Development of Tissue-Specific Promoters and Elaboration of Multi-tissue Effects of *MADS69* on Flowering Time and Plant Morphology in Maize

by

Nathaniel Schleif

A dissertation submitted in partial fulfillment of

the requirements for the degree of

Doctor of Philosophy

(Plant Breeding and Plant Genetics)

at the

UNIVERSITY OF WISCONSIN-MADISON

2023

Date of final oral examination: 10/16/2023

The dissertation is approved by the following members of the Final Examination Committee:

Heidi F. Kaeppler, Associate Professor, Plant and Agroecosystem Sciences
Jean-Michel Ané, Professor, Plant and Agroecosystem Sciences
Marisa Otegui, Associate Professor, Botany
Richard Amasino, Professor, Biochemistry
Shawn Kaeppler, Professor, Plant and Agroecosystem Sciences

DEDICATION

This work is dedicated foremostly to my wonderful wife, Ruby, who has been my greatest supporter throughout all of this. I am nothing without her. It is also dedicated to my parents and grandparents who have always encouraged and supported my interest in science. Finally, it is dedicated to my newborn Aurora; I hope someday she will find as much joy as I have in science.

ACKNOWLEDGMENTS

I would like to express my deepest gratitude to Dr. Heidi F. Kaeppler who has been an outstanding mentor before and throughout my graduate studies. She has taught me so much about how to be a scientist and I am forever indebted to her. Dr. Jean-Michel Ané has also played a pivotal role in my research career, both as an undergraduate and now on my committee. I am extremely grateful for all my other committee members, Dr. Shawn Kaeppler, Dr. Marisa Otegui, and Dr. Richard Amasino, each of which have provided vital feedback and ideas in the advancement of my research herein presented. I could not have undertaken this journey without the love and support of my wife Ruby Schleif, who been a constant source of emotional support throughout this. I am also grateful for my family in supporting my graduate studies. Special thanks to my fellow students who have uplifted me by learning by my side, including Dr. Frank McFarland and Dr. Lily Hislop. Many thanks to Dr. Brian Martinell and Dr. Ray Collier, each of which helped me learn and cultivate a wide swath of practical lab skills. More broadly, I am grateful for everyone at the Wisconsin Crop Innovation Center who has helped me to learn and grow. Finally, I'd like to recognize Zola Regele for her dedicated help as an undergraduate researcher.

ABSTRACT

Genetic manipulation of flowering time is vital for maximization of crop yield, especially for maize whose global cultivation demands finely tuned flowering. Underlying flowering time are complex genetic networks which span spatial and developmental space. Additionally, flowering time is not separate from other traits and developmental systems, making the characterization of theorized flowering time genes challenging. Disentangling a particular gene's effect on flowering time commonly requires the use of tissue-specific promoters which can isolate the potential effect of a gene to a given region and developmental phase. This approach is difficult in maize due to a general dearth of well characterized promoters. To address this need, we identified and characterized 6 tissue-specific promoters from an initial pool of 14 candidates. Tissue-specificity was identified through analysis of gene expression data and candidates were both qualitatively and quantitatively assessed through GUS staining and MUG assays. This cautious approach was vindicated as three of the promoters showed unexpected and off-target expression. Further work was undertaken to minimize the size of two of the promoters as well as create a synthetic promoter using the regulatory region of one, both of which were successful. Subsequently, these promoters were used to interrogate *mads69*, a flowering time gene that is expressed across many tissues in maize. With these, we demonstrated that expression in both leaf and embryo tissues induce earlier flowering. In addition, we broadly characterized the traits *mads69* expression impacts which included increased leaf and tassel angles, decreased root mass, and increased number of embryonic leaves. Finally, through DAP-seq we described the DNA-binding behavior of *mads69* which revealed binding at the *vgt* locus of *ZmRap2.7*. Through these studies we have expanded both the biotech toolbox of maize as well as our understanding of the complex role *mads69* plays in maize flowering time.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: A REVIEW OF THE FACTORS CONTROLLING GENE EXPRESSION IN MAIZE

The timing, cell or tissue location, and strength of plant gene expression can be controlled at many, often interacting, levels including at the levels of transcription, translation, post-translational processing, and chromatin remodeling via epigenetic mechanisms. DNA elements currently known to be involved in control of gene expression across those levels include promoters, introns, terminators, untranslated regions, and eu/heterochromatin. As plant genome engineering (transformation) and editing systems have been developed, the need for improved understanding of mechanisms underlying control of gene expression has significantly increased in order to achieve precise, tailored transgene expression for functional genomics research and for biotechnology-based crop improvement efforts. Expression which is not suited for the application can result in defects which limit growth and in extreme cases completely nullify the usefulness of engineered/edited plants for research and genetic enhancement applications. Furthermore, variations within the expression control mechanisms underlie many important traits, the understanding of which is key to making the improvements in crops that are necessary to meet the food, feed, fiber, and biofuel needs of a growing population in a sustainable manner in the face of climate change. Within this changing world, maize is a lynchpin of agriculture and is also a vital research subject in the history of genomics and is the focus of this review. Addressing the plant-related demands of the future will require a holistic understanding of the genetic expression process.

Measuring Gene Expression

"Gene expression" as a term generally refers to the amount of protein produced within a cell or tissue following gene transcription, translation, and processing. Because expression

generally acts as a two-phase process of mRNA being transcribed from DNA and subsequently translated into proteins, different techniques have been developed to analyze these two phases. Further differentiating the approaches to analysis is the relative focus of each on quantitative versus qualitative data. Some strategies toward characterizing gene expression provide a numeric amount of measured expression while others excel at identifying where expression occurs.

The forerunner of mRNA analysis is RNAseq, which has increasingly become more affordable with the advent of next-generation sequencing technologies (Stark et al. 2019). This technique provides a quantitative measure of all mRNA present within the sampled tissue. A more targeted method is quantitative real-time polymerase chain reaction (qRT-PCR) which examines the mRNA levels of a sample relative to that of reference genes (Bustin et al. 2005). A similarly targeted technique is digital droplet PCR which again relies on PCR, but does so on tissue-derived mRNA samples which are partitioned into millions of nanoliter wells (Hindson et al. 2011). Subsequently, expression levels are measured by comparing the number of amplification-positive to negative wells, resulting in the ability to detect only twofold differences (Baker 2012). An older but still occasionally used approach toward gene expression analysis is the use of microarrays and northern blots, both of which provide information on mRNA levels but have seen decreased use due to the development of other methods (Liang and Pardee 1992; Schena et al. 1995; Sekhon et al. 2011). An exciting and more recent development in gene expression analysis is single-cell RNA-seq-based techniques which provide measurements of all mRNAs present in the sample but at single-cell resolution (Giacomello, 2021). This, in part, bridges the gap between quantitative and qualitative approaches since mRNA levels can still be measured as well as tracked to specific cells and tissues in which the mRNA was produced, though it cannot identify sub-cellular localization.

Augmenting expression measurements are a host of other sequencing approaches that reveal the inner workings of genetic systems, focused on transcription factor binding and epigenetic features. Chromatin immunoprecipitation sequencing (ChIP-seq) utilizes antibodies which bind to at-interest transcription factors which are bound to target DNA, which is subsequently sequenced (Barski et al. 2007; Park 2009). This technique has been used to great effect to study the binding activity of histones such as H3K4me2 which is important in understanding the epigenetic landscape (Barski et al. 2007; He et al. 2014). Unfortunately, ChIP-seq relies on the development of antibodies, a labor-intensive practice. Contrastingly, DNA affinity purification sequencing (DAP-seq) only relies *in vitro* expression of the at-interest transcription factor and thus can be done at a much larger scale and has been used in many plants including maize to demonstrate the binding affinities of transcription factors (O'Malley et al. 2016; Bartlett et al. 2017). Finally, sequencing has been used to analyze accessible chromatin regions and includes DNase I hypersensitive sites sequencing (DNAse-seq), formaldehyde-assisted isolation of regulatory elements (FAIRE-seq), and assay for transposase-accessible chromatin with sequencing (ATAC-seq), each of which use different approaches to find what regions of the genome are open (Giresi et al. 2007; Boyle et al. 2008; Buenrostro et al. 2013).

Measurement of protein levels to analyze gene expression levels is divided by transgenic and molecular approaches. Transgenic techniques utilize transgene-mediated expression of proteins encoded by "reporter" genes with known properties, in combination with expression-related genetic elements that are to be tested. One of the most utilized reporter genes in plant gene expression analysis, GUS, encodes β-glucuronidase, which breaks down complex carbohydrate substrates forming visual and/or chemical products that can be measured in both qualitative and quantitative assays (Jefferson et al. 1987; Gallagher 1992). Qualitative

assessments can be made by utilizing 5-bromo-4-chloro-3-indolyl glucuronide (X-Gluc) as a substrate that, when hydrolyzed, is converted into 5,5'-dibromo-4,4'-dichloro-indigo (diX-indigo), a blue precipitate that is generally easily visible by eye and/or light microscopy. Quantitative assays can be conducted utilizing the substrate 4-methylumbelliferyl-β-D-galactopyranoside (4-MUG) that, when cleaved, turns into the fluorescent compound 4-methylumbelliferone (4-MU) which can be measured by a fluorimeter (Gallagher 1992).

Other common reporter genes include fluorescent proteins, luciferase, and proteins which create visible pigments or compounds such as RUBY (Gould and Subramani 1988; Stewart 2001; He et al. 2020). The first nonsubtrate-requiring fluorescent protein encoding gene widely utilized in plant gene expression analysis was green-fluorescent protein (GFP), which was originally isolated from the jelly fish *Aequorea victoria* and subsequently shown to be functional in orange suspension cultures (Shimomura et al. 1962; Prasher et al. 1992; Niedz et al. 1995). Subsequent modification of the GFP gene and discovery of other fluorescent proteins in the class Anthozoa led to an optimization and diversification of fluorescent protein sequences into a variety of excitation and emission spectra which work better in different plant tissues (Rodriguez et al. 2017). These proteins can be easily tracked visually at subcellular, cellular, tissue and whole plant levels making it a pivotal qualitative tool in assessing gene expression. Additionally, the lack of a destructive substrate means these proteins can be tracked in real time in living tissues. Unlike the fluorescent proteins described above, luciferase is not directly luminescent upon exposure to light-mediated excitation but instead must catalyze the reaction of luciferin to enable production of visible fluorescent light. Luciferin is added exogenously to cell/tissue samples enabling real time qualitative and quantitative measurements of luciferase activity (Ow et al. 1986; Xu et al. 2010; Smale 2010; Schaumberg et al. 2015).

Finally, there has been long-time interest using assays for gene expression that rely on expression of nonfluorescent proteins that produce signals that are visible to the naked eye such as anthocyanins and do not require substrates or destructive assays. However, transgene-mediated anthocyanin production can be a complicated process requiring expression of genes encoding multiple enzymes which may be susceptible to tissue and species-specific influences or may have negative effects on plant health when overexpressed, potentially making such marker genes non-ideal as reporters (Misyura et al. 2012; Li et al. 2016). Recently, another visible, pigment-producing reporter gene system for utilization in visual assays of gene expression n is the RUBY reporter gene set which can catalyze the conversion of tyrosine into betalain, a red-colored pigment that can be observed visible in many plants (He et al. 2020) and having reduced/no effect on plant health when overexpressed. However, the RUBY-based systems depend on the presence of tyrosine in sufficient quantities in cells and tissues of interest. If tyrosine levels are not sufficient, then pigment production can be significantly reduced, even when RUBY transgene expression can be very high, leading to misinterpretation of gene expression assay results.

Molecular approaches are also commonly used to study gene expression and resultant protein levels and include immunoassays, mass spectrometry, and electrophoretic separation. All these techniques act on bulked tissue samples and thus are quantitative in nature rather than qualitative. Enzyme-linked immunosorbent assays (ELISAs) have been one of the most popular immunoassay approaches to measure protein levels and do so by utilizing antibodies bound to detectible enzymes. Western blots utilize electrophoresis to separate proteins by their molecular weight and can be used to quantify expression when compared to control, "housekeeping" genes (Mahmood and Yang 2012). Electrophoresis is also used in combination with mass spectrometry

to provide proteome-wide measurements and has been repeatedly used in maize to elucidate the proteome across many tissues and developmental times (Pechanova et al. 2013; Walley et al. 2016; Chao et al. 2016; Hochholdinger et al. 2018).

