393)	A5U, A5I1, SI1(G4v2-44842, G4v2-44841, G4v2-53136)	GRMZM2G145715	3.05
Energy, oxoglutarate metabolism	*omt1, oxoglutarate malate transporter1*	Plastidic 2-oxoglutarate/malate transporter (PF00939)	2-AORF(G4v2-139656, G4v2-139657)	GRMZM2G383088	10.03
	gln1, glutamine synthetase1	Glutamine synthetase; glutamate-ammonia ligase, chloroplast (PF03951, PF00120)	A5U(G4v2-143971)	GRMZM2G098290	10.07
Energy, respiration	**sudh1, succinate dehydrogenase1*	Mitochondrial succinate dehydrogenase (PF00890, PF02910)	2-A5U(G4v2-108474, G4v2-108475)	GRMZM2G064799	7.00
	**nad3, NADH-ubiquinone oxidorectase3*	NADH-ubiquinone oxidoreductase 10.5 kDa subunit (PF05047)	AORF-Exon2(G4v2-117472)	GRMZM2G008464 (T02, P02)	8.03
	**ccr2, ubiquinol-cytochrome c reductase2*	Ubiquinol-cytochrome c reductase complex 8.0 kDa subunit (PF05365)	2-A5U(G4v2-27146, G4v2-27147)	GRMZM2G064896	2.05
	**cox6b, cytochrome-c oxidase subunit VIb*	Cytochrome oxidase c subunit VIb (PF02297)	A5U(G4v2-131796)	GRMZM5G815839 (T02, P02)	9.04
Energy signaling, TOR pathway	**raptor1, RAPTOR protein homolog1*	RAPTOR, TOR complex subunit (PF02985, PF00400)	2-A5U(G4v2-138708, G4v2-138709)	GRMZM2G048067	10.00

(*continued on next page*)

Table 1 (*continued*)

Functional class	Maize locus name[a]	Protein product[b] (PFAM domain)	Genic G4 motif class[c]	Gene model ID[d]	Genetic bin[e]
Energy signaling, AMPK/SnRK pathway	**snrk1a1, snf1-related kinase 1-like1*	Snf1-related AMPK, SnRK subunit (PF00069, PF07714, PF00627, PF02149)	A5U, SI1(G4v2-101507, G4v2-95276)	GRMZM2G077278	6.06
	**snrk1a2, snf1-related kinase 1-like2*	Snf1-related AMPK, SnRK subunit (PF00069, PF07714, PF02149)	SI10(G4v2-27350)	GRMZM2G180704	2.06
	**snrk1a3, snf1-related kinase 1-like3*	Snf1-related AMPK, SnRK subunit (PF04739)	A5U(G4v2-12761)	GRMZM2G138814	1.03
	**snrk1bc1, snf1-related kinase -like1*	AMPK, SnRK1 subunits with CMB48 and two CBS domains (PF00571)	2-A5U, A5I3(G4v2-21629, G4v2-21630, G4v2-21631)	GRMZM2G047774	1.12
	**snrk1bc2, snf1-related kinase -like2*	AMPK, SnRK1 subunits with CMB48 and two CBS domains (PF00571)	4-A5U(G4v2-73609, G4v2-73610, G4v2-73611, G4v2-73612)	GRMZM2G014170	5.00
	**snrkc2, snf4-related kinase 2*	SNF4-related AMPK, SnRK subunit 2 with two CBS domains (PF00571)	A5U(G4v2-47210)	GRMZM2G051764 (T03, P03)	3.09
Energy signaling, sucrose-regulated TF, bZIP11 family	**bzip11a, basic leucine zipper transcription factor 11a (ZmbZIP84)*	bZIP11 family TF with sucrose-regulated group 1 uORF transcript[g] (PF07716, PF00170)	AUG(G4v2-73521)	GRMZM2G361611	4.10
	**bzip11 b, basic leucine zipper transcription factor 11b (ZmbZIP60)*	bZIP11 family TF with sucrose-regulated group 1 uORF transcript (PF07716, PF00170)	2-A5U(G4v2-76359, G4v2-76360)	GRMZM2G444748	5.03
	**bzip11c, basic leucine zipper transcription factor 11c[i] (ZmbZIP12, lip15)*	bZIP11 family TF with sucrose-regulated group 1 uORF transcript (PF07716, PF00170)	A5U, A5I(G4v2-90387, G4v2-90386)	GRMZM2G448607	6.01
Oxidative stress response	*cat1, catalase1*	Catalase (PF00199, PF06628)	A5U(G4v2-83873)	GRMZM2G088212	5.03
Oxidative stress signaling	**dj1a1, DJ-1A GATase1-domains1*	DJ-1a/PARK7-like GATase1-like (PF01965)	A5U(G4v2-39277)	GRMZM2G102927	3.01
	**dj1a2, DJ-1A GATase1-domains2*	DJ-1a/PARK7-like GATase1-like (PF01965)	A5U(G4v2-67306)	GRMZM2G024959	4.05
	**dj1a3, DJ-1A GATase1-domains3*	DJ-1a/PARK7-like GATase1-like (PF01965)	A5U(G4v2-58314)	GRMZM2G127812	4.05
	**dj1c, DJ-1C GATase1-domains*	DJ-1c/PARK7-like GATase1-like (PF01965)	A5U(G4v2-18112)	GRMZM2G117189 (T02, P02)	1.07
Hypoxia transcription factor, group VII ERFs[h]	**hre1, hypoxia responsive ERF homologous1 (ZmEREB67)*	Hypoxia responsive ERF with AP2, CRIB domain, N-terminal MVLSAEI (PF00786, PF00847)	A5U(G4v2-137519)	GRMZM2G114820	9.04
	**hre2, hypoxia responsive ERF homologous2 (ZmEREB102)*	Hypoxia responsive ERF with AP2 domain, N-terminal MCGGAIL (PF00847)	A5U(G4v2-112559)	GRMZM2G052667	7.02
	**hre3, hypoxia responsive ERF homologous3 (ZmEREB202)*	Hypoxia responsive ERF with AP2 domain, N-terminal MCGGAIL (PF00847)	A5U(G4v2-37198)	GRMZM2G148333	2.06
	**hrap1, hypoxia responsive RAP2 homologous1 (ereb14)*	Hypoxia responsive ERF with AP2 domain and N-terminal MCGGAIL (PF00847)	A5U/TSS(G4v2-71198)	GRMZM2G018398	4.07
	**hrap2, hypoxia responsive RAP2 homologous2 (ZmEREB160)*	Hypoxia responsive ERF2 with AP2 domain protein and N-terminal MCGGAIL (PF00847)	3-A5U(G4v2-127886, G4v2-127887, G4v2-127888)	GRMZM2G171179	9.01

(*continued on next page*)

Table 1 (*continued*)

Functional class	Maize locus name[a]	Protein product[b] (PFAM domain)	Genic G4 motif class[c]	Gene model ID[d]	Genetic bin[e]
Transcription factor	*arf25, ARF-transcription factor 25 (ZmARF25)*	Auxin response factor (PF02362, PF06507, PF02309)	3-A5U(G4v2-120458, G4v2-120459, G4v2-120460)	GRMZM2G116557 (T02, P02)	8.06
	ca2p7, CCAAT-HAP2-transcription factor 27 (ZmCA2P7)	CCAAT box binding factor (PF02045, PF01406)	3-A5U(G4v2-149218, G4v2-149219, G4v2-149220)	GRMZM2G126957 (T02, P02)	10.06
Chromatin	*htr105, histone H3 105*	Histone H3.2 (PF00125)	A5U(G4v2-133864)	GRMZM2G130079	9.02
	htr103, histone H3 103	Histone H3.3 (PF00125)	A5U(G4v2-45648)	GRMZM2G078314	3.06
Chromatin	*chr112a, chromatin remodeling 112a*	SNF2 superfamily with HIRAN; DEAD box helicase, RING and RAD5 domains (PF08797, PF00176, PF00097, PF00271)	3-A5U(G4v2-67223, G4v2-67224, G4v2-67225)	GRMZM2G030768	4.05
Telomerase	*TERT, telomerase reverse transcriptase*	Telomerase, ribonucleoprotein complex-RNA binding domain; reverse transcriptase (PF12009, PF00078)	A5U, 2-AI5, AI6(G4v2-137652, G4v2-137653, G4v2-137654, G4v2-137655)	GRMZM2G167338	9.05
DNA repair, base excision repair (BER) pathway[j]	**ogg1, 8-oxoguanine DNA glycosylase*	N-glycosylase/DNA lyase; OGG1-like (PF07934, PF00730)	A5U(G4v2-79867)	GRMZM2G139031	5.04
	**endo3, endonuclease III*	Endonuclease III (PF00730, PF00633)	SORF-exon7(G4v2-138855)	GRMZM2G113228	10.02
	**dmag2a, DNA-2-methyladenine glycosylase II A*	DNA-3-methyladenine glycosylase II (PF00730)	A5U/TSS(G4v2-95442)	GRMZM2G117574	6.07
	**dmag2b, DNA-2-methyladenine glycosylase II B*	DNA-3-methyladenine glycosylase II (PF00730)	A5ORF(G4v2-101766)	GRMZM2G114592	6.07
	**dmag1, methyladenine glycosylase I*	DNA-3-methyladenine glycosylase I (PF03352)	A5U(G4v2-99259)	GRMZM2G171317	6.02
	**udg, uracil-DNA glycosylase*	Uracil-DNA glycosylase (PF03167)	A5U, A5ORF(G4v2-30578, G4v2-30579)	GRMZM2G040627	2.01
	hmg1, high mobility group protein1	High mobility group protein B1 (PF00505)	A5U(G4v2-75909)	GRMZM5G834758 (T03, P03)	5.03
	pcna1, proliferating cell nuclear antigen1	Proliferating cell nuclear antigen (PF00705, PF02747)	A5U, AP-140(G4v2-89286, G4v2-89287)	GRMZM2G030523	5.08
	dpole2, DNA polymerase epsilon subunit 2	DNA polymerase epsilon subunit 2 (PF12213, PF04042)	A5U/TSS(G4v2-18092)	GRMZM2G154267	1.06
	**dpold3a, DNA polymerase delta subunit 3 locus a*	DNA polymerase delta subunit 3 (PF00281, PF00673, PF09507)	AORF-exon6(G4v2-39377)	GRMZM2G435338	3.02
	**dpold3b, DNA polymerase delta subunit 3 locus b*	DNA polymerase delta subunit 3 (PF09507)	A5U(G4v2-115333)	GRMZM2G005536	8.03
	**dpold3c, DNA polymerase delta subunit 3 locus c*	DNA polymerase delta subunit 3 (PF09507)	A5U(G4v2-102166)	GRMZM2G148626 (T02, P02)	7.01