Each of the above-described techniques provide a different view into the expression behavior of a gene, which is important due to the independent influence transcription and translation have on final protein levels. For instance, in maize, correlations between mRNA and protein levels have been found to be only weakly positive across 23 different tissue types (Walley et al. 2016). Weak correlations have also been observed in other species such as *Arabidopsis*, yeast, and mice (Washburn et al. 2003; Ghazalpour et al. 2011; Mergner et al. 2020). In human-centered studies, detailed quantification of the proteome and transcriptome has shown that ratios of protein-to-mRNA varies by 200-fold across 80% of the genes (Eraslan et al. 2019), demonstrating the large impact the transcription/translation separation can have on determination of gene expression levels. The level of mRNA still holds predictive power, found to explain up to 43% of the protein expression variance in that study, but it does not present the full picture of expression levels. Exploring where control of expression is enacted and how it relates to gene expression in maize will be the focus of the remainder of this review.

Genetic Elements affecting Gene Expression: Core Promoters

Transcription requires the preinitiation complex (PIC) to be assembled at the transcription start site (TSS) of the gene coding sequence. The PIC is composed of RNA Polymerase II (RNAPII), general transcription factors (TFs) (Mediator, TFIID, TFIIB, TFIIA, TFIIF, TFIIE and TFIIH), and a variety of other transcription cofactors such as TATA binding proteins (TBP) and TBP-associated factors which interact with gene-specific transcription factors

(Hong 2016). These TFs recognize and bind to their cognate cis-regulatory elements (CRE). Once assembled, transcription enters the initiation phase, the transcriptional bubble is formed, and the first few RNA molecules are attached. Bringing this assembly to the TSS is mediated largely by the immediate upstream DNA sequence, called the "core promoter" of the gene.

Typically, for this DNA element sequence the broader term "promoter" is used, which refers to the whole 1kb to 2kb region upstream of the TSS. However, this also can contain upstream enhancer and silencer sequences that impact expression but are not sufficient to initiate transcription on their own and will be addressed separately (Juven-Gershon and Kadonaga 2010). Core promoters in mammals, yet to be clearly defined in plants, establish a baseline level of expression which happens in "bursts" of activity as transcription proceeds for a while only to quickly collapse. This baseline activity level is then modified by acting TFs which regulate the frequency of these bursts (Chubb et al. 2006; Bartman et al. 2016; Larsson et al. 2019). Several motifs which regulate this activity have been identified in plants.

TATA box, Y patch, initiator region (Inr), and CCAAT box are all elements that have been associated with core promoters and transcription. One of the first conserved motifs described, the TATA box, has a DNA sequence pattern TATA(A/T)A(A/T) which is bound by TFIID and other TFs resulting in RNAPII recruitment (Kwak and Lis 2013). Like many other plant species, TATA boxes are not common in maize and are only present in 14% of genes (Mejía-Guerra et al. 2015). Occupying a space nearly covering the TSS is the Inr, which, likewise, can recruit the TFIID and has been noted in plants as having transcription enhancing properties with a consensus sequence of yyyyTCAyyy (Zhu et al. 1995; Srivastava et al. 2014; Jores et al. 2021). Hypothesized to impact transcription efficiency is the CCAAT box, which has been widely found in plant promoters and seems to generally play a transcription enhancing role

(Laloum et al. 2013; Porto et al. 2014). Y patches, also called TC-motifs, have been found to increase core promoter expression by 10-15% (Jores et al. 2021).

The characteristics of the core promoter, in combination with other acting elements, create the placement of the TSS. This is not a precise placement however, and variation exists within each gene. In contrast to this prevalence of broad TSS seen in other plants, maize is dominated by precise TSS (Mejía-Guerra et al. 2015). These precise TSS are found even more enriched among genes that are expressed tissue-specifically, though given that the majority of maize genes have a precise TSS this holds little predictive value. Additionally, motif enrichment between the two promoter types was analyzed and TATA-boxes were found enriched in precise TSS, and GC-rich regions were enriched in broad TSS. However, this separation was not definitive and shows that much remains to be understood about what defines TSS selection. This differs strongly from metazoan promoters, which have strong correlations between TSS shape and motifs (Lenhard et al. 2012).

The core promoter sequence, when isolated, is often sufficient for a low level of gene expression and is thus sometimes called the “minimal promoter.” The minimal promoter sequences can be combined with characterized enhancers to create novel synthetic promoters (Huang et al. 2021). While many minimal promoter sequences have been validated within plants, most relevant for maize is the Cauliflower mosaic virus (CaMV) 35S minimal promoter and the maize Ubiquitin 1 minimal promoters (Odell et al. 1985; Kumar et al. 2015; Ali and Kim 2019). Very commonly used is the CaMV35S minimal promoter, which covers from -46 to +1 of the promoter sequence and is sufficient for expression when combined with enhancers. This unit is useful for transgene expression in dicot plants, but does not express strongly in monocots, though the exact mechanisms of why have not been elucidated (Amack and Antunes 2020). More

directly in maize, a 216nt minimal segment from the *ZmUbi1* promoter was isolated and used to create bidirectional promoters for use in plants (Kumar et al. 2015).

Epigenetic Factors

Prior to assembly of the PIC, the chromatin conformation of the promoter region must be accessible to interacting factors and the DNA needs to be in an unmethylated state. Regions of the genome which contribute to PIC assembly have generally been divided into enhancers and promoters, with promoters being proximal to the gene and enhancers being distal. In addition to this, activity levels are separated into active, poised, and repressed states. Challenging this separation, however, are results from research in the mammalian world showing enhancers with promoter activity and promoters with enhancing activity (Andersson and Sandelin 2020). Consequently, speaking about a "range of activities" is becoming the model for describing portions of the genome. Regardless, certain histone modifications have been found correlated with active gene expression regions and some tendencies found which separate enhancers and promoters.

Both enhancers and promoters that are active are associated with accessible-chromatin regions (ACRs) in both mammalian and plant systems (Li et al. 2007; Zhang et al. 2012; Thurman et al. 2012; Weber et al. 2016). Active enhancer sequences are marked by acetylation of histones (HAc) such as H3K9ac, H3K56ac, and H3K27ac as well as H2A.Z (Oka et al. 2017; Ricci et al. 2019; Sun et al. 2019). Chromatin markers of promoter regions have been found to have HAc and H2A.Z as well but also H3K4me3 with occupancy of RNAPII being the most discerning marker (Oka et al. 2017; Peng et al. 2019). Shared between promoters and enhancers (in both mammals and plants) is the repressive activity of H3K27me3 (Mikkelsen et al. 2007;

Shlyueva et al. 2014; Lu et al. 2019), with the same behavior found in maize (Oka et al. 2017; Peng et al. 2019; Ricci et al. 2019). While many studies on enhancers and promoters have been conducted in mammalian systems, caution needs to be exercised when comparing conclusions from those investigations to enhancer and promoter sequences and functions in plant systems. For example, H3K4me1 has been found to mark human active enhancers but not enhancers in *Arabidopsis* or rice (Heint*Zm*an et al. 2009; Sun et al. 2019; Yan et al. 2019).

Least well defined in current literature are "poised" enhancers and promoters. These have markers of both repressed and active sequence regions and likely exist in a transitory state for further alteration (Rada-Iglesias et al. 2010; Koenecke et al. 2017). Poised promoters have been studied in some systems where expression changes across development and found to be associated with RNAPII and changing HAc (Offermann et al. 2008; Gagete et al. 2011). Meanwhile, poised enhancers have low HAc levels and high H3K27me3 but also ACRs (Koenecke et al. 2017; Ricci et al. 2019; Sun et al. 2019). Chromatin conformation changes are highly associated with different developmental stages and environmental conditions and these poised genome pieces might be an essential part of those transitions (Marand et al. 2022).

In addition to these chromatin conditions, active regions of the genome are also generally unmethylated, with studies in maize confirming the trend (Crisp et al. 2020). Interestingly, these regions of unmethylation are relatively stable in maize across diverse tissue types, different from what is observed in mammalian systems. This stability also extends to stress conditions, with methylation remaining relatively unchanged due to heat, cold, and UV stress (Eichten and Springer 2015). DNA methylation in plants is *de novo* established via RNA-directed DNA methylation and it is maintained by a host of methyltransferases (Zhang et al. 2018). Most significant for transgenic approaches toward studying and manipulating gene expression is

homology-dependent gene silencing, which can induce silencing in transgenes in plants but can be managed (Rajeevkumar et al. 2015).

Most TFs are dependent on chromatin conditions to access DNA and enact their function. There are, however, a host of TFs called “pioneer TFs” which act on nucleosome bound and even methylated DNA (Zaret and Mango 2016). An excellent example of a pioneer TF in plants is LFY, which has the capacity to bind to methylated and nucleosome encircled DNA and instigate the floral transition (Jin et al. 2021; Lai et al. 2021). This pioneer activity has not been demonstrated specifically in maize but the maize homologs of the genes, ZFL1 and ZFL2, have been shown to play similar roles in flowering (Bomblies et al. 2003; Bomblies and Doebley 2006). Further pioneer TF genes identified in plants include LEAFY COTYLEDON, APETALA1, and SEPALLATA3 each of which influence flowering time in *Arabidopsis* (Pajoro et al. 2014; Tao et al. 2017).

Directly controlling these chromatin remodeling-related factors is increasingly possible using CRISPR/Cas9-based systems where deactivated (does not cut DNA) Cas9 protein been fused with methylation and chromatin remodelling-associated proteins. (Vojta et al. 2016; Xu et al. 2016; Fal et al. 2021). Outside of direct manipulation, transgenic approaches interact with these genetic-environmental factors most often due to position effects as a result of T-DNA insertion from *Agrobacterium*-based methods. These integration points have been investigated in Brachypodium, *Arabidopsis*, and rice and it has been found that transgene genomic insertion events favor transcriptionally active areas of the genome (Chen et al. 2003; Strizhov et al. 2003; Jeong et al. 2006a; Thole et al. 2010). However, subsequent studies have postulated that this observed phenomenon is due to the selection that occurs during the plant tissue culture and transformation process as transgenic plants under no selection pressure show insertion points

with a random distribution, instead of concentrated in euchromatin (Kim et al. 2007). Additionally, while the insertion point has been shown to be random, it has been found to not be a significant contributor to variation compared to other sources of variation such as somoclonal variation and greenhouse conditions in soybean and maize (Betts et al. 2019).

Enhancers and Silencers

Once the genome is accessible, most TFs can bind to their cognate CREs and contribute to gene expression. While generally separated into repressive or activating factors, it is increasingly apparent that TFs can act as both in a context dependent manner. An example is the plant TF *WUSCHEL* which primarily acts as a repressor but can act as an enhancer of transcription of the flowering-time gene *AGAMOUS* (Ikeda et al. 2009). The DNA motifs that TFs bind and recognize are short, only 6-10 base pairs long as identified in *Arabidopsis* (Franco-Zorrilla et al. 2014). On its own, 6 base pairs of recognition might be insufficient specificity since that motif at random would occur at a rate of 1 out of 1,296 base pairs. However, other mediating factors enhance the specificity of binding motifs, such as epigenetic factors (as already discussed) or through the formation of protein complexes. Many TFs such as LFY or the MADS-BOX family form homo- and heterodimers with other TFs, which increase the specificity of binding motifs (Davies et al. 1996; Hamès et al. 2008; Sayou et al. 2014). Additionally, TFs rarely act alone and instead are in concert with coregulators and other TFs which can impact which sites they recognize. Sites of TF binding can be identified by examining ACRs through ATAC-seq and MNase-seq, both of which have been used to great effect in maize to identify putative binding sites across many tissues (Oka et al. 2017; Ricci et al. 2019; Parvathaneni et al. 2020; Marand et al. 2022). Revealed by these inquiries is the vast distance that some CREs can

be away from the promoter sequences they influence with elements as far away as 100kb being found (Ricci et al. 2019). An example of this is the flowering time regulatory gene *ZmRap2.7* which is regulated by the *VGT* locus located more than 70kb away (Salvi et al. 2007).

Untranslated Regions

A common feature of plant genes is an untranslated sequence region at either the 5’ or 3’ end of the gene. These untranslated regions (UTRs) represent a large energy commitment, composing 18-19% of transcribed sequences in rice and *Arabidopsis* (Srivastava et al. 2018). Within maize, the median 5’-UTR length was 154 nt, running counter to previously established trends of genome size and 5’-UTR length (Liu et al. 2012; Mejía-Guerra et al. 2015). Like promoters, UTRs can be the source of CREs that influence transcription, but most of their influence is on post-transcriptional processes including riboswitches, secondary structure, alternative polyadenylation (APA), short open reading frames in the 5'-UTR (uORF), and nonsense mediated decay.