[a] Maize Locus Name; current gene name from MaizeGDB (MaizeGDB.org) annotation, listed as the short name followed by the full name. Genes named anew in this study are indicated with an asterisk preceding the short name. Synonyms are given for some loci/gene names in parentheses following the full name. Some genes are designated using suggested nomenclature from specialized databases, including those curated by "ChromDB" (the chromatin database, chromdb.org) for chromatin remodeling gene (prefix "*chr*") and histone H3 (prefix "*htr*") and those for plant transcription factors, curated by "Grassius" (grassius.org, Gray et al., 2009) for ethylene response element binding protein (prefix "*ZmEREB*"), basic region leucine zipper (prefix "*ZmbZIP*"), auxin response factor (prefix "*ZmARF*" or "*arf*") and CCAAT-HAP2 family (prefix "*ZmCA2P*" or "*ca2p*"). The bzip11c locus is also known to encode the gene known as *lip15, low temperature induced transcription factor15*; [b] The column named "Protein Product" refers to the name of the encoded gene product or enzyme known or predicted, along with the PFAM domain identifier in parenthesis, as reported from MaizeGDB. Predicted proteins are all from the only or 1st transcript model (T01, P01), unless otherwise designated. Abbreviations used in this column in order of appearance: PF/PFAM, protein family database (http://pfam.sanger.ac.uk/, Punta et al., 2011); NADPH, the

reduced form of nicotinamide adenine dinucleotide phosphate; PPP, pentose phosphate pathway; NADH, the reduced form of nicotinamide adenine dinucleotide; RAPTOR, regulatory-associated protein of mTOR (mammalian target of rapamycin); TOR, target of rapamycin; AMPK, 5′ adenosine monophosphate-activated protein kinase; SnRK, SNF-(sucrose non-fermenting)-related serine/threonine-protein kinase; CBM48, carbohydrate binding module 48, a domain that is a member of the N-terminal Early set domain, a glycogen binding domain associated with the catalytic domain of AMP-activated protein kinase beta subunit; CBS, cystathionine beta synthase domain (also known as Bateman domain) that regulates the activity of associated enzymatic and transporter domains in response to binding molecules with adenosyl groups, AMP and ATP, or s-adenosylmethionine; bZIP11, basic leucine zipper transcription factor of the bZIP11 family of the group S bZIP TFs (Jakoby et al., 2002); DJ-1/PARK7, oncogene DJ1/Parkinson Disease 7; GATase1; type 1 glutamine amidotransferase; ERF; ethylene response factor; AP2, APETALA2; CRIB, Cdc42/Rac interactive binding; SNF2, sucrose non-fermenting 2; HIRAN, HIP116, Rad5p N-terminal; RING, really interesting new gene; RAD5, radiation sensitive 5; OGG1, 8-oxoguanine DNA glycosylase 1; [c] Genic G4 motif classes are listed as defined in this study (Fig. 2). Multiple elements in the same class are listed by leading number and dash. For example 2-A5U indicates two G4 motifs within the Antisense 5′ UTR. Classifier abbreviations: A5U, antisense strand 5′ UTR; A5U/TSS, A5U/overlapping with the transcription start site; A5I1, antisense 5′ end of Intron 1; AUG, sense strand near the "AUG" start codon; S3U, sense strand 3′ UTR; AI, antisense strand non-first intron followed by the intron number; AORF, antisense ORF followed by the exon number for intron-containing genes; SI, sense strand intron followed by intron number; AP, antisense promoter followed by a dash and the number of bases upstream of the TSS. The name(s) of each G4 motifs is listed in parentheses; [d] Gene model ID: From maizesequence.org, from the filtered gene set of B73 line of maize. For genes with multiple transcript models, predicted proteins are all from the only or 1st transcript model (T01, P01), unless otherwise designated; [e] Genetic BIN refers to the chromosome number followed by the genetic linkage bin as defined by Davis et al. (1999) for the map UMC 98. Current markers delineating the genetic bins of maize are at MaizeGDB; [f] The protein encoded by the *sut1* locus lacks a PFAM domain but does contain the conserved domain cd06174; MFS, The Major Facilitator Superfamily (from Conserved Domain Database, CDD, at NCBI, http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml, Marchler-Bauer et al., 2013); [g] Group 1 refers to the uORF-containing "homology group 1" bZIP genes (Hayden and Jorgensen, 2007); [h] The group VII ERFs (from Nakano et al., 2006) include hypoxia responsive ERF TFs (loci/gene name "*hre*") and hypoxia responsive related to AP2 (RAP2.2/RAP2.12) TFs (loci/gene name "*hrap*"). The *hre* and *hrap* genes listed here fall in the AP2-EREBP (APETALA2 – ethylene response element binding protein) group as described by Gray et al. (2009) and curated by the grass TF database, grassius.org. The recommended grassius gene names are listed here as synonyms in parentheses; [i] The locus/gene *bzip11c* is also known as *lip15*, *low temperature induced protein15*; [j] Genes from the base excision repair pathway from the KEGG pathway zma03410 (KEGG, http://www.genome.jp/dbget-bin/www_bget?zma03410) are included here because of their reported role in G4-associated hypoxia-induced transcription (Clark et al., 2012).

Table 2
Antisense G4 motifs in select metabolic pathways

MaizeCyc pathway[a]	G4 motif relative abundance and significance[b]					
	Gene regions[c]			Whole gene[d]		
	A5U-List 1	A5U-List 2	A5I1-List 3	Both strands	Sense	Antisense
1. Aerobic respiration – electron donor II	2.0 (*p*, 9.7E-2)	1.2 (*p*, 4.5E-1)	2.4* (*p*, 2.6E-2)	0.6** (*p*, 6.8E-5)	0.3** (*p*, 1.1E-4)	0.8 (*p*, 2.4E-1)
2. Glycolysis III	1.1 (*p*, 7.8E-1)	1.6* (*p*, 1.2E-2)	1.8 (*p*, 2.1E-1)	0.9 (*p*, 1.6E-1)	0.8 (*p*, 1.5E-1)	1.0 (*p*, 8.7E-1)
3. Superpathway of cytosolic glycolysis (plants), pyruvate dehydrogenase and TCA cycle	1.6 (*p*, 1.8E-1)	1.7** (*p*, 5.3E-4)	1.5 (*p*, 3.1E-1)	0.9 (*p*, 2.1E-1)	1.1 (*p*, 5.2E-1)	0.8 (*p*, 1.8E-1)
4. Sucrose degradation III	5.1** (*p*, 1.8E-4)	1.4 (*p*, 2.5E-1)	1.6 (*p*, 6E-1)	1.3 (*p*, 9.0E-2)	1.2 (*p*, 3.1E-1)	1.6 (*p*, 5.2E-2)
5. 1D-myo-inosoitol hexakisphosphate biosynthesis V (from Ins(1,3,4)P3)	7.6* (*p*, 2.8E-2)	4.5** (*p*, 1.1E-3)	0.0 (*p*, 1.0)	1.9* (*p*, 2.4E-2)	2.1 (*p*, 8.3E-1)	2.0 (*p*, 5.6E-2)

[a] MaizeCyc pathways 1–5 are the same as those shown in Fig. 4C, and are from the Metabolic Pathways in Maize (MaizeCyc 2.2, http://maizecyc.maizegdb.org/); [b] G4 motif fold enrichment (>1.0) or depletion (<1.0) values were calculated as the ratio of observed (numerator) relative to the predicted (numerator) number based on the regional average for the filtered gene set. Fisher's exact test *p*-values, assuming the null hypothesis, are given in parentheses. Measurements are marked for significance levels of *p*-value 0.05 (one asterisk) or *p*-value 0.01 (two asterisks); [c] Gene regions: A5U, antisense 5′ UTR; A5I1, antisense 5′ end of intron 1; as defined in Fig. 2 and recorded in Tables S3 and S4; [d] Whole gene regions span the entire gene model or the designated canonical transcript for the genes with multiple transcript models.

possible relationships between inositol metabolism and plant hypoxia, especially in endosperm. We also note, however, that all G4 motifs need not be associated with a singular or shared function. In fact, given the myriad processes templated by nucleic acids, it is possible that we have detected some pathways merely coincidentally linked by shared G4 motifs.

To further explore the possible link between hypoxia-responsiveness and G4 motif-containing genes, we cross-referenced our gene lists with those from transcriptomic analysis of hypoxic responses (Bailey-Serres et al., 2012a) and those from a recent study showing that hypoxia triggers acquisition of male germ-cell fate (Kelliher and Walbot, 2012), and found that genes associated with hypoxia were more likely to have G4 motifs than were other genes, with a 1.5-fold enrichment for genes in our A5U-List 1 group. In addition to hypoxia-associated genes (Licausi et al., 2011), we found an intriguing abundance of G4 motifs, especially the A5U type, in signaling genes associated with energy homeostasis. Due to the overlapping nature of low-oxygen and low-sugar signaling (Zeng et al., 1998; Koch et al., 2000), as well as their response pathways, we included many of these genes in Table 1. For instance, many genes associated with the TOR, AMP kinase (AMPK)/Snf-related kinase (SnRK), and oxidative stress signaling (DJ-1/PARK7) pathways have one or more G4 motifs (Table 1), possibly signifying a shared capacity for G4 motifs to aid the regulatory, signaling, and metabolic adjustments to energy status (Xu et al., 2010; Robaglia et al., 2012; Dobrenel et al., 2013). In addition to energy-state-responsive genes, we also found G4 motifs associated with genes for base and excision repair pathways (Table 1). Interestingly, a role for G4s and DNA repair in redox-associated transcriptional regulation has recently been described for specific genes in humans (Ruchko et al., 2009; Clark et al., 2012). The functional grouping of genes with G4 motifs suggests the possibility that these sequences may provide a previously unrecognized link connecting several different gene response pathways in maize.

Oligonucleotides with maize G4 motifs can form quadruplex structures *in vitro*

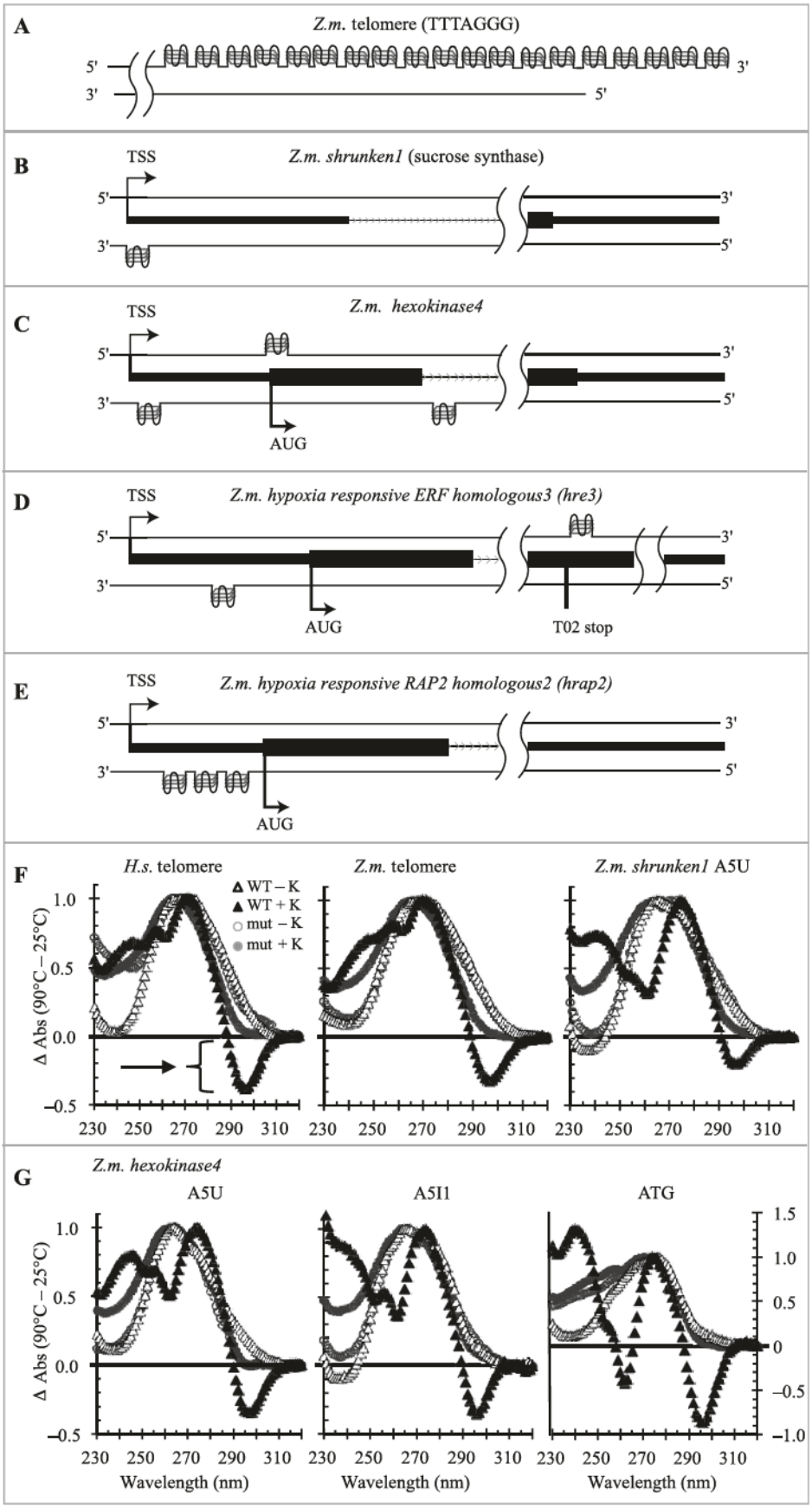
We tested several oligonucleotides with genic G4 motifs to see whether they could fold into G4 structures *in vitro* as summarized in Fig. 5A−E. Single-stranded synthetic oligonucleotides were incubated under G4-forming conditions, and thermal difference spectra showed a diagnostic G4-specific increase in absorption at 295 nm (Fig. 5G−F). This A_{295} signature was observed in a positive control sample of human telomere repeat DNA, $(TTAGGG)_4$, and also in a plant telomere oligonucleotide sample $(TTTAGGG)_4$, under the same conditions (Fig. 5F). The locations of several genic G4 motifs in the *shrunken1* and *hexokinase4* genes are shown. Oligonucleotides with mutations that altered the G-tracts ("mut" in Fig 5.), or wild-type oligonucleotides in the absence of potassium, failed to show spectra indicative of G4 structures in these assays. These results were corroborated using circular dichroism spectroscopy (data not shown). Together, these results confirmed our expectation that computationally-predicted G4 motifs can adopt quadruplex structures *in vitro*.

DISCUSSION

Computational prediction and analysis of G4 motifs from the genome sequence of a major crop species reveals that these elements in maize are widespread, predominantly located in genes, and present in those with shared functions in hypoxic responses, energy metabolism, and inositol phosphate metabolism. Using the Quadparser algorithm, we have named and classified nearly 150,000 G4 motifs, collectively covering over 2.22 Mb of the maize genome. The number of G4 motifs found in the repeat-masked region of the maize genome (~43,000) is similar in scale to the number ~40,000, found in rice (Takahashi et al., 2012). If these serve as functional DNA elements in plants and, more broadly, in all eukaryotic species, then the G4 motifs can be regarded as one of the largest groups of cis-acting genetic motifs known. They would thus be comparable to other major genic cis-acting elements including general transcriptional initiation motifs (TATA boxes) and RNA processing signals (splicing elements, polyA addition signals). This study, together with that of Takahashi et al. (2012), establishes G4 motifs as a major and largely uncharacterized entry in the encyclopedia of DNA elements in plants. The relative paucity of G4 motifs in *Arabidopsis* (Takahashi et al., 2012) is intriguing, possibly reflecting the reduced capacity of this species to resolve non-telomeric G4 motifs in comparison to maize or rice.

Systematic examination of locations for G4 motifs in maize genes revealed that for any given gene, there were two major hotspots at sites we designated A5U and A5I1. Together these two positions accounted for more than 90% of all the gene-associated G4 motifs. Given the locations of the A5U and A5I1 elements relative to TSSs and the first exon-intron boundary, we speculate a role for these maize G4 motif elements in transcription. More specifically, we find compelling the prediction that they impact the processivity of RNA polymerases. We find intriguing and unexpected the fact that maize and humans share G4 hotspot locations, but on opposite strands. Specifically, in humans, most of the TSS and 1st-intron G4 motifs are on the sense (coding/non-template) strand (summarized in Maizels and Gray, 2013), whereas in maize, the TSS- and 1st-intron-associated G4 motifs are almost entirely on the antisense (template) strand (Fig. 2). The biological significance of this difference is not known, but may reflect a divergent evolutionary deployment of G4 motifs for genetic functions in different phylogenetic taxa. These G4 motifs could also provide strand-independent cis-regulatory functions as structure-specific nucleation or recruitment sites for molecules involved in other processes such as DNA replication, strand-unwinding, DNA repair, epigenetic marking, chromatin-remodeling, or RNA metabolism.

A primary expectation at the outset of this study was that the identification of maize G4 motifs could reveal genes with shared regulation and function. Initially prompted by the detection of a G4 motif in the promoter of a maize sucrose


A
Z.m. telomere (TTTAGGG)
5'
3'
3'
5'
B
TSS
Z.m. shrunken1 (sucrose synthase)
C
TSS
Z.m. hexokinase4
AUG
D
TSS
Z.m. hypoxia responsive ERF homologous3 (hre3)
AUG
T02 stop
E
TSS
Z.m. hypoxia responsive RAP2 homologous2 (hrap2)
AUG
F
H.s. telomere
Z.m. telomere
Z.m. shrunken1 A5U
WT – K
WT + K
mut – K
mut + K
Δ Abs (90°C – 25°C)
1.0
0.5
0.0
–0.5
230 250 270 290 310
G
Z.m. hexokinase4
A5U
A5I1
ATG
1.5
1.0
0.5
0
–0.5
–1.0
Wavelength (nm)

synthase gene, *shrunken1*, we extended analyses to the entire genome. Metabolic pathway analysis, summarized in Fig. 4, showed a striking pattern of G4 motifs common to genes encoding key enzymes in energy metabolism. Prominent among these were enzymes for the altered carbohydrate metabolism and redox reactions that would be important for plant cell acclimation under hypoxia. Such hypoxic conditions can occur naturally, for example in some cells of phloem, endosperm, and anthers, or as a result of environmental conditions such as flooding (Rolletschek et al., 2005; Rolletschek et al., 2011; Bailey-Serres et al., 2012a, 2012b). Hallmarks of metabolic adjustments for hypoxic survival summarized in recent studies and reviews include substrate-level ATP production, increased glycolytic flux, tight regulation of mobile reductants (malate, oxoglutarate, and oxaloacetate) and the integration of sugar-signaling pathways (Bouche and Fromm, 2004; Foyer and Noctor, 2005, 2011; Bailey-Serres and Voesenek, 2008; Robaglia et al., 2012; Bailey-Serres et al., 2012a; Dobrenel et al., 2013). Remarkably, many of these same pathways were highlighted when G4 motif-containing genes were visualized using the omics viewer tools of the MaizeCyc database (Fig. 4 and Table 1). Associations persisted when we examined a more specific subset of genes including transcription factors involved in signaling of hypoxic and low-sugar status (Table 1).

In considering the two bodies of information discussed here—the G4 motif genic location non-random hotspots, and the functional classification of genes that contain them, we propose a generalized model in Fig. 6 to integrate our findings and suggest mechanistic hypotheses for future investigations. We have documented the prevalence of G4 motifs in many genes responsive to extreme energy states. From this, we predict that G4 motifs in maize may represent a common functional component in genetic responses to multiple overlapping signals of energy status (Fig. 6A). In support of this idea are the observations that G4 motifs are found in numerous genes encoding members of the AMPK/SnRK, NrF2-realated, and hypoxia-responsive transcription factors (Table 1). Furthermore, these signaling pathways affect downstream regulation of genes coding for core metabolic enzymes that produce ATP under low-oxygen or low-sugar conditions.

A simple mechanistic model for the possible function of G4 DNA in gene expression is summarized (Fig. 6B) for both transcription and translation. In this model, the physical impediment or "kink" in the template caused by the G4 could reduce or block RNA or protein synthesis along the G4-containing template. Resolution of the "kink" could then occur through the activity of trans-acting factors that respond to signals of energy status. Findings that support this sort of idea include transcriptional roles for G4 DNA in several animal genes, including *c-Myc*, *KRAS*, and *VEGF* (Cogoi and Xodo, 2006; Breit et al., 2008; Brown et al., 2011b). These data, together with findings from other studies, including those using polymerase stop-assays, were taken into consideration in developing this model (Sun and Hurley, 2010). There is little doubt that the mode of action may be more complex and multifactorial than a simple G4-kink-based block that is resolved for an increase gene expression. For instance, mechanistic models from research on G4 elements in animal gene promoters invoke interactions of multiple trans-acting factors in and around the quadruplex (reviewed in Brooks et al., 2010). For these reasons, we include the concept of "licensing" G4 motif-containing genes for adjustments coupled to extreme energy states (Fig. 6B). Direct resolution of G4s would be only one of several potential molecular mechanisms. More detailed molecular or biochemical analysis will be required to test aspects of this model and define additional contributing mechanisms. Evidence in support of the existence of G4-binding protein in maize comes from a recent G4 ligand-binding screen in which maize *nucleoside diphosphate kinase1* (*ZmNDPK1*) was discovered to exhibit high affinity G4 DNA binding activity *in vitro* (Kopylov, Bass, and Stroupe, unpublished observations) to a G4 DNA element from the A5U region of the maize *hexokinase4* gene (motif G4v2-53046, Table 1).

A particularly intriguing finding was the occurrence of A5U-type G4 motifs in maize homologs of hypoxia-responsive transcription factors belonging to the HRE and RAP2 group VII ethylene response factors (Nakano et al., 2006). These plant-specific, oxygen-sensing transcription factors were recently found to be stabilized under low-oxygen conditions *via* a Cys-dependent, N-end turnover pathway (Gibbs et al., 2011; Licausi et al., 2011). Once stabilized, they are proposed to enter the nucleus and mediate global transcriptional responses. Our findings point to a possible *cis*-acting element with which the HRE and HRAP2

Fig. 5. Location and folding of select G4 motifs.

Each schematic shows 750 bases after the TSS and the last 250 bases of the canonical transcript for each given gene model. The arrow on top strand denotes the TSS, arrow on bottom strand denotes AUG, location of G4 motifs are depicted as a three-sheet stack on the appropriate strand, coding regions are wide black boxes, UTR regions are narrow black boxes, and introns are arrowed lines. **A:** Maize telomere (TTTAGGG repeat); **B:** *shrunken1*, sucrose synthase with an A5U overlapping the TSS; **C:** Maize *hexokinase4*, a hexokinase domain protein with three quadruplexes: A5U near the TSS, AUG overlapping the start codon, and a A5I1 in the first intron; **D:** Maize *hre3* (GRMZM2G148333, *hypoxia responsive ERF homologous3*, Table 1) with one A5U G4 motif and one G4 motif on the template strand immediately after an alternative transcript coding stop site; **E:** Maize *hrap2* (GRMZM2G171179, *hypoxia responsive RAP2 homologous2*, Table 1), with three tandem A5U G4 motifs (a quadruplex patch; see Table S2 for a list of all patches) between the TSS and start codon. **F:** Normalized UV absorbance thermal difference spectra for selected synthetic oligonucleotides in human telomeric repeat, maize telomeric repeat, and *shrunken1* A5U. Human telomeric repeat was used in this experiment as a positive control. G4-characteristic TDS profile and prominent negative peak at A295 were obtained only for WT sequences annealed in the presence of 100 mmol/L potassium (filled triangles) but not for those annealed in TBA phosphate buffer alone (open triangles). Mutant sequences (mut) did not show G4-characteristic signature. The G4-characteristic signature is denoted by arrow and bracketed region for the first panel, *H.s.* telomere, and seen in other panels for the "WT + K" samples (filled triangles). **G:** Normalized UV absorbance thermal difference spectra for three different G4 motif oligos in maize *hexokinase4*.