Transcript maturation is a multi-phased process entailing 5’ capping of the mRNA, splicing of introns, and polyadenylation of the 3’ end of the transcript. Polyadenylation location can vary within a single transcript, with 76% of genes in maize demonstrating APA, similar to other higher plants (Jafar et al. 2019). However, these genes will typically display a primary transcript with alternatives making up a smaller proportion of the population. Sequence-wise, polyadenylation sites display a characteristic AAUAA motif with some variants found in maize and other plants (Li and Du 2014; Wang et al. 2018). These alternative sites can fall within the 3’-UTR itself or upstream in exons or introns, having significant impact on protein structure (Zhang et al. 2021). The length and content of the 3’-UTR is important because longer 3’-UTRs

exhibit greater secondary structure and more sites for micro RNAs, which triggers nonsense mediated decay, and the content can include RNA elements recognized byRNA-binding proteins which can affect mRNA stability (Millevoi and Vagner 2010; Berkovits and Mayr 2015; Mayr 2016; Tian and Manley 2016). Many APAs are differentially regulated, with polyadenylation sites changing in response to developmental pathways or in different tissues. A good example of this is the *Flowering Locus C*-mediated flowering process, which is regulated in part by *FPA* inducing APA in the antisense RNA of *FLC*, which induces *FLC* expression (Hornyik et al. 2010).

Messenger RNA has a 3-dimensional structure formed by the interactions between nucleic acids within the RNA strand. Within the 5'-UTR, this is thought to impact translation efficiency, measured in protein per mRNA per hour, with as high as 50-fold changes in efficiency caused by different 5'-UTRs (Babendure et al. 2006; Kudla et al. 2009; Kim et al. 2014; Cuperus et al. 2017). The 3-dimensional structure can also bind directly to small molecules, called a riboswitch, which results in structural changes in the RNA molecule. One of the most well-studied riboswitches are those sensitive to thiamin pyrophosphate (TPP), a derivative of vitamin B1 and is present in most species, including maize (Wachter et al. 2007; Yadav et al. 2015). Specifically, THIC catalyzes a key step in the thiamine synthesis pathway which results in TPP creation. When TPP is present, it binds to the 3'-UTR and makes available a splice site which otherwise is inaccessible due to the 3D conformation of the transcript. This splicing removes the primary polyadenylation site and leaves a longer 3'-UTR which is more unstable (Sudarsan et al. 2003; Bocobza et al. 2007; Wachter et al. 2007). Usage of riboswitches to control expression has been successful in *Arabidopsis* and the alga *Chlamydomonas*. In *Arabidopsis*, riboswitches sensitive to theophylline have been successfully used to regulate

expression (Shanidze et al. 2020). Additionally, mutation of the *THI4* TPP riboswitch in *Chlamydomonas reinhardtii* was used to create a variety of riboswitches sensitive to many different chemicals (Mehrshahi et al. 2020).

Upstream open reading frames (uORFs) are alternative reading frames present in the 5'-UTR. These are very common in plants, with 24% of rice and 30% of *Arabidopsis* cDNAs containing them (Shashikanth et al. 2008). The presence of these sequences on average reduce expression such as within humans, where uORFs reduce protein expression by 30 to 80% and mRNA levels by 30% on average (Calvo et al. 2009). The qualities of uORFs that determine their impact include the start codon's composition, the length of the uORF and the 5'-UTR, and the length of the alternative transcript created (Hayden and Jorgensen 2007).

Regulation of these sequences are generally associated with developmental and environmental responses which utilize uORFs to modify expression. In maize, drought stress enhances uORF usage, changing the expression landscape to deal with the stress (Lei et al. 2015). Within *Arabidopsis*, the ATB2/AtbZIP11 gene has a uOFR which is activated in the presence of sucrose, reducing expression levels (Wiese et al. 2004). Ignoring the potential impact of uORFs, especially in bZIP genes where they are common, can have unintended consequences as was found in transgenic banana plants overexpressing MusabZIP53 which experienced growth retardation due to coregulation of other genes through the uORF (Shekhawat and Ganapathi 2014). This can be avoided by excising uORFs from transgene sequences as has been done in tomato plants engineered with ZIP genes, though the exciting possibility to take advantage of this phenomena has not yet been realized.

Introns

Most genetic sequences are only relevant within one step of the transcription to translation paradigm; introns stand out as they touch all phases of the gene expression process, ranging from transcription initiation, transcript stability, to translational efficiency of gene expression (Gallegos and Rose 2015; Laxa 2017; Shaul 2017; Dwyer et al. 2021). While introns are broadly defined as any nucleotide segments that are not expressed in a final gene product, there are only four proposed families of introns including spliceosomal, tRNA, self-splicing group I, and self-splicing group II. Expression-enhancing introns have several properties that make them interesting research subjects as well as useful genetic parts. The expression enhancing activity of introns is not tied to specific promoters and can be used independent of its generating sequences (Rose 2002). Important to note is that the expression enhancing properties of introns will not necessarily respect the tissue-specificity of the promoter it is paired with and can, for example, instead enhance expression in all tissues (Jeong et al. 2006b; Emami et al. 2013). This is contingent on the mode of activity of the intron sequences with transcriptional enhancers being more likely to enhance broadly than translational enhancers such as that provided by the AtMHX intron which should have less of an impact in off-target tissues (Akua and Shaul 2013). Ultimately, introns are sequences destined to be excised from the final mRNA transcript, but as part of the transcription and post-transcription processes may shape the behavior of the transcript.

At the simplest level, introns can house CREs which influence gene expression levels. Notable about these sequences are that their expression enhancement is independent of their location relative to the promoter and transcript, first demonstrated in the mouse immunoglobulin heavy chain gene (Mercola et al. 1983; Banerji et al. 1983; Gillies et al. 1983). Within *Arabidopsis*, an example is the *AGAMOUS* gene, whose expression was found to be mediated by

the second intron sequence being acted on by LFY in concert with WUS (Sieburth and Meyerowitz 1997; Busch et al. 1999; Lohmann et al. 2001). Genome-wide detection of CREs in introns in *Arabidopsis* has been performed multiple times by leveraging sequence, expression, and open chromatin data (Parra et al. 2011; Gallegos and Rose 2015; Rose 2019; Meng et al. 2021; Back and Walther 2021). It seems that CREs that act in a tissue-specific manner are housed within introns (Borsari et al. 2021). While many of these are purely bioinformatic inquiry, transgenic assays have confirmed the location-independent effect of some of the CREs identified (Meng et al. 2021). Within maize, the knotted1 gene was originally traced from a mu-insertion event to an intron, which was shown to negatively regulate expression of the gene (Vollbrecht et al. 1991; Greene et al. 1994). Furthermore, the knotted1 gene as a transcription factor has been shown to act on its own intron in a self-regulatory fashion as well as the introns of other genes (though not exclusively on introns) (Bolduc and Hake 2009; Bolduc et al. 2012).

This enhancement of transcription from introns is tied to the distance of the intron from the TSS (Snowden et al. 1996; Rose et al. 2008). This can be seen clearly in intron-containing UTRs, with 5'-UTRs having many more introns than 3'-UTRs (Chung et al. 2006). A study of the TRP1 enhancing introns in *Arabidopsis* showed the enhancement effect disappeared when the intron was more than 1.1kb away from the TSS (Rose 2004). Based on this tendency, the IMEter algorithm was developed to calculate how similar a given intron is to promoter-proximal ones (Rose et al. 2008; Gallegos and Rose 2015) but has had mixed results when applied to practical applications in rice (Morello et al. 2011). Complicating this is the variety of mechanisms by which introns can influence expression. Even in the TRP1 study a difference in mechanism was noted for the four introns studied (Rose 2004). Thus, this general tendency of intron-mediated transcription enhancement being tied to the TSS should not be treated as a rule.

Outside of the expected CRE enhancement pathways is a host of expression-enhancement broadly termed “intron mediated enhancement” (IME) and are dependent on splicing to function. First studied in maize, the intron of the Adh1 gene was found to enhance expression 50 to 100 fold (Callis et al. 1987). In the intervening 35 years, much has been learned about the mode of actions IME has by carefully studying the activity of select introns. At the simplest level, splicing alone is enough to provide enhancing activity with the 5’ splice site (5’ SS) being sufficient in human cells to enhance expression (Damgaard et al. 2008). Mutations in the 5’ SS can reduce transcription on its own (Furger et al. 2002). This is due to U1 small nuclear RNA (snRNA), the DNA-recognizing segment of the RNA-protein complex that is the spliceosome. Transcription initiation is enhanced due to the U1 snRNA recruiting TFIIH near the TSS, which phosphorylates the large subunit carboxyl-terminal domain of RNPII, promoting initiation (Kwek et al. 2002; Akhtar et al. 2009; Glover-Cutter et al. 2009). Transcripts are also protected from premature cleavage and polyadenylation by U1 snRNA, seemingly independent of its role in splicing (Kaida et al. 2010; Berg et al. 2012; Almada et al. 2013). Sites of U1 snRNA binding do not have to be exact matches, with binding being found in similar motifs which coincidentally are enriched in introns (Engreitz et al. 2014).

In addition to U1 snRNA, other splicing factors have been shown to enhance expression through interactions with spliceosome components. Human transcription elongation factor TAT-SF1 stimulates transcription elongation and is an interaction partner with the spliceosome U1 and U2 small nuclear ribonucleoproteins (Fong and Zhou 2001). SRSF1 and SRSF2, two splicing regulators, have also been implicated in transcription in mammalian cells. Specifically at the elongation phase, SRSF2 depletion was found correlated with insufficient recruitment of positive transcription elongation factor b to the RNPII (Lin et al. 2008).

The other major route through which introns can influence expression levels is through the exon-junction complex (EJC) which has many impacts by its presence. Firstly, EJC enhances export to the cytosol for translation through its interaction. The rate of mRNA export is up to 6 to 10 fold for mRNA that has been spliced compared to cDNA (Valencia et al. 2008). Secondly, the EJC has been found to make mRNA more stable, with correlations between intron number and mRNA stability found in mammals and *Arabidopsis* (Narsai et al. 2007; Sharova et al. 2009; Duan et al. 2013). The mode of action appears to be the EJC core component Y14, which interacts with the 5' cap as well as decapping factor Dcp2, which inhibits mRNA-decapping (Chuang et al. 2013). Finally, the EJC and related peripheral factors appear to enhance translation through the recruitment of the translation machinery (Wiegand et al. 2003; Nott et al. 2004; Lee et al. 2009; Hir et al. 2015).

IME also is contingent on position of the intron relative to other genomic regions, with misplacement of the intron causing gene expression repercussions. The movement of an intron in *Arabidopsis* resulted in novel TSS creation (Gallegos and Rose 2017). This also extends to sense vs antisense transcript expression as U1 snRNA binding sites need to be in the sense orientation of the TSS for proper expression (Almada et al. 2013). Finally, if introns are positioned more than 50 nucleotides downstream of the termination sequence in *Arabidopsis* and mammals, they can trigger nonsense-mediated decay, halting expression altogether (Zhang et al. 1998; Nyikó et al. 2013; Shaul 2015).

These points about IME may give the impression that introns uniformly enhance expression by their presence, however, this is not the case. Efficiently spliced introns present in the same location will have very different impacts on expression (Rose 2002). Analysis of introns proximal to the TSS have resulted in the identification of two sequence motifs (CGATT

and TTNGATYTG), which were subsequently tested and found to enhance expression even controlling for intron efficiency (Parra et al. 2011; Rose et al. 2016). An additional 35 nt motif has been identified in the maize Sh1 gene, which is known to enhance expression, reliant on splicing occurring, but does not mediate the splicing process itself (Vasil et al. 1989; Clancy and Curtis Hannah 2002). For many introns, though, the exact mechanism of its' IME is not elucidated. A very widely used enhancing intron, intron 1 derived from the *ZmUbi1* gene, is known to enhance transgene expression in monocots and will even act as an independent promoter, but the specific mechanism of how it does this has not been characterized (Christensen et al. 1992; Cornejo et al. 1993; Salgueiro et al. 2000). Introns represent a great deal of potential for control of gene expression; however, further research is required to reach that potential.

Terminators

While the termination sequence of a gene signals the end of transcription it can also potentially affect the level of gene expression. Most studies on terminator sequences also include the 3'-UTR which complicates the interpretation of their effect. Regardless, in maize several expression-enhancing terminators have been identified, derived from a variety of monocot species. Useful from a design perspective, terminators appear to act independently of promoters in their enhancing activity, enabling pairing between the two parts in a transgene vector to be based on design intentions rather than necessity (Sarrion-Perdigones et al. 2013).

Mechanistically, terminators appear to enhance expression through the action of TFs that bind to the sequences and the ability to interact with promoter sequences. A common factor of control in mammals is pausing of the RNAPII once assembled (Jonkers and Lis 2015). However, in plants this does not seem nearly as significant, with studies in *Arabidopsis* showing much

more pausing occurring at the 3’ end of the transcript (Hetzel et al. 2016). This may be due to the 3D structure DNA forms when transcription is active, during which, the terminator is near the TSS (Misteli 2007; Al Husini et al. 2013). This is mediated by the plethora of RNAPII TFs and terminator TFs which have been found to interact with one another (Al-Husini et al. 2020). As a result, expression is enhanced through polymerase reinitiation (Tan-Wong et al. 2012; Al Husini et al. 2013).

A Holistic Approach to Plant Transgene Vector Design

The sheer variety of mechanisms through which control of gene expression can be exerted can be overwhelming when designing optimal expression cassettes for transgenic and gene editing plant applications. General principles to keep in mind when designing plant transgene expression cassettes include:

1. Clear definition of expression needs (timing, location and strength of expression) is essential prior to vector assembly. What expression pattern is ideal for the application and what is acceptable? A vision of what type of expression is needed will greatly aid in vector design.
2. Transgene constructs should not be unnecessarily complicated. Elements can interact in unpredictable ways and may result in undesirable expression. Do not assume something will work in combination unless it has been demonstrated.
3. Consider what simplifications have been made to the transgene sequence/transcript. Commonly, cDNA is expressed instead of full genomic transcripts (which would include introns). The removal of regulatory information might impact how well a transcript is expressed.

4. Knowledge of the control of gene expression is steadily advancing, and new information is available. Staying abreast of even commonly used gene expression components is vital.

Now is a time of constantly expanding tools/components for optimal design of transgene/editing vectors. Simple overexpression vectors and components of the past are giving way to tissue- and timing-specific expression in plants beyond the *Arabidopsis* model (Nuccio 2018). As more expression cassette components are characterized, there become more opportunities for creation of synthetically designed parts, optimally designed for their applications (Ali and Kim 2019). The application of CRISPR-Cas9 for direct modification of expression to generate novel variation is ever broadening (Rodríguez-Leal et al. 2017). The advancement of single-cell sequencing techniques is enabling more precise examinations of gene expression in plants and will continue to improve as techniques are refined. These advancements and others will enable plant scientists to accelerate and improve plant functional genomics research and applications toward genetic improvement of crops.

References

Akhtar MS, Heidemann M, Tietjen JR, et al (2009) TFIIH Kinase Places Bivalent Marks on the Carboxy-Terminal Domain of RNA Polymerase II. Mol Cell 34:387–393. https://doi.org/10.1016/J.MOLCEL.2009.04.016

Akua T, Shaul O (2013) The Arabidopsis thaliana MHX gene includes an intronic element that boosts translation when localized in a 5′ UTR intron. J Exp Bot 64:4255–4270. https://doi.org/10.1093/JXB/ERT235

Al-Husini N, Medler S, Ansari A (2020) Crosstalk of promoter and terminator during RNA polymerase II transcription cycle. Biochim Biophys Acta - Gene Regul Mech 1863:194657. https://doi.org/10.1016/J.BBAGRM.2020.194657

Al Husini N, Kudla P, Ansari A (2013) A Role for CF1A 3′ End Processing Complex in Promoter-Associated Transcription. PLOS Genet 9:e1003722. https://doi.org/10.1371/JOURNAL.PGEN.1003722

Ali S, Kim WC (2019) A Fruitful Decade Using Synthetic Promoters in the Improvement of Transgenic Plants. Front Plant Sci 10:1–14. https://doi.org/10.3389/fpls.2019.01433

Almada AE, Wu X, Kriz AJ, et al (2013) Promoter directionality is controlled by U1 snRNP and polyadenylation signals. Nat 2013 4997458 499:360–363. https://doi.org/10.1038/nature12349

Amack SC, Antunes MS (2020) CaMV35S promoter – A plant biology and biotechnology workhorse in the era of synthetic biology. Curr Plant Biol 24:100179. https://doi.org/10.1016/J.CPB.2020.100179

Andersson R, Sandelin A (2020) Determinants of enhancer and promoter activities of regulatory elements. Nat Rev Genet 21:71–87. https://doi.org/10.1038/s41576-019-0173-8

Babendure JR, Babendure JL, Ding JH, Tsien RY (2006) Control of mammalian translation by mRNA structure near caps. RNA 12:851–861. https://doi.org/10.1261/RNA.2309906

Back G, Walther D (2021) Identification of cis-regulatory motifs in first introns and the prediction of intron-mediated enhancement of gene expression in Arabidopsis thaliana. BMC Genomics 22:1–24. https://doi.org/10.1186/S12864-021-07711-1/FIGURES/8

Baker M (2012) Digital PCR hits its stride. Nat Methods 2012 96 9:541–544. https://doi.org/10.1038/NMETH.2027

Banerji J, Olson L, Schaffner W (1983) A lymphocyte-specific cellular enhancer is located downstream of the joining region in immunoglobulin heavy chain genes. Cell 33:729–740. https://doi.org/10.1016/0092-8674(83)90015-6

Barski A, Cuddapah S, Cui K, et al (2007) High-resolution profiling of histone methylations in the human genome. Cell 129:823–837. https://doi.org/10.1016/J.CELL.2007.05.009

Bartlett A, O'Malley RC, Huang SSC, et al (2017) Mapping genome-wide transcription-factor binding sites using DAP-seq. Nat Protoc 2017 128 12:1659–1672. https://doi.org/10.1038/nprot.2017.055

Bartman CR, Hsu SC, Hsiung CCS, et al (2016) Enhancer Regulation of Transcriptional Bursting Parameters Revealed by Forced Chromatin Looping. Mol Cell 62:237–247. https://doi.org/10.1016/J.MOLCEL.2016.03.007

Berg MG, Singh LN, Younis I, et al (2012) U1 snRNP Determines mRNA Length and Regulates Isoform Expression. Cell 150:53–64. https://doi.org/10.1016/J.CELL.2012.05.029

Berkovits BD, Mayr C (2015) Alternative 3′ UTRs act as scaffolds to regulate membrane protein localization. Nat 2015 5227556 522:363–367. https://doi.org/10.1038/nature14321

Betts SD, Basu S, Bolar J, et al (2019) Uniform Expression and Relatively Small Position Effects Characterize Sister Transformants in Maize and Soybean. Front Plant Sci 10:1–20. https://doi.org/10.3389/fpls.2019.01209

Bocobza S, Adato A, Mandel T, et al (2007) Riboswitch-dependent gene regulation and its evolution in the plant kingdom. Genes Dev 21:2874–2879. https://doi.org/10.1101/GAD.443907

Bolduc N, Hake S (2009) The Maize Transcription Factor KNOTTED1 Directly Regulates the Gibberellin Catabolism Gene ga2ox1. Plant Cell 21:1647–1658. https://doi.org/10.1105/TPC.109.068221

Bolduc N, Yilmaz A, Mejia-Guerra MK, et al (2012) Unraveling the KNOTTED1 regulatory network in maize meristems. Genes Dev 26:1685–1690. https://doi.org/10.1101/GAD.193433.112

Bomblies K, Doebley JF (2006) Pleiotropic Effects of the Duplicate Maize FLORICAULA/LEAFY Genes zfl1 and zfl2 on Traits Under Selection During Maize Domestication. Genetics 172:519–531. https://doi.org/10.1534/GENETICS.105.048595

Bomblies K, Wang R-L, Ambrose B, et al (2003) Duplicate FLORICAULA/LEAFY homologs zfl1 and zfl2 control inflorescence architecture and flower patterning in maize. Development 130:2385–2395. https://doi.org/10.1242/dev.00457

Borsari B, Villegas-Mirón P, Pérez-Lluch S, et al (2021) Enhancers with tissue-specific activity are enriched in intronic regions. Genome Res 31:1325–1336. https://doi.org/10.1101/GR.270371.120/-/DC1

Boyle AP, Davis S, Shulha HP, et al (2008) High-Resolution Mapping and Characterization of Open Chromatin across the Genome. Cell 132:311. https://doi.org/10.1016/J.CELL.2007.12.014

Buenrostro JD, Giresi PG, Zaba LC, et al (2013) Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat Methods 2013 1012 10:1213–1218. https://doi.org/10.1038/nmeth.2688

Busch MA, Bomblies K, Weigel D (1999) Activation of a floral homeotic gene in Arabidopsis. Science (80-) 285:585–587. https://doi.org/10.1126/SCIENCE.285.5427.585/ASSET/863C10CC-A8C4-49A3-92E2-4127C71D0642/ASSETS/GRAPHIC/SE2897682004.JPEG

Bustin SA, Benes V, Nolan T, Pfaffl MW (2005) Quantitative real-time RT-PCR - A perspective. J Mol Endocrinol 34:597–601. https://doi.org/10.1677/jme.1.01755

Callis J, Fromm M, Walbot V (1987) Introns increase gene expression in cultured maize cells. Genes Dev 1:1183–1200. https://doi.org/10.1101/GAD.1.10.1183

Calvo SE, Pagliarini DJ, Mootha VK (2009) Upstream open reading frames cause widespread reduction of protein expression and are polymorphic among humans. Proc Natl Acad Sci U S A 106:7507–7512. https://doi.org/10.1073/PNAS.0810916106/SUPPL_FILE/0810916106SI.PDF

Chao Q, Gao Z fang, Wang Y feng, et al (2016) The proteome and phosphoproteome of maize pollen uncovers fertility candidate proteins. Plant Mol Biol 2016 913 91:287–304. https://doi.org/10.1007/S11103-016-0466-7

Chen S, Jin W, Wang M, et al (2003) Distribution and characterization of over 1000 T-DNA tags in rice genome. Plant J 36:105–113. https://doi.org/10.1046/j.1365-313X.2003.01860.x

Christensen AH, Sharrock RA, Quail PH (1992) Maize polyubiquitin genes: structure, thermal perturbation of expression and transcript splicing, and promoter activity following transfer to protoplasts by electroporation. Plant Mol Biol 18:675–689. https://doi.org/10.1007/BF00020010/METRICS

Chuang TW, Chang WL, Lee KM, Tarn WY (2013) The RNA-binding protein Y14 inhibits mRNA decapping and modulates processing body formation. Mol Biol Cell 24:1–13. https://doi.org/10.1091/MBC.E12-03-0217/ASSET/IMAGES/LARGE/1FIG7.JPEG

Chubb JR, Trcek T, Shenoy SM, Singer RH (2006) Transcriptional pulsing of a developmental gene. Curr Biol 16:1018–1025. https://doi.org/10.1016/J.CUB.2006.03.092

Chung BYW, Simons C, Firth AE, et al (2006) Effect of 5′UTR introns on gene expression in Arabidopsis thaliana. BMC Genomics 7:1–13. https://doi.org/10.1186/1471-2164-7-120/TABLES/2

Clancy M, Curtis Hannah L (2002) Splicing of the Maize Sh1 First Intron Is Essential for Enhancement of Gene Expression, and a T-Rich Motif Increases Expression without Affecting Splicing. Plant Physiol 130:918–929. https://doi.org/10.1104/PP.008235

Cornejo MJ, Luth D, Blankenship KM, et al (1993) Activity of a maize ubiquitin promoter in transgenic rice. Plant Mol Biol 23:567–581. https://doi.org/10.1007/BF00019304/METRICS

Crisp PA, Marand AP, Noshay JM, et al (2020) Stable unmethylated DNA demarcates expressed genes and their cis-regulatory space in plant genomes. Proc Natl Acad Sci U S A 117:23991–24000. https://doi.org/10.1073/pnas.2010250117

Cuperus JT, Groves B, Kuchina A, et al (2017) Deep learning of the regulatory grammar of yeast 5′ untranslated regions from 500,000 random sequences. Genome Res 27:2015–2024. https://doi.org/10.1101/GR.224964.117

Damgaard CK, Kahns S, Lykke-Andersen S, et al (2008) A 5′ Splice Site Enhances the Recruitment of Basal Transcription Initiation Factors In Vivo. Mol Cell 29:271–278. https://doi.org/10.1016/J.MOLCEL.2007.11.035

Davies B, Egea-Cortines M, De Andrade Silva E, et al (1996) Multiple interactions amongst floral homeotic MADS box proteins. EMBO J 15:4330–4343. https://doi.org/10.1002/J.1460-2075.1996.TB00807.X

Duan J, Shi J, Ge X, et al (2013) Genome-wide survey of interindividual differences of RNA stability in human lymphoblastoid cell lines Motivation Analysis Result Conclusion. https://doi.org/10.1038/srep01318

Dwyer K, Agarwal N, Pile L, Ansari A (2021) Gene Architecture Facilitates Intron-Mediated Enhancement of Transcription. Front Mol Biosci 8:276. https://doi.org/10.3389/FMOLB.2021.669004/BIBTEX