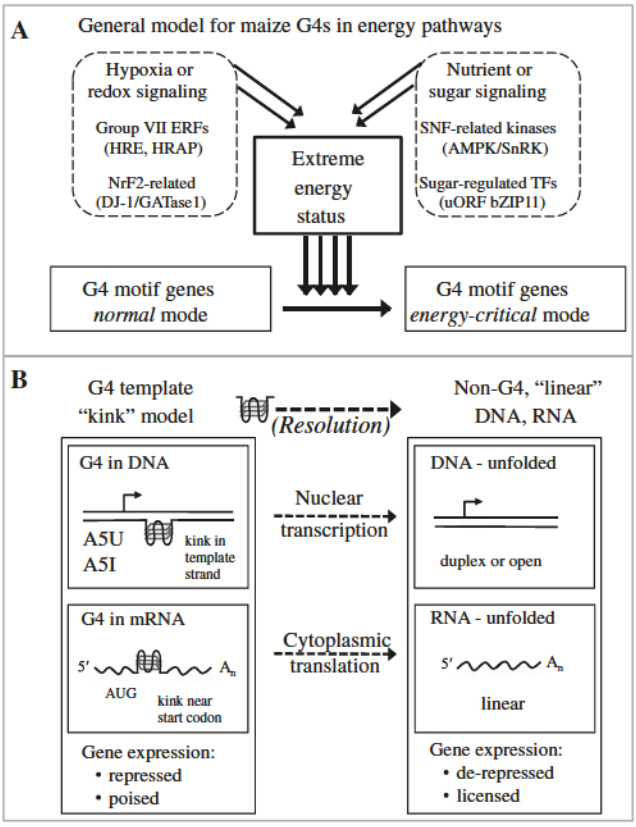


Fig. 6. Models for role of G4 in expression of genes responsive to extreme energy states.

In these models, various and primary signals of energy status are transduced *via* G-quadruplexes that affect the expression of their resident genes. The G4 motifs are depicted as a shared genetic component in genes that can respond to low-oxygen (hypoxia), low sugars (starvation/famine), or other signals (redox and nutrient balance). **A:** A generalized model with several different causes of energy extremes (dashed boxes) and their signaling pathway components (parentheses). Depicted here is the transduction of extreme energy status signals to nuclear G4 motif genes *via* unknown trans-acting factors. The prevalence of G4 motifs in the genes (in parenthesis) for the signaling pathway proteins implies auto-regulatory or feedback control (not depicted). In this general model, the G4 motif confers a capacity for prioritizing gene function in response to energy status information. **B:** A simple mechanistic model (G4 template "kink" model) depicting how G4 elements might influence gene expression levels on the basis of the position and strand of the G4 motifs classified as A5U-, A5I-, or AUG- (defined in Fig. 2). Transcription and translation are depicted separately, but both illustrate the concept of the G4 as a template strand physical impediment. For transcription, the G4 would prevent, reduce, or stall RNA polymerase (any RNA pol) activity at a G4 "kink" structure in the template. Similarly, for translation, the G4 would prevent, reduce, or stall ribosomal protein synthesis activity at a G4 "kink" structure in the template. "*Resolution*" denotes any processes that resolves, "unkinks", or unwinds G4 DNA or RNA into a structure more favorable for expression, up-regulation, or transcriptional licensing of the associated gene. The "*resolution*" factors are not presumed to be the same for transcription and translation. Other modes of G4 action known to occur in nature could also act in maize, but are not depicted in this model.

transcription factors may interact to modulate expression of specific sets of target genes. This situation is analogous to the regulation of a different, hypoxia-response transcription factor in animals, HIF-1α. This gene is also is stabilized under low-oxygen conditions, but by a different mechanism involving an oxygen-sensing prolyl-hydroxylase (Epstein et al., 2001; Ivan et al., 2001; Jaakkola et al., 2001).

Our study also points to possible roles for G4 motifs in genes for biosynthesis of *myo*-inositol and phytic acid metabolism, and in genes for SnRK signaling pathways (Fig. 4, pathways 5 and 6). We speculate that conservation and positioning of these G4 motifs are not coincidental, but rather represent functional commonalities among these pathways. Two important ways in which G4 motifs in these genes might aid survival are through direct metabolic effects, or through effects on signaling systems that ensure low-energy metabolic programs are operative. First, recent evidence indicates an increasingly close relationship between myo-inositol metabolism and stresses involving cellular energy status (Bailey-Serres and Voesenek, 2008; Bailey-Serres et al., 2012a). Levels of this sugar-alcohol rise during hypoxia, nutrient deprivation, and perturbations of related sensing systems (Valluru and Van den Ende, 2011; Robaglia et al., 2012; Dobrenel et al., 2013). The same stresses also increase levels of the myo-inositol metabolites, galactinol, raffinose, and phytate (Sulpice et al., 2010; Dobrenel et al., 2013; Wouters et al., 2013), and likely also involve the phosphatidyl inositols (Sulpice et al., 2010; Valluru and Van den Ende, 2011; Dobrenel et al., 2013). Suggested stress roles of raffinose, galactinol, and phytate include a possible parallel to their well-known storage functions in seeds (Valluru and Van den Ende, 2011).

However, other roles may be more important to survival during starvation or hypoxia. Raffinose and galactinol, for example, can also arise during remobilization of starch or cell wall polysaccharides (Valluru and Van den Ende, 2011), and further, can scavenge reactive oxygen species *in vitro* and *in vivo* (Valluru and Van den Ende, 2011). In addition, sugar-alcohol-forming reactions can serve as sinks for excess reductant (Nishizawa et al., 2008; Keunen et al., 2013), and biosynthesis of myo-inositol can compete with starch formation in a potentially advantageous way under stress (Loescher, 1987; Shen et al., 1999; Sickler et al., 2007). The non-storage, direct metabolic roles of phytate also affect the balance of Pi, PPi, and adenylates (Sulpice et al., 2010; Dobrenel et al., 2013), which would be especially important for low-oxygen or starvation conditions as well as sieve tube or PPi-dependent tonoplast functions (Huber and Akazawa, 1986; Koch, 2004; Bailey-Serres and Voesenek, 2008; Fukao et al., 2011; Gibbs et al., 2011; Valluru and Van den Ende, 2011; Bailey-Serres et al., 2012a).

Second, new research is revealing an even greater importance for the rapid, indirect action of myo-inositol metabolism in signaling systems. Previous work demonstrated diverse roles for myo-inositol-derived signaling molecules that include inositol tri-phosphate ($Ins(1,4,5)P_3$) and phosphatidyl-inositol cascades (Ferjani et al., 2011). In addition, raffinose synthase can facilitate sucrose metabolism and myo-inositol cycling without production of hexoses (Valluru and Van den Ende, 2011), an advantageous pathway under stress, because hexose-based signals of carbohydrate abundance could otherwise counter adjustments to their limited supply (Dobrenel

et al., 2013). Recently, myo-inositol metabolism has been recognized as a hub of interaction between two major sensing systems for nutrient and energy status of cells (Koch, 1996; 2004; Ruan et al., 2010; Tiessen and Padilla-Chacon, 2012). The TOR and SnRK complexes, respectively, sense nutrient abundance and deprivation (Dobrenel et al., 2013; Robaglia et al., 2012), thus mediating feast and famine responses from cellular to whole organism levels (Robaglia et al., 2012; Caldana et al., 2013; Dobrenel et al., 2013; Xiong et al., 2013). The responsiveness of both systems is vital to the adjustment of metabolism for survival of the changing cellular energy status (Koch, 1996, 2004; Ruan et al., 2010; Tiessen and Padilla-Chacon, 2012; Bihmidine et al., 2013). Both are also integrally involved in responses to hypoxia and nutrient deprivation (Dobrenel et al., 2013). In this regard, sugar-modifying enzymes are not viewed as merely metabolic substrates, but also as key signaling molecules, possibly requiring stringent control.

The G4 DNA elements are also known to affect telomere structure or maintenance for many eukaryotes. Telomere length is affected by hypoxia, and known to up-regulate the TERT subunit of telomerase in eukaryotes from fishes to humans (Yatabe et al., 2004; Yu et al., 2006). Interestingly, in humans, the *hTERT* gene contains G4 elements (Palumbo et al., 2009). Analysis of the maize *TERT* gene (GRMZM2G167338) shows that it exists as a single copy in the maize genome, with several G4 motifs near the promoter and in the introns (Table 1). This intriguing similarity points to a possible connection between G4 motifs, telomerase regulation, telomere metabolism and hypoxia. In plants, these relationships are further evidenced by the discovery of hypoxia as inductive to meiosis and germ-cell fate specification, a developmental time frame during which telomerase is active and genotype-based telomere length may be reset (Brown et al., 2011a; Kelliher and Walbot, 2012).

Taken together, these data support a new hypothesis for functional significance of G4 motifs in maize and potentially elsewhere. Analysis of G4 motif distribution revealed their widespread association with genes for adjustment of cellular energy status, especially under extreme conditions. On the basis of these observations, we propose that G4 motifs reflect a previously unrecognized but widely conserved set of DNA sequences that could function in regulating genes central to the energy state of the cell.

MATERIALS AND METHODS

The maize genome, gene models, and syntenic grass gene lists

Maize reference genome

The maize inbred line B73 genome sequence was obtained from GenBank Accessions CM000777−CM000786 and GK000031−GK000034 (Schnable et al., 2012), as an annotated assembly (B73 RefGen_v2: Release 5b.60) from MaizeGDB http://www.maizegdb.org (Sen et al., 2009); it consists of 10 pseudomolecule chromosomes (Chromosome 1−Chromosome 10), 2 plasmid genomes (chloroplast and mitochondria), and a set of unmapped contigs (UNMAPPED). The 10 pseudomolecule chromosomes and unmapped contigs comprise approximately 2.07 billion base pairs, including 25.8 million unsequenced gap spaces, designated as "N"s and collectively constituting 1.2% of the genome sequence build.

Repeat-masked genome

The non-repetitive portion of the maize genome was analyzed as a "masked" subset of the genomic data from http://gramene.org/Zea_mays/Info/Index (Schnable et al., 2009; Youens-Clark et al., 2011). The mask was defined by the Munich Information Centre for Protein Sequences Repeat Catalog (MIPS/REcat) repeats (Youens-Clark et al., 2011) from http://mips.helmholtz-muenchen.de/plant/and the Transposable Elements (TE) Consortium repeats, including LTR exemplars (Nussbaumer et al., 2013). Collectively, the masked repetitive regions account for approximately 83% of the maize genome. The remaining portion (358 Mb in total), referred to in this study as the "repeat-masked genome," was used to limit some analyses to the low copy, gene-rich portion of the maize genome.

Gene models

The B73 RefGen_v2 maize reference genome includes two sets of predicted gene models. The working gene set (WGS) consists of 109,704 non-overlapping candidate gene elements, produced by the union of genes from two different gene finding approaches − evidence-based gene predictions using GeneBuilder (SanMiguel et al., 1998; Schnable et al., 2009) and *ab initio* gene prediction models using FGENESH (Milanesi et al., 1999; Salamov and Solovyev, 2000). The filtered gene set (FGS) is a 39,570 gene model subset of the WGS that excludes transposons, pseudogenes, contaminants, and other low-confidence annotations. For gene models with more than one transcript model, the transcript with the largest coding sequence, i.e., the 'canonical' transcript model, was analyzed or tabulated. Consequently, some G4 motifs may be associated with two or more different structures (intron *versus* exon, for example) for genes with multiple transcript models.