Eichten SR, Springer NM (2015) Minimal evidence for consistent changes in maize DNA methylation patterns following environmental stress. Front Plant Sci 6:1–10. https://doi.org/10.3389/fpls.2015.00308

Emami S, Arumainayagam D, Korf I, Rose AB (2013) The effects of a stimulating intron on the expression of heterologous genes in Arabidopsis thaliana. Plant Biotechnol J 11:555–563. https://doi.org/10.1111/PBI.12043

Engreitz JM, Sirokman K, McDonel P, et al (2014) RNA-RNA Interactions Enable Specific Targeting of Noncoding RNAs to Nascent Pre-mRNAs and Chromatin Sites. Cell 159:188–199. https://doi.org/10.1016/J.CELL.2014.08.018

Eraslan B, Wang D, Gusic M, et al (2019) Quantification and discovery of sequence determinants of protein-per-mRNA amount in 29 human tissues. Mol Syst Biol 15:1–25. https://doi.org/10.15252/msb.20188513

Fal K, Tomkova D, Vachon G, et al (2021) Chromatin Manipulation and Editing: Challenges, New Technologies and Their Use in Plants. Int J Mol Sci 22:1–24. https://doi.org/10.3390/IJMS22020512

Fong YW, Zhou Q (2001) Stimulatory effect of splicing factors on transcriptional elongation. Nat 2001 4146866 414:929–933. https://doi.org/10.1038/414929a

Franco-Zorrilla JM, López-Vidriero I, Carrasco JL, et al (2014) DNA-binding specificities of plant transcription factors and their potential to define target genes. Proc Natl Acad Sci U S A 111:2367–2372. https://doi.org/10.1073/PNAS.1316278111/SUPPL_FILE/SAPP.PDF

Furger A, O'Sullivan JM, Binnie A, et al (2002) Promoter proximal splice sites enhance transcription. Genes Dev 16:2792–2799. https://doi.org/10.1101/GAD.983602

Gagete AP, Franco L, Isabel Rodrigo M (2011) The Pisum sativum psp54 gene requires ABI3 and a chromatin remodeller to switch from a poised to a transcriptionally active state. New Phytol 192:353–363. https://doi.org/10.1111/J.1469-8137.2011.03818.X

Gallagher S (1992) GUS protocols: Using the GUS gene as a reporter of gene expression

Gallegos JE, Rose AB (2015) The enduring mystery of intron-mediated enhancement. Plant Sci 237:8–15. https://doi.org/10.1016/J.PLANTSCI.2015.04.017

Gallegos JE, Rose AB (2017) Intron DNA Sequences Can Be More Important Than the Proximal Promoter in Determining the Site of Transcript Initiation. Plant Cell 29:843–853. https://doi.org/10.1105/TPC.17.00020

Ghazalpour A, Bennett B, Petyuk VA, et al (2011) Comparative analysis of proteome and transcriptome variation in mouse. PLoS Genet 7:. https://doi.org/10.1371/journal.pgen.1001393

Giacomello S (2021) A new era for plant science: spatial single-cell transcriptomics. Curr Opin Plant Biol 60:102041. https://doi.org/10.1016/j.pbi.2021.102041

Gillies SD, Morrison SL, Oi VT, Tonegawa S (1983) A tissue-specific transcription enhancer element is located in the major intron of a rearranged immunoglobulin heavy chain gene. Cell 33:717–728. https://doi.org/10.1016/0092-8674(83)90014-4

Giresi PG, Kim J, McDaniell RM, et al (2007) FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin. Genome Res 17:877–885. https://doi.org/10.1101/GR.5533506

Glover-Cutter K, Larochelle S, Erickson B, et al (2009) TFIIH-Associated Cdk7 Kinase Functions in Phosphorylation of C-Terminal Domain Ser7 Residues, Promoter-Proximal Pausing, and Termination by RNA Polymerase II. Mol Cell Biol 29:5455–5464. https://doi.org/10.1128/MCB.00637-09/SUPPL_FILE/GLOVER_CUTTER_REV_SUPP_LEGENDS.PDF

Gould SJ, Subramani S (1988) Firefly luciferase as a tool in molecular and cell biology. Anal Biochem 175:5–13. https://doi.org/10.1016/0003-2697(88)90353-3

Greene B, Walko R, Hake S (1994) Mutator insertions in an intron of the maize knotted1 gene result in dominant suppressible mutations. Genetics 138:1275–1285. https://doi.org/10.1093/GENETICS/138.4.1275

Hamès C, Ptchelkine D, Grimm C, et al (2008) Structural basis for LEAFY floral switch function and similarity with helix-turn-helix proteins. EMBO J 27:2628–2637. https://doi.org/10.1038/EMBOJ.2008.184

Hayden CA, Jorgensen RA (2007) Identification of novel conserved peptide uORF homology groups in Arabidopsis and rice reveals ancient eukaryotic origin of select groups and preferential association with transcription factor-encoding genes. BMC Biol 5:1–30. https://doi.org/10.1186/1741-7007-5-32/FIGURES/9

He S, Yan S, Wang P, et al (2014) Comparative analysis of genome-wide chromosomal histone modification patterns in maize cultivars and their wild relatives. PLoS One 9:e97364. https://doi.org/10.1371/JOURNAL.PONE.0097364

He Y, Zhang T, Sun H, et al (2020) A reporter for noninvasively monitoring gene expression and plant transformation. Hortic Res 7:1–6. https://doi.org/10.1038/s41438-020-00390-1

Heintzman ND, Hon GC, Hawkins RD, et al (2009) Histone modifications at human enhancers reflect global cell-type-specific gene expression. Nat 2009 4597243 459:108–112. https://doi.org/10.1038/nature07829

Hetzel J, Duttke SH, Benner C, Chory J (2016) Nascent RNA sequencing reveals distinct features in plant transcription. Proc Natl Acad Sci U S A 113:12316–12321. https://doi.org/10.1073/pnas.1603217113

Hindson BJ, Ness KD, Masquelier DA, et al (2011) High-throughput droplet digital PCR system for absolute quantitation of DNA copy number. Anal Chem 83:8604–8610. https://doi.org/10.1021/ac202028g

Hir H Le, Saulière J, Wang Z (2015) The exon junction complex as a node of post-transcriptional networks. Nat Rev Mol Cell Biol 2015 171 17:41–54. https://doi.org/10.1038/nrm.2015.7

Hochholdinger F, Marcon C, Baldauf JA, et al (2018) Proteomics of Maize Root Development. Front Plant Sci 9:. https://doi.org/10.3389/FPLS.2018.00143

Hong JC (2016) General Aspects of Plant Transcription Factor Families. Elsevier Inc.

Hornyik C, Terzi LC, Simpson GG (2010) The Spen Family Protein FPA Controls Alternative Cleavage and Polyadenylation of RNA. Dev Cell 18:203–213. https://doi.org/10.1016/J.DEVCEL.2009.12.009

Huang D, Kosentka PZ, Liu W (2021) Synthetic biology approaches in regulation of targeted gene expression. Curr Opin Plant Biol 63:102036. https://doi.org/10.1016/J.PBI.2021.102036

Ikeda M, Mitsuda N, Ohme-Takagi M (2009) Arabidopsis WUSCHEL Is a Bifunctional Transcription Factor That Acts as a Repressor in Stem Cell Regulation and as an Activator in Floral Patterning. Plant Cell 21:3493–3505. https://doi.org/10.1105/TPC.109.069997

Jafar Z, Tariq S, Sadiq I, et al (2019) Genome-Wide Profiling of Polyadenylation Events in Maize Using High-Throughput Transcriptomic Sequences. G3 Genes|Genomes|Genetics 9:2749–2760. https://doi.org/10.1534/G3.119.400196

Jefferson RA, Kavanagh TA, Bevan MW (1987) GUS fusions: beta-glucuronidase as a sensitive and versatile gene fusion marker in higher plants. EMBO J 6:3901. https://doi.org/10.1002/J.1460-2075.1987.TB02730.X

Jeong DH, An S, Park S, et al (2006a) Generation of a flanking sequence-tag database for activation-tagging lines in japonica rice. Plant J 45:123–132. https://doi.org/10.1111/j.1365-313X.2005.02610.x

Jeong YM, Mun JH, Lee I, et al (2006b) Distinct Roles of the First Introns on the Expression of Arabidopsis Profilin Gene Family Members. Plant Physiol 140:196–209. https://doi.org/10.1104/PP.105.071316

Jin R, Klasfeld S, Zhu Y, et al (2021) LEAFY is a pioneer transcription factor and licenses cell reprogramming to floral fate. Nat Commun 12:1–14. https://doi.org/10.1038/s41467-020-20883-w

Jonkers I, Lis JT (2015) Getting up to speed with transcription elongation by RNA polymerase II. Nat Rev Mol Cell Biol 16:167–177. https://doi.org/10.1038/nrm3953

Jores T, Tonnies J, Wrightsman T, et al (2021) Synthetic promoter designs enabled by a comprehensive analysis of plant core promoters. Nat Plants 7:842–855. https://doi.org/10.1038/s41477-021-00932-y

Juven-Gershon T, Kadonaga JT (2010) Regulation of gene expression via the core promoter and the basal transcriptional machinery. Dev Biol 339:225–229. https://doi.org/10.1016/j.ydbio.2009.08.009

Kaida D, Berg MG, Younis I, et al (2010) U1 snRNP protects pre-mRNAs from premature cleavage and polyadenylation. Nat 2010 4687324 468:664–668. https://doi.org/10.1038/nature09479

Kim SI, Veena, Gelvin SB (2007) Genome-wide analysis of Agrobacterium T-DNA integration sites in the Arabidopsis genome generated under non-selective conditions. Plant J 51:779–791. https://doi.org/10.1111/j.1365-313X.2007.03183.x

Kim Y, Lee G, Jeon E, et al (2014) The immediate upstream region of the 5′-UTR from the AUG start codon has a pronounced effect on the translational efficiency in Arabidopsis thaliana. Nucleic Acids Res 42:485–498. https://doi.org/10.1093/NAR/GKT864

Koenecke N, Johnston J, He Q, et al (2017) Drosophila poised enhancers are generated during tissue patterning with the help of repression. Genome Res 27:64–74. https://doi.org/10.1101/GR.209486.116

Kudla G, Murray AW, Tollervey D, Plotkin JB (2009) Coding-sequence determinants of expression in escherichia coli. Science (80-) 324:255–258. https://doi.org/10.1126/SCIENCE.1170160/SUPPL_FILE/KUDLA.SOM.PDF

Kumar S, AlAbed D, Whitteck JT, et al (2015) A combinatorial bidirectional and bicistronic approach for coordinated multi-gene expression in corn. Plant Mol Biol 87:341–353. https://doi.org/10.1007/S11103-015-0281-6/FIGURES/10

Kwak H, Lis JT (2013) Control of Transcriptional Elongation. https://doi-org.ezproxy.library.wisc.edu/101146/annurev-genet-110711-155440 47:483–508. https://doi.org/10.1146/ANNUREV-GENET-110711-155440

Kwek KY, Murphy S, Furger A, et al (2002) U1 snRNA associates with TFIIH and regulates transcriptional initiation. Nat Struct Biol 2002 911 9:800–805. https://doi.org/10.1038/nsb862

Lai X, Blanc-Mathieu R, GrandVuillemin L, et al (2021) The LEAFY floral regulator displays pioneer transcription factor properties. Mol Plant 14:829–837. https://doi.org/10.1016/J.MOLP.2021.03.004

Laloum T, De Mita S, Gamas P, et al (2013) CCAAT-box binding transcription factors in plants: Y so many? Trends Plant Sci 18:157–166. https://doi.org/10.1016/J.TPLANTS.2012.07.004

Larsson AJM, Johnsson P, Hagemann-Jensen M, et al (2019) Genomic encoding of transcriptional burst kinetics. Nature 565:251–254. https://doi.org/10.1038/S41586-018-0836-1

Laxa M (2017) Intron-mediated enhancement: A tool for heterologous gene expression in plants? Front Plant Sci 7:1977. https://doi.org/10.3389/FPLS.2016.01977/BIBTEX

Lee HC, Choe J, Chi SG, Kim YK (2009) Exon junction complex enhances translation of spliced mRNAs at multiple steps. Biochem Biophys Res Commun 384:334–340. https://doi.org/10.1016/J.BBRC.2009.04.123

Lei L, Shi J, Chen J, et al (2015) Ribosome profiling reveals dynamic translational landscape in maize seedlings under drought stress. Plant J 84:1206–1218. https://doi.org/10.1111/TPJ.13073