Maize syntenic orthologs

Syntenic orthologs (also called syntelogs) of maize genes were analyzed from four grass (*Poaceae* family) species: *Sorghum bicolor* (sorghum), *Setaria italica* (foxtail millet), *Oryza sativa* (rice), or *Brachypodium distachyon* (purple false brome or brachy). This set of pan-grass syntenic orthologs, compiled as described by Schnable et al. (2012), were obtained as gene model transcript sequences from Phytozome at http://www.phytozome.net (Goodstein et al., 2012). The numbers of gene models from this set were 25,507 from *S. bicolor* (release 79), 23,216 from *S. italica* (release 164), 66,338 from *O. sativa* (release 193), and 31,229 from *B. distachyon* (release 192). In this study, maize genes were called "conserved" only if they had at least one designated syntelog in any of the other four grass species, whereas they were called "highly-conserved" maize genes if they had designated syntelogs in all four of the other grass species.

G4 quadruplex motif nomenclature

In this study, G4 motifs in maize were defined as contiguous single-stranded sequences that match the $G_3L_{1\text{-}7}$ pattern, sometimes referred to as the default or strict G4 motif formula, using the Quadparser software (Huppert and Balasubramanian, 2005). In this formula, $G_{3+}L_{1\text{-}7}G_{3+}L_{1\text{-}7}G_{3+}L_{1\text{-}7}G_{3+}$, the "3+" refers to 3 or more sequential Gs and "L" refers to loops of any nucleotide length from 1–7. The resulting 149,520 non-overlapping maize G4 motifs were assigned unique names ranging from G4v2_1 to G4v2_149520. Each G4 motif site spans from the first to the last G and is defined by its assignment to a chromosome, with designated start and stop chromosomal coordinates and strand, plus(+) or minus(−) (149,520 G4 motifs listed in Table S1). For each of the chromosomes, the plus strand refers to the "top strand" which has its 5′ end in the short arm displayed at left in browsers and at top for maize karyotype diagrams. The G4 motifs were numbered sequentially one chromosome at a time, starting with the top strand, in the short to long arm direction, and ending with bottom strand, also in the short to long arm direction. The "strand" assignment is relative to the chromosomal strand, not relative to the transcript direction. To distinguish those, gene-associated strand assignments were designated sense (coding) or antisense (template) such that a plus-stranded G4 motif might also have a gene-related designation of sense or antisense strand.

A large number of G4 motifs were found to occur in clusters on the same strand. We designated these "G4 patch motifs", defined by us as requiring the first base of three or more G4 motifs to reside on the same strand within a 300 bp window (6911 G4 motif patches listed in Table S2). G4 motifs associated with genes are listed and annotated in Table S3. The G4 motifs from the gene lists (List 1–List 5, described in the Results) are provided in Table S4.

Analysis of observed and simulated G4 motif distributions relative to gene structure

The relative distributions of G4 motifs were characterized in two ways. First, they were mapped by assignment to six bins defined in relation to the transcription start site (TSS, +1 bp) for each gene model using the designated canonical transcript. These six bins included the TSS-relative bp segments at −1000 to −301, −300 to −101, −100 to −1, +1 to +100, +101 to +300, and +301 to +1000. The G4 motif density was calculated as the total number per Mbp. Second, the G4 motif distributions were quantified in relation to gene structural boundaries and strand (sense *versus* antisense), using windows centered at the boundary of interest: the TSS (−1//+1), the start codon (5′ UTR//AUG), or the first exon-intron junction (3′ end of exon1//5′ end of intron 1). Regions with sequence assembly gaps or unresolved BAC boundaries were excluded from the analysis. Each base position overlapping with a G4 motif was incremented by +1 for plotting (Fig. 2).

For randomized-sequence gene model simulations, a 40,000-gene randomized set of sequence simulations was produced for each of the three boundaries by sampling from the position-specific base plus strand frequency tables produced from the empirical dataset of *Zea mays* B73 RefGen_v2 Filtered Gene Set (downloaded from Phytozome 9 Genomes from Phytozome Biomart). In this way, the base composition was preserved in relation to strand plus location in the randomized sequence simulations. The Quadparser program (Huppert and Balasubramanian, 2005) was used to analyze the 40,000 random gene-simulation sequences using the same parameters as those for the maize genome to allow side-by-side plotting (Fig. 2) of the abundances in observed *versus* predicted G4 motifs.

Metabolic pathway analysis

Genes with G4 motifs were examined for possible functional similarities using two online interactive databases, MaizeCyc (Monaco et al., 2013) and MapMan, accessed *via* http://mapman.gabipd.org and http://maizecyc.maizegdb.org, or http://pathway.gramene.org/MAIZE/class-tree?object=Pathways. The cellular omics tool of version 2.0.2 of MaizeCyc (built on RefGen_v2) and the MapMan program (version 3.5.0 standalone program, Zm_GENOME_RELEASE_09 mapping dataset for B73 RefGen_v2) were used. Both tools, designed for visualizing microarray data, were used with the gene lists (Lists 1–5, defined in the Results) as inputs to generate color-coded representation maps (URLs provided in Table S5). In this study, we produced MaizeCyc output web-browser (html) graphics files using red or orange to denote enzymes from genes with G4 motifs, blue to denote those that lack G4 motifs, and grey to denote those lacking any assigned maize gene model.

Pathway enrichment was quantified statistically using the Fisher's exact test, similar to the *chi*-square test, for non-random associations between G4 motifs found in gene models or specific sub-regions of genes (e.g., A5U, List 1, List 2) and their relative abundance in particular metabolic pathways. Fisher's exact test was used because of the relatively small sample number of genes per pathway. For select pathways described in this paper, we tabulated the enrichment (relative abundance values > 1.0) or depletion (relative abundance values <1.0) and the *p*-value for the corresponding enrichment or depletion. The values for the whole genes or for specified sub-regions of genes (Lists 1–5, Fig. 1) are summarized in Table 1.

Nomenclature for genes named in this study

In maize, the locus (gene) is represented by a unique italicized lowercase word or phrase. Each gene model locus in maize is assigned a unique gene ID (GRMZM number) and the associated gene locus pages list other synonyms for each gene or locus. Experimentally characterized dominant alleles are represented with an upper case first letter (e.g., *Shrunken1*) or lowercase for known recessive alleles (e.g., *shrunken1*). In the absence of evidence on mode of inheritance, lowercase italics gene names are used by convention (http://maizegdb.org/maize_nomenclature.php). Genes that were newly-named in this study were done so

in consultation with MaizeGDB curators (courtesy M. Schaeffer) and on the basis of named homologs from other species or protein domain composition collated by the Conserved Domain Database collection (Marchler-Bauer et al., 2013).

In vitro folding assays for oligonucleotides with G4 motifs

Oligonucleotide folding conditions

All oligonucleotides (Eurofins-MWG-Operon, Huntsville, Alabama, USA) were received salt-free, purified and hydrated to 300 μmol/L in water. Oligonucleotide sequences for the G-rich strands were either wild-type (wt) sequences mutants (mut, changed bases underlined) for use as G4-incompatible controls. The oligonucleotides are given 5′ to 3′ and designed to represent human telomere repeat DNA: HsTelo4xG_wt, GGATACTTAGGGTTAGGGTTAGGGTTAGGGCGAGTC; HsTelo4xG_mut, GGATACTTAGAGTTAGCGTTAGCGTTAGCGCGAGTC; maize telomere repeat DNA: ZmTelo4xG_wt, GGATACTTTAGGGTTTAGGGTTTAGGGTTTAGGGCGAGTC; ZmTelo4xG_mut, GGATACTTTAGCGTTTAGAGTTTAGCGTTTAGAGCGAGTC; the *shrunken1* A5U: sh1_A5U_wt, GGGAGGGAGGGTTTCTCTGGGACGGGAGAGGG, sh1_A5U_mut, GGGAGTGAGGGTTTCTCTGTGACGGGAGAGTG: the *hexokinase4* A5U: hex4_A5U_wt, CGGGGGTGTTGAAGGGAGGAGGAGGGAGGGG; hex4_A5U_mut CGACGGTGTTGAAGCGAGGAGGAGCGAGCGG; the *hexokinase4* A5I1: hex4_A5I1_wt, TGGGGTGGGGGGGGAGCGGG; hex4_A5I1_mut, TGGAGTCGGAGGAGAGCGCG: and the *hexokinase4* ATG: hex4_ATG_wt, CGGGGGGATGGGGCGGGTCGGG: hex4_ATG_mut, CGAGGCGATGAGGCGAGTCGAG. Oligonucleotides were diluted to 10 μmol/L in 10 mmol/L tetrabutylammonium phosphate buffer pH 7.5 supplemented with or without 100 mmol/L KCl, heated in a 1.5 mL polypropylene microfuge tube to 95°C for 15 min on a heat block, then slowly cooled *via* ~8 h drift down to room temperature, then used for spectra or stored at 4°C.

Thermal difference absorption spectroscopy

Oligonucleotides were subjected to folding conditions, diluted to 2.5 μmol/L, and added to capped quartz cuvettes with 10 mm optical path length for spectroscopy using a Cary 300 Bio UV-Vis spectrophotometer equipped with a Varian Cary Peltier cooler. Each measurement was determined from the average of three spectra spanning 230–330 nm at 30 nm/min and 1 nm data intervals. Spectra were taken at 25°C and again after heating to 90°C for 20 min. The averaged thermal difference spectra (values at 90°C minus values at 25°C) were calculated and normalized by setting the value at 330 nm to zero and the highest positive peak value to 1.0 before plotting.

Circular dichroism (CD) spectroscopy

CD spectra (Fig. S1) were also collected on the 10 μmol/L folded oligonucleotide samples, without dilution, at 25°C on an AVIV 202 CD spectrometer using a quartz cuvette with 1 mm optical path. Data were collected at a 200–330 nm range using 3 scans at 15 nm/min, 0.33 s settling time and 1 nm bandwidth. Buffer baseline was recorded using the same parameters and subtracted from the sample spectra before plotting.

ACKNOWLEDGEMENTS

We are grateful to M. Schaeffer (at MaizeGDB.org) for assistance with maize gene nomenclature, to B.P. Chadwick, H. Cui, J.H. Dennis and G.G. Hoffman for helpful discussions, and to J. Kennedy for editorial assistance. This work was supported by USDA-ARS and grants from the National Science Foundation (PGRP IOS-1025954 to HWB, PGRP IOS-1116561 to KEK and coworkers), the U.S. Department of Agriculture (NRI-Plant Biochemistry 07-03580 to KEK and coworkers) and The Florida State University (CRC Planning Grant to HWB, OMNI award No. 0000025471).

SUPPLEMENTARY DATA

Table S1. All maize G4 motifs, names and locations.
Table S2. G4 patch motifs, names and locations.
Table S3. Maize Genes with G4 motifs and related annotations.
Table S4. Gene lists classified by G4 motif locations.
Table S5. Links to URL figures of gene lists displayed in MaizeCyc.
Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.jgg.2014.10.004.

REFERENCES

Armanios, M., Blackburn, E.H., 2012. The telomere syndromes. Nat. Rev. Genet. 13, 693–704.
Bailey-Serres, J., Fukao, T., Gibbs, D.J., Holdsworth, M.J., Lee, S.C., Licausi, F., Perata, P., Voesenek, L.A., van Dongen, J.T., 2012a. Making sense of low oxygen sensing. Trends Plant Sci. 17, 129–138.
Bailey-Serres, J., Lee, S.C., Brinton, E., 2012b. Waterproofing crops: effective flooding survival strategies. Plant Physiol. 160, 1698–1709.
Bailey-Serres, J., Voesenek, L.A., 2008. Flooding stress: acclimations and genetic diversity. Ann. Rev. Plant Biol. 59, 313–339.
Beckett, J., Burns, J., Broxson, C., Tornaletti, S., 2012. Spontaneous DNA lesions modulate DNA structural transitions occurring at nuclease hypersensitive element III(1) of the human c-myc proto-oncogene. Biochemistry 51, 5257–5268.
Bennetzen, J.L., Hake, S.C., 2009. Handbook of Maize: Genetics and Genomics. Springer, New York.
Biffi, G., Tannahill, D., McCafferty, J., Balasubramanian, S., 2013. Quantitative visualization of DNA G-quadruplex structures in human cells. Nat. Chem. 5, 182–186.
Bihmidine, S., Lin, J., Stone, J.M., Awada, T., Specht, J.E., Clemente, T.E., 2013. Activity of the *Arabidopsis rd29a* and *rd29b* promoter elements in soybean under water stress. Planta 237, 55–64.
Blackburn, E.H., Epel, E.S., 2012. Telomeres and adversity: too toxic to ignore. Nature 490, 169–171.
Blackburn, E.H., Greider, C.W., Szostak, J.W., 2006. Telomeres and telomerase: the path from maize, tetrahymena and yeast to human cancer and aging. Nat. Med. 12, 1133–1138.

Bochman, M.L., Paeschke, K., Zakian, V.A., 2012. DNA secondary structures: stability and function of G-quadruplex structures. Nat. Rev. Genet. 13, 770−780.

Bouche, N., Fromm, H., 2004. GABA in plants: just a metabolite? Trends Plant Sci. 9, 110−115.

Bourdoncle, A., Estevez Torres, A., Gosse, C., Lacroix, L., Vekhoff, P., Le Saux, T., Jullien, L., Mergny, J.L., 2006. Quadruplex-based molecular beacons as tunable DNA probes. J. Am. Chem. Soc. 128, 11094−11105.

Breit, J.F., Ault-Ziel, K., Al-Mehdi, A.B., Gillespie, M.N., 2008. Nuclear protein-induced bending and flexing of the hypoxic response element of the rat vascular endothelial growth factor promoter. FASEB J. 22, 19−29.

Brooks, T.A., Hurley, L.H., 2009. The role of supercoiling in transcriptional control of myc and its importance in molecular therapeutics. Nat. Rev. Cancer 9, 849−861.

Brooks, T.A., Hurley, L.H., 2010. Targeting MYC expression through G-quadruplexes. Genes Cancer 1, 641−649.

Brooks, T.A., Kendrick, S., Hurley, L., 2010. Making sense of G-quadruplex and i-motif functions in oncogene promoters. FEBS J. 277, 3459−3469.

Brown, A.N., Lauter, N., Vera, D.L., McLaughlin-Large, K.A., Steele, T.M., Fredette, N.C., Bass, H.W., 2011a. QTL mapping and candidate gene analysis of telomere length control factors in maize (*Zea mays* L.). G3: Genes, Genomes, Genetics 1, 437−450.

Brown, R.V., Danford, F.L., Gokhale, V., Hurley, L.H., Brooks, T.A., 2011b. Demonstration that drug-targeted down-regulation of MYC in non-Hodgkins lymphoma is directly mediated through the promoter G-quadruplex. J. Biol. Chem. 286, 41018−41027.

Bugaut, A., Balasubramanian, S., 2008. A sequence-independent study of the influence of short loop lengths on the stability and topology of intramolecular DNA G-quadruplexes. Biochemistry 47, 689−697.

Burge, S., Parkinson, G.N., Hazel, P., Todd, A.K., Neidle, S., 2006. Quadruplex DNA: sequence, topology and structure. Nucleic Acids Res. 34, 5402−5415.

Cahoon, L.A., Seifert, H.S., 2013. Transcription of a cis-acting, noncoding, small RNA is required for pilin antigenic variation in *Neisseria gonorrhoeae*. PLoS Pathog. 9, e1003074.

Caldana, C., Li, Y., Leisse, A., Zhang, Y., Bartholomaeus, L., Fernie, A.R., Willmitzer, L., Giavalisco, P., 2013. Systemic analysis of inducible target of rapamycin mutants reveal a general metabolic switch controlling growth in *Arabidopsis thaliana*. Plant J. 73, 897−909.

Capra, J.A., Paeschke, K., Singh, M., Zakian, V.A., 2010. G-quadruplex DNA sequences are evolutionarily conserved and associated with distinct genomic features in *Saccharomyces cerevisiae*. PLoS Comput. Biol. 6, e1000861.

Chen, Y., Yang, D., 2012. Sequence, stability, and structure of G-quadruplexes and their interactions with drugs. Curr. Protoc. Nucleic Acid Chem. Chapter 17, Unit 17.5.

Chinnapen, D.J., Sen, D., 2004. A deoxyribozyme that harnesses light to repair thymine dimers in DNA. Proc. Natl. Acad. Sci. USA 101, 65−69.

Clark, D.W., Phang, T., Edwards, M.G., Geraci, M.W., Gillespie, M.N., 2012. Promoter G-quadruplex sequences are targets for base oxidation and strand cleavage during hypoxia-induced transcription. Free Radic. Biol. Med. 53, 51−59.

Cogoi, S., Xodo, L.E., 2006. G-quadruplex formation within the promoter of the KRAS proto-oncogene and its effect on transcription. Nucleic Acids Res. 34, 2536−2549.

Davis, G.L., McMullen, M.D., Baysdorfer, C., Musket, T., Grant, D., Staebell, M., Xu, G., Polacco, M., Koster, L., Melia-Hancock, S., Houchins, K., Chao, S., Coe Jr., E.H., 1999. A maize map standard with sequenced core markers, grass genome reference points and 932 expressed sequence tagged sites (ESTs) in a 1736-locus map. Genetics 152, 1137−1172.

De Armond, R., Wood, S., Sun, D., Hurley, L.H., Ebbinghaus, S.W., 2005. Evidence for the presence of a guanine quadruplex forming region within a polypurine tract of the hypoxia inducible factor 1alpha promoter. Biochemistry 44, 16341−16350.

Dexheimer, T.S., Sun, D., Hurley, L.H., 2006. Deconvoluting the structural and drug-recognition complexity of the G-quadruplex-forming region upstream of the bcl-2 P1 promoter. J. Am. Chem. Soc. 128, 5404−5415.

Dobrenel, T., Marchive, C., Azzopardi, M., Clement, G., Moreau, M., Sormani, R., Robaglia, C., Meyer, C., 2013. Sugar metabolism and the plant target of rapamycin kinase: a sweet operaTOR? Front. Plant Sci. 4, 93.

Du, Z., Zhao, Y., Li, N., 2008. Genome-wide analysis reveals regulatory role of G4 DNA in gene transcription. Genome Res. 18, 233−241.

Du, Z., Zhao, Y., Li, N., 2009. Genome-wide colonization of gene regulatory elements by G4 DNA motifs. Nucleic Acids Res. 37, 6784−6798.

Eddy, J., Maizels, N., 2006. Gene function correlates with potential for G4 DNA formation in the human genome. Nucleic Acids Res. 34, 3887−3896.

Eddy, J., Maizels, N., 2008. Conserved elements with potential to form polymorphic G-quadruplex structures in the first intron of human genes. Nucleic Acids Res. 36, 1321−1333.

Eddy, J., Vallur, A.C., Varma, S., Liu, H., Reinhold, W.C., Pommier, Y., Maizels, N., 2011. G4 motifs correlate with promoter-proximal transcriptional pausing in human genes. Nucleic Acids Res. 39, 4975−4983.

Epstein, A.C., Gleadle, J.M., McNeill, L.A., Hewitson, K.S., O'Rourke, J., Mole, D.R., Mukherji, M., Metzen, E., Wilson, M.I., Dhanda, A., Tian, Y.M., Masson, N., Hamilton, D.L., Jaakkola, P., Barstead, R., Hodgkin, J., Maxwell, P.H., Pugh, C.W., Schofield, C.J., Ratcliffe, P.J., 2001. *C. elegans* EGL-9 and mammalian homologs define a family of dioxygenases that regulate HIF by prolyl hydroxylation. Cell 107, 43−54.

Ferjani, A., Segami, S., Horiguchi, G., Muto, Y., Maeshima, M., Tsukaya, H., 2011. Keep an eye on PPi: the vacuolar-type H^+-pyrophosphatase regulates postgerminative development in *Arabidopsis*. Plant Cell 23, 2895−2908.

Fernando, H., Reszka, A.P., Huppert, J., Ladame, S., Rankin, S., Venkitaraman, A.R., Neidle, S., Balasubramanian, S., 2006. A conserved quadruplex motif located in a transcription activation site of the human c-kit oncogene. Biochemistry 45, 7854−7860.

Foyer, C.H., Noctor, G., 2005. Redox homeostasis and antioxidant signaling: a metabolic interface between stress perception and physiological responses. Plant Cell 17, 1866−1875.

Foyer, C.H., Noctor, G., 2011. Ascorbate and glutathione: the heart of the redox hub. Plant Physiol. 155, 2−18.

Fukao, T., Yeung, E., Bailey-Serres, J., 2011. The submergence tolerance regulator SUB1A mediates crosstalk between submergence and drought tolerance in rice. Plant Cell 23, 412−427.

Gibbs, D.J., Lee, S.C., Isa, N.M., Gramuglia, S., Fukao, T., Bassel, G.W., Correia, C.S., Corbineau, F., Theodoulou, F.L., Bailey-Serres, J., Holdsworth, M.J., 2011. Homeostatic response to hypoxia is regulated by the N-end rule pathway in plants. Nature 479, 415−418.

Goodstein, D.M., Shu, S., Howson, R., Neupane, R., Hayes, R.D., Fazo, J., Mitros, T., Dirks, W., Hellsten, U., Putnam, N., Rokhsar, D.S., 2012. Phytozome: a comparative platform for green plant genomics. Nucleic Acids Res. 40, D1178−D1186.

Gray, J., Bevan, M., Brutnell, T., Buell, C.R., Cone, K., Hake, S., Jackson, D., Kellogg, E., Lawrence, C., McCouch, S., Mockler, T., Moose, S., Paterson, A., Peterson, T., Rokshar, D., Souza, G.M., Springer, N., Stein, N., Timmermans, M., Wang, G.L., Grotewold, E., 2009. A recommendation for naming transcription factor proteins in the grasses. Plant Physiol. 149, 4−6.

Guo, K., Gokhale, V., Hurley, L.H., Sun, D., 2008. Intramolecularly folded G-quadruplex and i-motif structures in the proximal promoter of the vascular endothelial growth factor gene. Nucleic Acids Res. 36, 4598−4608.

Guo, K., Pourpak, A., Beetz-Rogers, K., Gokhale, V., Sun, D., Hurley, L.H., 2007. Formation of pseudosymmetrical G-quadruplex and i-motif structures in the proximal promoter region of the *RET* oncogene. J. Am. Chem. Soc. 129, 10220−10228.

Halder, K., Halder, R., Chowdhury, S., 2009. Genome-wide analysis predicts DNA structural motifs as nucleosome exclusion signals. Mol. Biosyst 5, 1703−1712.

Hazel, P., Huppert, J., Balasubramanian, S., Neidle, S., 2004. Loop-length-dependent folding of G-quadruplexes. J. Am. Chem. Soc. 126, 16405−16415.