Lenhard B, Sandelin A, Carninci P (2012) Metazoan promoters: emerging characteristics and insights into transcriptional regulation. Nat Rev Genet 2012 134 13:233–245. https://doi.org/10.1038/NRG3163

Li B, Carey M, Workman JL (2007) The Role of Chromatin during Transcription. Cell 128:707–719. https://doi.org/10.1016/j.cell.2007.01.015

Li P, Chen B, Zhang G, et al (2016) Regulation of anthocyanin and proanthocyanidin biosynthesis by Medicago truncatula bHLH transcription factor MtTT8. New Phytol 210:905–921. https://doi.org/10.1111/nph.13816

Li XQ, Du D (2014) Motif types, motif locations and base composition patterns around the RNA polyadenylation site in microorganisms, plants and animals. BMC Evol Biol 14:1–24. https://doi.org/10.1186/S12862-014-0162-7/FIGURES/10

Liang P, Pardee AB (1992) Differential Display of Eukaryotic Messenger RNA by Means of the Polymerase Chain Reaction. Science (80-) 257:967–971. https://doi.org/10.1126/SCIENCE.1354393

Lin S, Coutinho-Mansfield G, Wang D, et al (2008) The splicing factor SC35 has an active role in transcriptional elongation. Nat Struct Mol Biol 2008 158 15:819–826. https://doi.org/10.1038/nsmb.1461

Liu H, Yin J, Xiao M, et al (2012) Characterization and evolution of 5′ and 3′ untranslated regions in eukaryotes. Gene 507:106–111. https://doi.org/10.1016/J.GENE.2012.07.034

Lohmann JU, Hong RL, Hobe M, et al (2001) A Molecular Link between Stem Cell Regulation and Floral Patterning in Arabidopsis. Cell 105:793–803. https://doi.org/10.1016/S0092-8674(01)00384-1

Lu Z, Marand AP, Ricci WA, et al (2019) The prevalence, evolution and chromatin signatures of plant regulatory elements. Nat Plants 5:1250–1259. https://doi.org/10.1038/s41477-019-0548-z

Mahmood T, Yang PC (2012) Western Blot: Technique, Theory, and Trouble Shooting. N Am J Med Sci 4:429. https://doi.org/10.4103/1947-2714.100998

Marand AP, Eveland AL, Kaufmann K, Springer NM (2022) cis-Regulatory Elements in Plant Development, Adaptation, and Evolution. Annu Rev Plant Biol 74:. https://doi.org/10.1146/ANNUREV-ARPLANT-070122-030236

Mayr C (2016) Evolution and Biological Roles of Alternative 3′UTRs. Trends Cell Biol 26:227–237. https://doi.org/10.1016/J.TCB.2015.10.012

Mehrshahi P, Nguyen GTDT, Gorchs Rovira A, et al (2020) Development of Novel Riboswitches for Synthetic Biology in the Green Alga Chlamydomonas. ACS Synth Biol 9:1406–1417. https://doi.org/10.1021/ACSSYNBIO.0C00082/SUPPL_FILE/SB0C00082_SI_003.ZIP

Mejía-Guerra MK, Li W, Galeano NF, et al (2015) Core Promoter Plasticity Between Maize Tissues and Genotypes Contrasts with Predominance of Sharp Transcription Initiation Sites. Plant Cell 27:3309–3320. https://doi.org/10.1105/tpc.15.00630

Meng F, Zhao H, Zhu B, et al (2021) Genomic editing of intronic enhancers unveils their role in fine-tuning tissue-specific gene expression in Arabidopsis thaliana. Plant Cell 33:1997–2014. https://doi.org/10.1093/PLCELL/KOAB093

Mercola M, Wang XF, Olsen J, Calame K (1983) Transcriptional Enhancer Elements in the Mouse Immunoglobulin Heavy Chain Locus. Science (80-) 221:663–665. https://doi.org/10.1126/SCIENCE.6306772

Mergner J, Frejno M, List M, et al (2020) Mass-spectrometry-based draft of the Arabidopsis proteome. Nature 579:409–414. https://doi.org/10.1038/s41586-020-2094-2

Mikkelsen TS, Ku M, Jaffe DB, et al (2007) Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature 448:553–560. https://doi.org/10.1038/nature06008

Millevoi S, Vagner S (2010) Molecular mechanisms of eukaryotic pre-mRNA 3′ end processing regulation. Nucleic Acids Res 38:2757–2774. https://doi.org/10.1093/NAR/GKP1176

Misteli T (2007) Beyond the Sequence: Cellular Organization of Genome Function. Cell 128:787–800. https://doi.org/10.1016/J.CELL.2007.01.028

Misyura M, Colasanti J, Rothstein S (2012) Physiological and genetic analysis of Arabidopsis thaliana methylation and chromatin patterning anthocyanin biosynthesis mutants under chronic adverse environmental conditions. J Exp Bot 64:219–240. https://doi.org/10.1093/jxb/err313

Morello L, Gianí S, Troina F, Breviario D (2011) Testing the IMEter on rice introns and other aspects of intron-mediated enhancement of gene expression. J Exp Bot 62:533–544. https://doi.org/10.1093/JXB/ERQ273

Narsai R, Howell KA, Millar AH, et al (2007) Genome-Wide Analysis of mRNA Decay Rates and Their Determinants in Arabidopsis thaliana. Plant Cell 19:3418–3436. https://doi.org/10.1105/TPC.107.055046

Niedz RP, Sussman MR, Satterlee JS (1995) Green fluorescent protein: an in vivo reporter of plant gene expression. Plant Cell Rep 14:403–406. https://doi.org/10.1007/BF00234043

Nott A, Le Hir H, Moore MJ (2004) Splicing enhances translation in mammalian cells: an additional function of the exon junction complex. Genes Dev 18:210–222. https://doi.org/10.1101/GAD.1163204

Nuccio ML (2018) A Brief History of Promoter Development for Use in Transgenic Maize Applications. Methods Mol Biol 1676:61–93. https://doi.org/10.1007/978-1-4939-7315-6_4

Nyikó T, Kerényi F, Szabadkai L, et al (2013) Plant nonsense-mediated mRNA decay is controlled by different autoregulatory circuits and can be induced by an EJC-like complex. Nucleic Acids Res 41:6715–6728. https://doi.org/10.1093/NAR/GKT366

O'Malley RC, Huang S-SC, Song L, et al (2016) Cistrome and Epicistrome Features Shape the Regulatory DNA Landscape. Cell 165:1280–1292. https://doi.org/10.1016/j.cell.2016.04.038

Odell JT, Nagy F, Chua N-H (1985) Identification of DNA sequences required for activity of the cauliflower mosaic virus 35S promoter. Nature 313:810–812

Offermann S, Dreesen B, Horst I, et al (2008) Developmental and Environmental Signals Induce Distinct Histone Acetylation Profiles on Distal and Proximal Promoter Elements of the C4-Pepc Gene in Maize. Genetics 179:1891–1901. https://doi.org/10.1534/GENETICS.108.087411

Oka R, Zicola J, Weber B, et al (2017) Genome-wide mapping of transcriptional enhancer candidates using DNA and chromatin features in maize. Genome Biol 18:1–24. https://doi.org/10.1186/s13059-017-1273-4

Ow DW, Wood K V., Deluca M, et al (1986) Transient and Stable Expression of the Firefly Luciferase Gene in Plant Cells and Transgenic Plants. Science (80-) 234:856–859. https://doi.org/10.1126/SCIENCE.234.4778.856

Pajoro A, Madrigal P, Muiño JM, et al (2014) Dynamics of chromatin accessibility and gene regulation by MADS-domain transcription factors in flower development. Genome Biol 15:. https://doi.org/10.1186/gb-2014-15-3-r41

Park PJ (2009) ChIP–seq: advantages and challenges of a maturing technology. Nat Rev Genet 2009 1010 10:669–680. https://doi.org/10.1038/NRG2641

Parra G, Bradnam K, Rose AB, Korf I (2011) Comparative and functional analysis of intron-mediated enhancement signals reveals conserved features among plants. Nucleic Acids Res 39:5328–5337. https://doi.org/10.1093/NAR/GKR043

Parvathaneni RK, Bertolini E, Shamimuzzaman M, et al (2020) The regulatory landscape of early maize inflorescence development. Genome Biol 2020 211 21:1–33. https://doi.org/10.1186/S13059-020-02070-8

Pechanova O, Tom´ T, Tomáˇ T, et al (2013) Maize proteomics: An insight into the biology of an important cereal crop. Proteomics 13:637–662. https://doi.org/10.1002/PMIC.201200275

Peng Y, Xiong D, Zhao L, et al (2019) Chromatin interaction maps reveal genetic regulation for quantitative traits in maize. Nat Commun 10:1–11. https://doi.org/10.1038/s41467-019-10602-5

Porto MS, Pinheiro MPN, Batista VGL, et al (2014) Plant promoters: An approach of structure and function. Mol Biotechnol 56:38–49. https://doi.org/10.1007/S12033-013-9713-1/TABLES/2

Prasher DC, Eckenrode VK, Ward WW, et al (1992) Primary structure of the Aequorea victoria green-fluorescent protein. Gene 111:229–233. https://doi.org/10.1016/0378-1119(92)90691-H

Rada-Iglesias A, Bajpai R, Swigut T, et al (2010) A unique chromatin signature uncovers early developmental enhancers in humans. Nat 2010 4707333 470:279–283. https://doi.org/10.1038/NATURE09692

Rajeevkumar S, Anunanthini P, Ramalingam S (2015) Epigenetic silencing in transgenic plants. Front Plant Sci 6:693. https://doi.org/10.3389/FPLS.2015.00693

Ricci WA, Lu Z, Ji L, et al (2019) Widespread long-range cis-regulatory elements in the maize genome. Nat Plants 5:1237–1249. https://doi.org/10.1038/s41477-019-0547-0

Rodríguez-Leal D, Lemmon ZH, Man J, et al (2017) Engineering Quantitative Trait Variation for Crop Improvement by Genome Editing. Cell 171:470–480. https://doi.org/10.1016/j.cell.2017.08.030

Rodriguez EA, Campbell RE, Lin JY, et al (2017) The Growing and Glowing Toolbox of Fluorescent and Photoactive Proteins. Trends Biochem Sci 42:111–129. https://doi.org/10.1016/J.TIBS.2016.09.010

Rose AB (2002) Requirements for intron-mediated enhancement of gene expression in Arabidopsis. RNA 8:1444–1453. https://doi.org/10.1017/S1355838202020551

Rose AB (2019) Introns as gene regulators: A brick on the accelerator. Front Genet 10:672. https://doi.org/10.3389/FGENE.2018.00672/BIBTEX

Rose AB (2004) The effect of intron location on intron-mediated enhancement of gene expression in Arabidopsis. Plant J 40:744–751. https://doi.org/10.1111/J.1365-313X.2004.02247.X

Rose AB, Carter A, Korf I, Kojima N (2016) Intron sequences that stimulate gene expression in Arabidopsis. Plant Mol Biol 92:337–346. https://doi.org/10.1007/S11103-016-0516-1/FIGURES/4

Rose AB, Elfersi T, Parra G, Korf I (2008) Promoter-Proximal Introns in Arabidopsis thaliana Are Enriched in Dispersed Signals that Elevate Gene Expression. Plant Cell 20:543–551. https://doi.org/10.1105/TPC.107.057190

Salgueiro S, Pignocchi C, Parry MAJ (2000) Intron-mediated gusA expression in Tritordeum and wheat resulting from particle bombardment. Artic Plant Mol Biol. https://doi.org/10.1023/A:1006331831858

Salvi S, Sponza G, Morgante M, et al (2007) Conserved noncoding genomic sequences associated with a flowering-time quantitative trait locus in maize. Proc Natl Acad Sci 104:11376–11381. https://doi.org/10.1073/pnas.0704145104

Sarrion-Perdigones A, Vazquez-Vilar M, Palací J, et al (2013) Goldenbraid 2.0: A comprehensive DNA assembly framework for plant synthetic biology. Plant Physiol 162:1618–1631. https://doi.org/10.1104/pp.113.217661

Sayou C, Monniaux M, Nanao MH, et al (2014) A promiscuous intermediate underlies the evolution of LEAFY DNA binding specificity. Science (80-) 343:645–648. https://doi.org/10.1126/SCIENCE.1248229/SUPPL_FILE/SAYOU.SM.PDF

Schaumberg KA, Antunes MS, Kassaw TK, et al (2015) Quantitative characterization of genetic parts and circuits for plant synthetic biology. Nat Methods 2016 131 13:94–100. https://doi.org/10.1038/NMETH.3659