Henderson, A., Wu, Y., Huang, Y.C., Chavez, E.A., Platt, J., Johnson, F.B., Brosh Jr., R.M., Sen, D., Lansdorp, P.M., 2014. Detection of G-quadruplex DNA in mammalian cells. Nucleic Acids Res. 42, 860−869.

Hershman, S.G., Chen, Q., Lee, J.Y., Kozak, M.L., Yue, P., Wang, L.S., Johnson, F.B., 2008. Genomic distribution and functional analyses of potential G-quadruplex-forming sequences in *Saccharomyces cerevisiae*. Nucleic Acids Res. 36, 144−156.

Huber, M.D., Duquette, M.L., Shiels, J.C., Maizels, N., 2006. A conserved G4 DNA binding domain in RecQ family helicases. J. Mol. Biol. 358, 1071−1080.

Huber, S.C., Akazawa, T., 1986. A novel sucrose synthase pathway for sucrose degradation in cultured sycamore cells. Plant Physiol. 81, 1008−1013.

Huppert, J.L., 2007. Four-stranded DNA: cancer, gene regulation and drug development. Philos. Trans. A. Math. Phys. Eng. Sci. 365, 2969−2984.

Huppert, J.L., 2010. Structure, location and interactions of G-quadruplexes. FEBS J. 277, 3452−3458.

Huppert, J.L., Balasubramanian, S., 2005. Prevalence of quadruplexes in the human genome. Nucleic Acids Res. 33, 2908−2916.

Ivan, M., Kondo, K., Yang, H., Kim, W., Valiando, J., Ohh, M., Salic, A., Asara, J.M., Lane, W.S., Kaelin Jr., W.G., 2001. HIFalpha targeted for VHL-mediated destruction by proline hydroxylation: implications for O2 sensing. Science 292, 464−468.

Jaakkola, P., Mole, D.R., Tian, Y.M., Wilson, M.I., Gielbert, J., Gaskell, S.J., von Kriegsheim, A., Hebestreit, H.F., Mukherji, M., Schofield, C.J., Maxwell, P.H., Pugh, C.W., Ratcliffe, P.J., 2001. Targeting of HIF-alpha to the von Hippel-Lindau ubiquitylation complex by O2-regulated prolyl hydroxylation. Science 292, 468−472.

Jakoby, M., Weisshaar, B., Droge-Laser, W., Vicente-Carbajosa, J., Tiedemann, J., Kroj, T., Parcy, F., 2002. bZIP transcription factors in *Arabidopsis*. Trends Plant Sci. 7, 106−111.

Juranek, S.A., Paeschke, K., 2012. Cell cycle regulation of G-quadruplex DNA structures at telomeres. Curr. Pharm. Des. 18, 1867−1872.

Karp, P.D., Paley, S.M., Krummenacker, M., Latendresse, M., Dale, J.M., Lee, T.J., Kaipa, P., Gilham, F., Spaulding, A., Popescu, L., Altman, T., Paulsen, I., Keseler, I.M., Caspi, R., 2010. Pathway tools version 13.0: integrated software for pathway/genome informatics and systems biology. Brief. Bioinform. 11, 40−79.

Kelliher, T., Walbot, V., 2012. Hypoxia triggers meiotic fate acquisition in maize. Science 337, 345−348.

Keunen, E., Peshev, D., Vangronsveld, J., W, V.D.E., Cuypers, A., 2013. Plant sugars are crucial players in the oxidative challenge during abiotic stress: extending the traditional concept. Plant Cell Environ. 36, 242−1255.

Kikin, O., D'Antonio, L., Bagga, P.S., 2006. QGRS mapper: a web-based server for predicting G-quadruplexes in nucleotide sequences. Nucleic Acids Res. 34, W676−W682.

Koch, K.E., 2004. Sucrose metabolism: regulatory mechanisms and pivotal roles in sugar sensing and plant development. Curr. Opin. Plant Biol. 7, 235−246.

Koch, K.E., 1996. Carbohydrate-modulated gene expression in plants. Annu. Rev. Plant Physiol. Plant Mol. Biol. 47, 509−540.

Koch, K.E., Ying, Z., Wu, Y., Avigne, W.T., 2000. Multiple paths of sugar-sensing and a sugar/oxygen overlap for genes of sucrose and ethanol metabolism. J. Exp. Bot. 51, 417−427.

Lam, E.Y., Beraldi, D., Tannahill, D., Balasubramanian, S., 2013. G-quadruplex structures are stable and detectable in human genomic DNA. Nat. Commun. 4, 1796.

Laurie, D.A., Bennett, M.D., 1985. Nuclear DNA content in the genera *Zea* and *Sorghum*. Intergeneric, interspecific and intraspecific variation. Heredity 55, 307−313.

Lexa, M., Kejnovsky, E., Steflova, P., Konvalinova, H., Vorlickova, M., Vyskot, B., 2014. Quadruplex-forming sequences occupy discrete regions inside plant LTR retrotransposons. Nucleic Acids Res. 42, 968−978.

Licausi, F., Kosmacz, M., Weits, D.A., Giuntoli, B., Giorgi, F.M., Voesenek, L.A., Perata, P., van Dongen, J.T., 2011. Oxygen sensing in plants is mediated by an N-end rule pathway for protein destabilization. Nature 479, 419−422.

Loescher, W.H., 1987. Physiology and metabolism of sugar alcohols in higher plants. Physiol. Plant. 70, 553−557.

Maizels, N., 2006. Dynamic roles for G4 DNA in the biology of eukaryotic cells. Nat. Struct. Mol. Biol. 13, 1055−1059.

Maizels, N., Gray, L.T., 2013. The G4 genome. PLoS Genet. 9, e1003468.

Mani, P., Yadav, V.K., Das, S.K., Chowdhury, S., 2009. Genome-wide analyses of recombination prone regions predict role of DNA structural motif in recombination. PLoS ONE 4, e4399.

Marchler-Bauer, A., Zheng, C., Chitsaz, F., Derbyshire, M.K., Geer, L.Y., Geer, R.C., Gonzales, N.R., Gwadz, M., Hurwitz, D.I., Lanczycki, C.J., Lu, F., Lu, S., Marchler, G.H., Song, J.S., Thanki, N., Yamashita, R.A., Zhang, D., Bryant, S.H., 2013. CDD: conserved domains and protein three-dimensional structure. Nucleic Acids Res. 41, D348−D352.

Menendez, C., Frees, S., Bagga, P.S., 2012. QGRS-H predictor: a web server for predicting homologous quadruplex forming G-rich sequence motifs in nucleotide sequences. Nucleic Acids Res. 40, W96−W103.

Milanesi, L., D'Angelo, D., Rogozin, I.B., 1999. GeneBuilder: interactive in silico prediction of gene structure. Bioinformatics 15, 612−621.

Monaco, M.K., Sen, T.Z., Dharmawardhana, P.D., Ren, L., Schaeffer, M., Naithani, S., Amarasinghe, V., Thomason, J., Harper, L., Gardiner, J., Cannon, E.K.S., Lawrence, C.J., Ware, D., Jaiswal, P., 2013. Maize metabolic network construction and transcriptome analysis. Plant Genome 6. http://dx.doi.org/10.3835/plantgenome2012.09.0025.

Mullen, M.A., Olson, K.J., Dallaire, P., Major, F., Assmann, S.M., Bevilacqua, P.C., 2010. RNA G-quadruplexes in the model plant species *Arabidopsis thaliana*: prevalence and possible functional roles. Nucleic Acids Res. 38, 8149−8163.

Nakano, T., Suzuki, K., Fujimura, T., Shinshi, H., 2006. Genome-wide analysis of the ERF gene family in *Arabidopsis* and rice. Plant Physiol. 140, 411−432.

Nishizawa, A., Yabuta, Y., Shigeoka, S., 2008. Galactinol and raffinose constitute a novel function to protect plants from oxidative damage. Plant Physiol. 147, 1251−1263.

Nussbaumer, T., Martis, M.M., Roessner, S.K., Pfeifer, M., Bader, K.C., Sharma, S., Gundlach, H., Spannagl, M., 2013. MIPS PlantsDB: a database framework for comparative plant genome research. Nucleic Acids Res. 41, D1144−D1151.

Palumbo, S.L., Ebbinghaus, S.W., Hurley, L.H., 2009. Formation of a unique end-to-end stacked pair of G-quadruplexes in the *hTERT* core promoter with implications for inhibition of telomerase by G-quadruplex-interactive ligands. J. Am. Chem. Soc. 131, 10878−10891.

Palumbo, S.L., Memmott, R.M., Uribe, D.J., Krotova-Khan, Y., Hurley, L.H., Ebbinghaus, S.W., 2008. A novel G-quadruplex-forming GGA repeat region in the c-myb promoter is a critical regulator of promoter activity. Nucleic Acids Res. 36, 1755−1769.

Pontier, D.B., Kruisselbrink, E., Guryev, V., Tijsterman, M., 2009. Isolation of deletion alleles by G4 DNA-induced mutagenesis. Nat. Meth. 6, 655−657.

Punta, M., Coggill, P.C., Eberhardt, R.Y., Mistry, J., Tate, J., Boursnell, C., Pang, N., Forslund, K., Ceric, G., Clements, J., Heger, A., Holm, L., Sonnhammer, E.L., Eddy, S.R., Bateman, A., Finn, R.D., 2012. The PFAM protein families database. Nucleic Acids Res. 40, D290−D301.

Qin, Y., Fortin, J.S., Tye, D., Gleason-Guzman, M., Brooks, T.A., Hurley, L.H., 2010. Molecular cloning of the human platelet-derived growth factor receptor beta (PDGFR-beta) promoter and drug targeting of the G-quadruplex-forming region to repress pdgfr-beta expression. Biochemistry 49, 4208−4219.

Qin, Y., Hurley, L.H., 2008. Structures, folding patterns, and functions of intramolecular DNA G-quadruplexes found in eukaryotic promoter regions. Biochimie 90, 1149−1171.

Rawal, P., Kummarasetti, V.B., Ravindran, J., Kumar, N., Halder, K., Sharma, R., Mukerji, M., Das, S.K., Chowdhury, S., 2006. Genome-wide prediction of G4 DNA as regulatory motifs: role in *Escherichia coli* global regulation. Genome Res. 16, 644−655.

Robaglia, C., Thomas, M., Meyer, C., 2012. Sensing nutrient and energy status by SnRK1 and TOR kinases. Curr. Opin. Plant Biol. 15, 301−307.

Rolletschek, H., Koch, K., Wobus, U., Borisjuk, L., 2005. Positional cues for the starch/lipid balance in maize kernels and resource partitioning to the embryo. Plant J. 42, 69−83.

Rolletschek, H., Melkus, G., Grafahrend-Belau, E., Fuchs, J., Heinzel, N., Schreiber, F., Jakob, P.M., Borisjuk, L., 2011. Combined noninvasive

imaging and modeling approaches reveal metabolic compartmentation in the barley endosperm. Plant Cell 23, 3041–3054.

Ruan, Y.L., Jin, Y., Yang, Y.J., Li, G.J., Boyer, J.S., 2010. Sugar input, metabolism, and signaling mediated by invertase: roles in development, yield potential, and response to drought and heat. Mol. Plant 3, 942–955.

Ruchko, M.V., Gorodnya, O.M., Pastukh, V.M., Swiger, B.M., Middleton, N.S., Wilson, G.L., Gillespie, M.N., 2009. Hypoxia-induced oxidative base modifications in the VEGF hypoxia-response element are associated with transcriptionally active nucleosomes. Free Radic. Biol. Med. 46, 352–359.

Sachs, M.M., Freeling, M., Okimoto, R., 1980. The anaerobic proteins of maize. Cell 20, 761–767.