Schena M, Shalon D, Davis RW, Brown PO (1995) Quantitative Monitoring of Gene Expression Patterns with a Complementary DNA Microarray. Science (80-) 270:467–470. https://doi.org/10.1126/SCIENCE.270.5235.467

Sekhon RS, Lin H, Childs KL, et al (2011) Genome-wide atlas of transcription during maize development. Plant J 66:553–563. https://doi.org/10.1111/j.1365-313X.2011.04527.x

Shanidze N, Lenkeit F, Hartig JS, Funck D (2020) A Theophylline-Responsive Riboswitch Regulates Expression of Nuclear-Encoded Genes. Plant Physiol 182:123–135. https://doi.org/10.1104/PP.19.00625

Sharova L V., Sharov AA, Nedorezov T, et al (2009) Database for mRNA Half-Life of 19 977 Genes Obtained by DNA Microarray Analysis of Pluripotent and Differentiating Mouse Embryonic Stem Cells. DNA Res 16:45–58. https://doi.org/10.1093/DNARES/DSN030

Shashikanth M, Krishna AR, Ramya G, et al (2008) Genome-wide comparative analysis of Oryza sativa (japonica) and Arabidopsis thaliana 5′-UTR sequences for translational regulatory signals. Plant Biotechnol 25:553–563. https://doi.org/10.5511/PLANTBIOTECHNOLOGY.25.553

Shaul O (2017) How introns enhance gene expression. Int J Biochem Cell Biol 91:145–155. https://doi.org/10.1016/J.BIOCEL.2017.06.016

Shaul O (2015) Unique Aspects of Plant Nonsense-Mediated mRNA Decay. Trends Plant Sci 20:767–779. https://doi.org/10.1016/J.TPLANTS.2015.08.011

Shekhawat UKS, Ganapathi TR (2014) Transgenic banana plants overexpressing MusabZIP53 display severe growth retardation with enhanced sucrose and polyphenol oxidase activity. Plant Cell Tissue Organ Cult 116:387–402. https://doi.org/10.1007/S11240-013-0414-Z/FIGURES/8

Shimomura O, Johnson FH, Saiga Y (1962) Extraction, Purification and Properties of Aequorin, a Bioluminescent Protein from the Luminous Hydromedusan, Aequorea. J Cell Comp Physiol 59:223–239. https://doi.org/10.1002/JCP.1030590302

Shlyueva D, Stampfel G, Stark A (2014) Transcriptional enhancers: From properties to genome-wide predictions. Nat Rev Genet 15:272–286. https://doi.org/10.1038/nrg3682

Sieburth LE, Meyerowitz EM (1997) Molecular dissection of the AGAMOUS control region shows that cis elements for spatial regulation are located intragenically. Plant Cell 9:355–365. https://doi.org/10.1105/TPC.9.3.355

Smale ST (2010) Luciferase Assay. Cold Spring Harb Protoc 2010:pdb.prot5421. https://doi.org/10.1101/PDB.PROT5421

Snowden KC, Buchhholz WG, Hall TC (1996) Intron position affects expression from the tpi promoter in rice. Plant Mol Biol 31:689–692. https://doi.org/10.1007/BF00042241/METRICS

Srivastava AK, Lu Y, Zinta G, et al (2018) UTR-Dependent Control of Gene Expression in Plants. Trends Plant Sci 23:248–259. https://doi.org/10.1016/J.TPLANTS.2017.11.003

Srivastava R, Rai KM, Srivastava M, et al (2014) Distinct Role of Core Promoter Architecture in Regulation of Light-Mediated Responses in Plant Genes. Mol Plant 7:626–641. https://doi.org/10.1093/MP/SST146

Stark R, Grzelak M, Hadfield J (2019) RNA sequencing: the teenage years. Nat Rev Genet 20:631–656. https://doi.org/10.1038/s41576-019-0150-2

Stewart J (2001) The utility of green fluorescent protein in transgenic plants. Plant Cell Rep 20:376–382. https://doi.org/10.1007/S002990100346/METRICS

Strizhov N, Li Y, Rosso MG, et al (2003) High-throughput generation of sequence indexes from T-DNA mutagenized Arabidopsis thaliana lines. Biotechniques 35:1164–1168. https://doi.org/10.2144/03356st01

Sudarsan N, Barrick JE, Breaker RR (2003) Metabolite-binding RNA domains are present in the genes of eukaryotes. RNA 9:644–647. https://doi.org/10.1261/RNA.5090103

Sun J, He N, Niu L, et al (2019) Global Quantitative Mapping of Enhancers in Rice by STARR-seq. Genomics Proteomics Bioinformatics 17:140–153. https://doi.org/10.1016/J.GPB.2018.11.003

Tan-Wong SM, Zaugg JB, Camblong J, et al (2012) Gene loops enhance transcriptional directionality. Science (80-) 338:671–675. https://doi.org/10.1126/SCIENCE.1224350/SUPPL_FILE/TANWONG.SM.PDF

Tao Z, Shen L, Gu X, et al (2017) Embryonic epigenetic reprogramming by a pioneer transcription factor in plants. Nature 551:124–128. https://doi.org/10.1038/nature24300

Thole V, Worland B, Wright J, et al (2010) Distribution and characterization of more than 1000 T-DNA tags in the genome of Brachypodium distachyon community standard line Bd21. Plant Biotechnol J 8:734–747. https://doi.org/10.1111/j.1467-7652.2010.00518.x

Thurman RE, Rynes E, Humbert R, et al (2012) The accessible chromatin landscape of the human genome. Nat 2012 4897414 489:75–82. https://doi.org/10.1038/nature11232

Tian B, Manley JL (2016) Alternative polyadenylation of mRNA precursors. Nat Rev Mol Cell Biol 2016 181 18:18–30. https://doi.org/10.1038/nrm.2016.116

Valencia P, Dias AP, Reed R (2008) Splicing promotes rapid and efficient mRNA export in mammalian cells. Proc Natl Acad Sci U S A 105:3386–3391. https://doi.org/10.1073/PNAS.0800250105/SUPPL_FILE/00250FIG8.PDF

Vasil V, Clancy M, Ferl RJ, et al (1989) Increased Gene Expression by the First Intron of Maize Shrunken-1 Locus in Grass Species. Plant Physiol 91:1575–1579. https://doi.org/10.1104/PP.91.4.1575

Vojta A, Dobrinic P, Tadic V, et al (2016) Repurposing the CRISPR-Cas9 system for targeted DNA methylation. Nucleic Acids Res 44:5615–5628. https://doi.org/10.1093/nar/gkw159

Vollbrecht E, Veit B, Sinha N, Hake S (1991) The developmental gene Knotted-1 is a member of a maize homeobox gene family. Nat 1991 3506315 350:241–243. https://doi.org/10.1038/350241A0

Wachter A, Tunc-Ozdemir M, Grove BC, et al (2007) Riboswitch Control of Gene Expression in Plants by Splicing and Alternative 3′ End Processing of mRNAs. Plant Cell 19:3437–3450. https://doi.org/10.1105/TPC.107.053645

Walley J, Sartor R, Shen Z, et al (2016) Integration of omic networks in adevelopmental atlas of maize. Science (80-) 353:0–5. https://doi.org/10.5061/dryad.v8969

Wang B, Regulski M, Tseng E, et al (2018) A comparative transcriptional landscape of maize and sorghum obtained by single-molecule sequencing. Genome Res 28:921–932. https://doi.org/10.1101/GR.227462.117

Washburn MP, Koller A, Oshiro G, et al (2003) Protein pathway and complex clustering of correlated mRNA and protein expression analyses in Saccharomyces cerevisiae. Proc Natl Acad Sci U S A 100:3107–3112. https://doi.org/10.1073/pnas.0634629100

Weber B, Zicola J, Oka R, Stam M (2016) Plant Enhancers: A Call for Discovery. Trends Plant Sci 21:974–987. https://doi.org/10.1016/j.tplants.2016.07.013

Wiegand HL, Lu S, Cullen BR (2003) Exon junction complexes mediate the enhancing effect of splicing on mRNA expression. Proc Natl Acad Sci 100:11327–11332. https://doi.org/10.1073/PNAS.1934877100

Wiese A, Elzinga N, Wobbes B, Srneekens S (2004) A Conserved Upstream Open Reading Frame Mediates Sucrose-Induced Repression of Translation. Plant Cell 16:1717–1729. https://doi.org/10.1105/TPC.019349

Xu X, Tao Y, Gao X, et al (2016) A CRISPR-based approach for targeted DNA demethylation. Cell Discov 2:. https://doi.org/10.1038/celldisc.2016.9

Xu X, Xie Q, McClung CR (2010) Robust Circadian Rhythms of Gene Expression in Brassica rapa Tissue Culture. Plant Physiol 153:841. https://doi.org/10.1104/PP.110.155465

Yadav S, Swati D, Chandrasekharan H (2015) Thiamine pyrophosphate riboswitch in some representative plant species: A bioinformatics study. J Comput Biol 22:1–9. https://doi.org/10.1089/CMB.2014.0169/ASSET/IMAGES/LARGE/FIGURE5.JPEG

Yan W, Chen D, Schumacher J, et al (2019) Dynamic control of enhancer activity drives stage-specific gene expression during flower morphogenesis. Nat Commun 2019 101 10:1–16. https://doi.org/10.1038/s41467-019-09513-2

Zaret KS, Mango SE (2016) Pioneer transcription factors, chromatin dynamics, and cell fate control. Curr Opin Genet Dev 37:76–81. https://doi.org/10.1016/j.gde.2015.12.003

Zhang H, Lang Z, Zhu JK (2018) Dynamics and function of DNA methylation in plants. Nat Rev Mol Cell Biol 19:489–506. https://doi.org/10.1038/s41580-018-0016-z

Zhang J, Sun X, Qian Y, et al (1998) At Least One Intron Is Required for the Nonsense-Mediated Decay of Triosephosphate Isomerase mRNA: a Possible Link between Nuclear Splicing and Cytoplasmic Translation. Mol Cell Biol 18:5272–5283. https://doi.org/10.1128/MCB.18.9.5272/ASSET/24D56646-672B-4A63-A12E-ACA6BF7E61E6/ASSETS/GRAPHIC/MB0981709007.JPEG

Zhang W, Zhang T, Wu Y, Jiang J (2012) Genome-Wide Identification of Regulatory DNA Elements and Protein-Binding Footprints Using Signatures of Open Chromatin in Arabidopsis. Plant Cell 24:2719–2731. https://doi.org/10.1105/TPC.112.098061

Zhang Y, Liu L, Qiu Q, et al (2021) Alternative polyadenylation: methods, mechanism, function, and role in cancer. J Exp Clin Cancer Res 40:1–19. https://doi.org/10.1186/S13046-021-01852-7/TABLES/3

Zheng N, Li T, Dittman JD, et al (2020) CRISPR/Cas9-Based Gene Editing Using Egg Cell-Specific Promoters in Arabidopsis and Soybean. Front Plant Sci 11:800. https://doi.org/10.3389/FPLS.2020.00800/BIBTEX

Zhu Q, Dabi T, Lamb' C (1995) TATA box and initiator functions in the accurate transcription of a plant minimal promoter in vitro. Plant Cell 7:1681–1689. https://doi.org/10.1105/TPC.7.10.1681

CHAPTER 2: NOVEL TISSUE-SPECIFIC PROMOTERS CHARACTERIZED IN MAIZE

Abstract

Promoter sequences control the timing, location, and level of gene transcription. Lack of identification and characterization of plant promoters can limit and define the questions that can be asked in plant genetics research and biotechnology applications broadly. For some plant species, such as the model species *Arabidopsis thaliana*, characterized promoters are plentiful. However, for other species such as the globally important crop, maize (*Zea mays* L.), such resources are lagging. We have identified and developed novel promoter elements which confer tissue-specific expression in maize as well as demonstrated the importance of a holistic approach to assessing promoters. Gene expression data was utilized to identify candidate promoters which were subsequently verified through qualitative GUS staining and quantitative MUG assays. From the initial pool of 14 candidate promoters, 6 showed potentially useful tissue-specific expression. Among these, 3 showed expression in unexpected and very specific off-target tissues, demonstrating the necessity of checking a wide range of tissues when characterizing promoter function. Based on that data, the full-length cloned promoters were reduced in sequence length based on measured accessible chromatin, resulting in smaller promoters that still functioned. Finally, the promoter *lhcb10* was shown to function as a cis-regulatory region and can be used to create synthetic promoters in combination with the CaMV53 minimal promoter. This work expands the range of available molecular tools available to control transgene expression in maize while demonstrating the importance of vigilance in characterization of promoters.