Salamov, A.A., Solovyev, V.V., 2000. *Ab initio* gene finding in Drosophila genomic DNA. Genome Res. 10, 516–522.

SanMiguel, P., Gaut, B.S., Tikhonov, A., Nakajima, Y., Bennetzen, J.L., 1998. The paleontology of intergene retrotransposons of maize. Nat. Genet. 20, 43–45.

Schnable, J.C., Freeling, M., Lyons, E., 2012. Genome-wide analysis of syntenic gene deletion in the grasses. Genome Biol. Evol. 4, 265–277.

Schnable, P.S., Ware, D., Fulton, R.S., Stein, J.C., Wei, F., Pasternak, S., Liang, C., Zhang, J., Fulton, L., Graves, T.A., Minx, P., Reily, A.D., Courtney, L., Kruchowski, S.S., Tomlinson, C., Strong, C., Delehaunty, K., Fronick, C., Courtney, B., Rock, S.M., Belter, E., Du, F., Kim, K., Abbott, R.M., Cotton, M., Levy, A., Marchetto, P., Ochoa, K., Jackson, S.M., Gillam, B., Chen, W., Yan, L., Higginbotham, J., Cardenas, M., Waligorski, J., Applebaum, E., Phelps, L., Falcone, J., Kanchi, K., Thane, T., Scimone, A., Thane, N., Henke, J., Wang, T., Ruppert, J., Shah, N., Rotter, K., Hodges, J., Ingenthron, E., Cordes, M., Kohlberg, S., Sgro, J., Delgado, B., Mead, K., Chinwalla, A., Leonard, S., Crouse, K., Collura, K., Kudrna, D., Currie, J., He, R., Angelova, A., Rajasekar, S., Mueller, T., Lomeli, R., Scara, G., Ko, A., Delaney, K., Wissotski, M., Lopez, G., Campos, D., Braidotti, M., Ashley, E., Golser, W., Kim, H., Lee, S., Lin, J., Dujmic, Z., Kim, W., Talag, J., Zuccolo, A., Fan, C., Sebastian, A., Kramer, M., Spiegel, L., Nascimento, L., Zutavern, T., Miller, B., Ambroise, C., Muller, S., Spooner, W., Narechania, A., Ren, L., Wei, S., Kumari, S., Faga, B., Levy, M.J., McMahan, L., Van Buren, P., Vaughn, M.W., Ying, K., Yeh, C.T., Emrich, S.J., Jia, Y., Kalyanaraman, A., Hsia, A.P., Barbazuk, W.B., Baucom, R.S., Brutnell, T.P., Carpita, N.C., Chaparro, C., Chia, J.M., Deragon, J.M., Estill, J.C., Fu, Y., Jeddeloh, J.A., Han, Y., Lee, H., Li, P., Lisch, D.R., Liu, S., Liu, Z., Nagel, D.H., McCann, M.C., SanMiguel, P., Myers, A.M., Nettleton, D., Nguyen, J., Penning, B.W., Ponnala, L., Schneider, K.L., Schwartz, D.C., Sharma, A., Soderlund, C., Springer, N.M., Sun, Q., Wang, H., Waterman, M., Westerman, R., Wolfgruber, T.K., Yang, L., Yu, Y., Zhang, L., Zhou, S., Zhu, Q., Bennetzen, J.L., Dawe, R.K., Jiang, J., Jiang, N., Presting, G.G., Wessler, S.R., Aluru, S., Martienssen, R.A., Clifton, S.W., McCombie, W.R., Wing, R.A., Wilson, R.K., 2009. The B73 maize genome: complexity, diversity, and dynamics. Science 326, 1112–1115.

Sen, D., Gilbert, W., 1988. Formation of parallel four-stranded complexes by guanine-rich motifs in DNA and its implications for meiosis. Nature 334, 364–366.

Sen, T.Z., Andorf, C.M., Schaeffer, M.L., Harper, L.C., Sparks, M.E., Duvick, J., Brendel, V.P., Cannon, E., Campbell, D.A., Lawrence, C.J., 2009. MaizeGDB becomes 'sequence-centric'. Database (Oxford) 2009, bap020.

Shen, B., Hohmann, S., Jensen, R.G., Bohnert, a, H., 1999. Roles of sugar alcohols in osmotic stress adaptation. Replacement of glycerol by mannitol and sorbitol in yeast. Plant Physiol. 121, 45–52.

Sickler, C.M., Edwards, G.E., Kiirats, O., Gao, Z., Loescher, W., 2007. Response of mannitol-producing *Arabidopsis thaliana* to abiotic stress. Funct. Plant Biol. 34, 382–391.

Siddiqui-Jain, A., Grand, C.L., Bearss, D.J., Hurley, L.H., 2002. Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c-MYC transcription. Proc. Natl. Acad. Sci. USA 99, 11593–11598.

Simonsson, T., 2001. G-quadruplex DNA structures–variations on a theme. Biol. Chem. 382, 621–628.

Stegle, O., Payet, L., Mergny, J.L., MacKay, D.J., Leon, J.H., 2009. Predicting and understanding the stability of G-quadruplexes. Bioinformatics 25, i374–382.

Sulpice, R., Trenkamp, S., Steinfath, M., Usadel, B., Gibon, Y., Witucka-Wall, H., Pyl, E.T., Tschoep, H., Steinhauser, M.C., Guenther, M., Hoehne, M., Rohwer, J.M., Altmann, T., Fernie, A.R., Stitt, M., 2010. Network analysis of enzyme activities and metabolite levels and their relationship to biomass in a large panel of *Arabidopsis* accessions. Plant Cell 22, 2872–2893.

Sun, D., Hurley, L.H., 2010. Biochemical techniques for the characterization of G-quadruplex structures: EMSA, DMS footprinting, and DNA polymerase stop assay. Methods Mol. Biol. 608, 65–79.

Sundquist, W.I., Klug, A., 1989. Telomeric DNA dimerizes by formation of guanine tetrads between hairpin loops. Nature 342, 825–829.

Takahashi, H., Nakagawa, A., Kojima, S., Takahashi, A., Cha, B.Y., Woo, J.T., Nagai, K., Machida, Y., Machida, C., 2012. Discovery of novel rules for G-quadruplex-forming sequences in plants by using bioinformatics methods. J. Biosci. Bioeng. 114, 570–575.

Thimm, O., Blasing, O., Gibon, Y., Nagel, A., Meyer, S., Kruger, P., Selbig, J., Muller, L.A., Rhee, S.Y., Stitt, M., 2004. MAPMAN: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. Plant J. 37, 914–939.

Tiessen, A., Padilla-Chacon, D., 2012. Subcellular compartmentation of sugar signaling: links among carbon cellular status, route of sucrolysis, sink-source allocation, and metabolic partitioning. Front. Plant Sci. 3, 306.

Todd, A.K., Johnston, M., Neidle, S., 2005. Highly prevalent putative quadruplex sequence motifs in human DNA. Nucleic Acids Res. 33, 2901–2907.

Todd, A.K., Neidle, S., 2011. Mapping the sequences of potential guanine quadruplex motifs. Nucleic Acids Res. 39, 4917–4927.

Valluru, R., Van den Ende, W., 2011. Myo-inositol and beyond–emerging networks under stress. Plant Sci. 181, 387–400.

Verma, A., Halder, K., Halder, R., Yadav, V.K., Rawal, P., Thakur, R.K., Mohd, F., Sharma, A., Chowdhury, S., 2008. Genome-wide computational and expression analyses reveal G-quadruplex DNA motifs as conserved cis-regulatory elements in human and related species. J. Med. Chem. 51, 5641–5649.

Verma, A., Yadav, V.K., Basundra, R., Kumar, A., Chowdhury, S., 2009. Evidence of genome-wide G4 DNA-mediated gene expression in human cancer cells. Nucleic Acids Res. 37, 4194–4204.

Weng, H.Y., Huang, H.L., Zhao, P.P., Zhou, H., Qu, L.H., 2012. Translational repression of cyclin D3 by a stable G-quadruplex in its 5′ UTR: implications for cell cycle regulation. RNA Biol. 9, 1099–1109.

Williamson, J.R., Raghuraman, M.K., Cech, T.R., 1989. Monovalent cation-induced structure of telomeric DNA: the G-quartet model. Cell 59, 871–880.

Wong, H.M., Stegle, O., Rodgers, S., Huppert, J.L., 2010. A toolbox for predicting G-quadruplex formation and stability. J. Nucleic Acids 2010.

Wouters, A., Boeckx, C., Vermorken, J.B., Van den Weyngaert, D., Peeters, M., Lardon, F., 2013. The intriguing interplay between therapies targeting the epidermal growth factor receptor, the hypoxic microenvironment and hypoxia-inducible factors. Curr. Pharm. Des. 19, 907–917.

Xiong, Y., McCormack, M., Li, L., Hall, Q., Xiang, C., Sheen, J., 2013. Glucose-TOR signalling reprograms the transcriptome and activates meristems. Nature 496, 181–186.

Xu, X.M., Lin, H., Maple, J., Bjorkblom, B., Alves, G., Larsen, J.P., Moller, S.G., 2010. The Arabidopsis DJ-1a protein confers stress protection through cytosolic SOD activation. J. Cell Sci. 123, 1644–1651.

Xu, Y., Sugiyama, H., 2006. Formation of the G-quadruplex and i-motif structures in retinoblastoma susceptibility genes (Rb). Nucleic Acids Res. 34, 949–954.

Yadav, V.K., Abraham, J.K., Mani, P., Kulshrestha, R., Chowdhury, S., 2008. QuadBase: genome-wide database of G4 DNA–occurrence and conservation in human, chimpanzee, mouse and rat promoters and 146 microbes. Nucleic Acids Res. 36, D381–D385.

Yang, D., Okamoto, K., 2010. Structural insights into G-quadruplexes: towards new anticancer drugs. Future Med. Chem. 2, 619–646.

Yatabe, N., Kyo, S., Maida, Y., Nishi, H., Nakamura, M., Kanaya, T., Tanaka, M., Isaka, K., Ogawa, S., Inoue, M., 2004. HIF-1-mediated activation of telomerase in cervical cancer cells. Oncogene 23, 3708−3715.

Youens-Clark, K., Buckler, E., Casstevens, T., Chen, C., Declerck, G., Derwent, P., Dharmawardhana, P., Jaiswal, P., Kersey, P., Karthikeyan, A.S., Lu, J., McCouch, S.R., Ren, L., Spooner, W., Stein, J.C., Thomason, J., Wei, S., Ware, D., 2011. Gramene database in 2010: updates and extensions. Nucleic Acids Res. 39, D1085−D1094.

Yu, R.M., Chen, E.X., Kong, R.Y., Ng, P.K., Mok, H.O., Au, D.W., 2006. Hypoxia induces telomerase reverse transcriptase (TERT) gene expression in non-tumor fish tissues *in vivo*: the marine medaka (*Oryzias melastigma*) model. BMC Mol. Biol. 7, 27.

Yuan, L., Tian, T., Chen, Y., Yan, S., Xing, X., Zhang, Z., Zhai, Q., Xu, L., Wang, S., Weng, X., Yuan, B., Feng, Y., Zhou, X., 2013. Existence of G-quadruplex structures in promoter region of oncogenes confirmed by G-quadruplex DNA cross-linking strategy. Sci. Rep. 3, 1811.

Zeng, Y., Wu, Y., Avigne, W.T., Koch, K.E., 1998. Differential regulation of sugar-sensitive sucrose synthases by hypoxia and anoxia indicate complementary transcriptional and posttranscriptional responses. Plant Physiol. 116, 1573−1583.