Significance Statement

Six tissue-specific promoters were identified and characterized in maize and demonstrated to control leaf-, embryo-, and root-specific gene expression. Promoters are a foundational biotechnology tool that enables scientists to both dissect gene function and bestow novel properties through genetic engineering. While many promoters have been developed and tested in model species such as *Arabidopsis thaliana*, crops like maize (*Zea mays* L.) still lack well-characterized options. We identified several promoters from genes showing tissue-specific expression via analysis of RNA-seq data collected in maize across the entire development of the plant. Subsequently the promoter regions were cloned and transformed into inbred maize line, LH244, where qualitative and quantitative assessments demonstrated both tissue-specific expression, and instances of off-target expression. Finally, some promoters were further developed into shortened, "minimal promoter" forms and one promoter, *lhcb10*, was demonstrated to function as a mesophyll-specifying cis-regulatory region when combined with the CaMV35S core promoter sequence.

Introduction

Maize is a vital global crop with 197 M ha of land dedicated to its growth, second only to wheat (Erenstein et al. 2022). As dry weight, maize production totals 1,137 million tons, 50% higher than either wheat or rice (Erenstein et al. 2022). Much of this production goes into the feeding of cattle and creation of a range of products from 3D printing filament to ethanol. Maize production is projected to continue to grow as it has for decades (Erenstein et al. 2022). Challenging this growth, though, is continued ecological degradation and anthropogenic climate change, both of which necessitate the continued genetic improvement of maize and development of biotechnology tools to support it (Pörtner et al. 2022).

Transgenic plants were first introduced to the market in 1994 with the “Flavr Savr” tomato and have since grown year over year, with 439 transgenic events now commercialized (Kumar et al. 2020). Among these, 152 are maize events, with traits ranging from enhanced ethanol production, herbicide resistance, and insect resistance (ISAAA). Most of these have relied on overexpression of the transgenes, turning on expression at high levels in all tissues (Nuccio 2018). This approach is sufficient with single events for certain traits but is complicated when multiple events are combined (stacked) into a single genotype, which is commonly done in commercial lines from industry. Stacked transgenes may interfere with one another, inducing silencing in cases where a promoter is duplicated (Rajeevkumar et al. 2015). Transgenic approaches to crop improvement are increasingly more ambitious, such as introduction nitrogen fixation and C4 photosynthesis pathways into new crops (Mahmud et al. 2020; Ermakova et al. 2021). In tandem with and fueling these direct applications is the ever-expanding field of functional genomics research. Next generation sequencing has brought to the fore many causative genes, especially through GWAS studies (Gupta et al. 2019; Nguyen et al. 2019; Tibbs Cortes et al. 2021). To express and study these genes optimally, a diverse array of promoters has been characterized spanning across developmental time and physiological space (as reviewed in Nuccio et al. 2018). Despite these advancements, there are still gaps in the depth of characterization and breadth of targetable tissues in maize promoters.

The goal of this study was to identify tissue-specific promoters and demonstrate their efficacy using quantitative and qualitative assays. Promoters were identified through RNA sequencing data, still the best measure of promoter performance since that feature predominantly acts transcriptionally. RNA data was taken from bulk tissue samples as a part of the Maize Gene Atlas (Sekhon et al. 2011, 2013; Stelpflug et al. 2016); single-cell RNA sequencing data had not

yet come to the fore when this project began. Other data sources such as proteomics were considered but ultimately discarded as it would only add variation to the analysis. Identified promoters were cloned into vectors driving expression of the GUSPlus gene (Broothaerts et al. 2005). This enabled qualitative and quantitative analysis of promoter strength, tissue specificity and timing via qualitative staining with 5-bromo-4-chloro-3-indolyl glucuronide (X-Gluc) and quantitative expression measurements with 4-methylumbelliferyl-beta-D-glucuronide (MUG). Promoters were further refined utilizing assay for transposase-accessible chromatin with sequencing (ATAC-seq), which indicates which regions of the genome are open and accessible. Maize inbred line LH244 was chosen for transformation with the promoter test vectors due to its ease of transformation and genetic similarity to maize genetic model line, B73, though new techniques with morphogenic genes have opened the potential of direct transformation into B73 (Lowe et al. 2016). Six promoters were shown to have tissue-specific expression: *ris2*, *lhcb10,* and *clo2b* each show expression exclusive to their designed tissues while *nas2*, *aa2m*, and *expb14* showed off-target expression.

Results

Identification, Isolation, and Cloning of Candidate Tissue-Specific Promoters

Candidate promoters were identified from the pool of maize tissue expression data in the Maize Gene Atlas (Sekhon et al. 2011, 2013; Stelpflug et al. 2016). Tissue-specific expression was identified by looking for genes with high expression levels in desired tissue and low off-target expression, ranking candidates based on their Euclidean distance from an “ideal expression profile”. Manual curation of candidates resulted in 14 promoter sequences being identified that appeared to drive leaf-, seed-, or root-specific gene expression (Table 2-2).

Identified promoters were synthesized based on their sequence in version 4 of the B73 genome, with the first 2kb upstream sequence cloned along with any 5'-UTR if present (Jiao et al. 2017) . The promoters to be tested were placed in front of GUSPlus (Brothaerts et al. 2005) gene sequence to drive expression. For comparison, a T-DNA containing the *ZmUbi1* strong, constitutive promoter driving GUSPlus was also created (Christensen et al. 1992). T-DNA expression vectors were constructed using Golden Gate cloning and included the tdTomato gene under the control of the *ZmUbi1* or the *CmYLCV* promoter to visually track events and the bar gene (providing resistance to the herbicide, bialophos) driven by the *OsUbq2* promoter to enable selection of transgenic cells/plants (Thompson et al. 1987; Christensen et al. 1992; Wang et al. 2000a; Stavolone et al. 2003; Engler et al. 2014). Promoter testing transcriptional units were separated from other transcriptional units within the T-DNA region of the vector by the petunia transformation boost sequence (TBS) insulator. This insulator has been shown to block enhancing activity it separates in *Arabidopsis thaliana* and *Nicotiana tabacum*, though some studies have shown it can mildly enhance expression and might not reduce variability (Hily et al. 2009; Singer et al. 2011; Dietz-Pfeilstetter et al. 2016; Pérez-González and Caro 2019). All candidate promoters were successfully cloned into expression vectors (**Error! Reference source not found.**). Transformation treatments were successful with all promoters in LH244 and at least 3 independent single-copy events were generated for each promoter, taken from a larger generated pool of events.

Assessment of LH244 Transcriptome Compared to B73

RNA-sequencing data was successfully generated from a reduced set of tissues from the B73-based Gene Atlas, specifically, in 11 "days after pollination (DAP) whole seeds, 24 DAP

embryos, V1 pooled leaves, R1 12th leaf, V1 stem & shoot, V7 crown roots, and VT anthers (Abendroth et al. 2011). Generated reads were aligned to the B73 version 5 transcriptome alongside reads from the Maize Gene Atlas (Hufford et al. 2021). Initial comparison was conducted using principal component analysis on the "transcripts per million" for each gene which did not result in like-tissue samples grouped with one another but instead were separated by their variety. This lack of grouping was further confirmed by Pearson correlations between the samples, which ranged from 0.43 to 0.79 with "V7 crown roots" and "R2 12th Leaf" comparing the worst (David Freedman, Robert Pisani 2007). The pattern of expression of candidate promoters was checked visually between the two varieties, with parity found between each in their tissue specificity. Two example promoter profiles are shown in Figure 2-2.

Qualitative and Quantitative Assessment of Promoters for Tissue-Specificity

Testing of promoters for strength and specificity necessitated development of a set of tissues that would be sufficient for capturing potential off target expression while minimizing the number of tissues collected. Principal component analysis was performed on RNA sequencing data, again from the Maize Gene Atlas, to identify which tissues separate based on expression. This resulted in 14 tissues being identified to cover the whole development cycle of maize (Table 2-1). Initial assessment was conducted on T0 plant tissues which were stained with X-Gluc in the targeted tissues with further investigation conducted on those promoters which showed promise. Subsequent investigation was done on T1 plants which were hemizygous for the transgene and single copy. Tissue samples from the growing plants were taken at each phase listed in Table 2-1 and stained with X-Gluc. Additionally, V1 leaves and primary roots from *lhcb10* and *ris2* were measured for GUS activity through a MUG assay. Analysis of promoters with cis-regulatory

identification programs like PlantPan (Chow et al. 2019) was considered but the low density of experimental evidence for maize transcription factor binding sites makes the process too uncertain.

Among the leaf samples, only *lhcb10* and *ris2* showed visible GUS staining results. Examples of visual GUS assay results for the *lhcb10* promoter and *ris2* promoter are shown in Figure 2-4 and Figure 2-5, respectively. Both showed consistent staining specifically in the mesophyll tissue of leaves across the entire life cycle of the plant. The only off-target expression detected was in the green tissue of anthers. MUG assay results for the two promoters confirm the visible GUS assay results for activity level and show that *lhcb10* is a much stronger promoter than *ris2* (Table 2-3).

Three embryo-specific promoters were identified: *aa2m* (Figure 2-6), *clo2b* (Figure 2-7), and *expb14* (Figure 2-8). The *aa2m* promoter was most active in the 11 DAP embryo though still showed activity/staining at 24 DAP, most prevalently along the scutellar surface. It was also very active in the leaf vasculature at the base of the 6th leaf from the V5 growth phase. There was no activity detected elsewhere including the tip of the V5 6th leaf. The *clo2b* promoter displayed high activity in the 11 DAP embryo, but was very faint by 24 DAP, with only some GUS staining occurring in the aleurone. No other GUS expression was detected in any other tissues tested. Finally, the *expb14* promoter was shown via GUS assay to be active in the 24 DAP embryo, within the scutellum itself and at its juncture with the nascent plant. Staining was also strongly visible within the still-growing vasculature of above-ground tissues, present in the V1 leaves, V5 6th leaf base, and the fourth internode. Expression faded within mature tissue such as the tip of the V5 6th leaf. No staining was observed in root tissues.

Only one root-specific promoter was identified, *nas2* (Figure 2-9). Visual GUS assays demonstrated strong *nas2*-driven GUS expression in the primary root, specifically in the cortex. The promoter was shown to be active in mature root tissue and did not appear in the elongation zone or at the root cap. In crown roots, GUS expression was detected in the secondary root, though much weaker compared to the primary root. Expression was not observed in brace roots or any reproductive tissues. The *nas2* promoter was also found to be active in the vasculature specifically in the V9 8th leaf, which was sampled close to the ligule. Expression was not detected elsewhere on the leaf such as the V1 leaves or the V5 6th leaf.

Generation and Assessment of Minimal Promoter Sequences and Synthetic Promoters

The *lhcb10* and *ris2* promoters were reduced from their original 2kb size to 499bp and 400bp, respectively, based on chromatin accessibility data in maize leaf (Ricci et al. 2019). In addition, a fragment of *Lhcb10* was isolated based on accessible chromatin from leaf tissue (Ricci et al. 2019). A 400bp segment from -499 to -99 from the transcription start site in the B73 V4 genome was selected (Figure 2-3). However, in the V5 annotation for maize, this fragment corresponds with -395 to +4 of the transcription start site as a 5'UTR had been added compared to the V4 version. This fragment was combined with CamV35S and *ZmUbi1* core promoter sequences to construct a complete promoter. The core sequences were also cloned on their own into a T-DNA expression cassette as a control. All the promoters were synthesized then cloned into T-DNA expression cassettes like previous constructs (**Error! Reference source not found.**) and transformed into LH244.

GUS staining results indicated activity of both minimal promoters exclusively in the shoot but not the root (Figure 2-10). MUG assay results were like the visual assay in that that

both minimal promoters maintained their rank relative to one another in strength and their strength compared to the longer versions (Table 2-3). Neither core promoter showed GUS staining in any of the tissues sampled. The *lhcb10* fragment with the *ZmUbi1*-core sequence also did not demonstrate positive GUS staining in any tissues. The fragment combined with CamV35S-core showed positive staining in the leaves, but none in roots (Figure 2-10). Results from MUG assays indicated that the expression level of the *lhcb10* fragment with the *ZmUbi1*-core sequence was greatly decreased relative to either the full length or minimized promoter sequence.

Discussion

Maize Inbred LH244 as a Transformation Platform

Analysis of the LH244 expression data within the context of B73 reveals the difficulty in experiment duplication and the incomplete picture genomic data provides. While B73 and LH244 are very similar genetically with 97.1% parity in SNPs, clear expression differences between the two were observed, specifically in the reproductive-phase leaf and crown root tissues. LH244's genomic background is closer to that of elite lines and may explain some of the differences observed specifically in the reproductive phase leaf and crown roots. An additional difference might be in sampling practices and conditions, despite best efforts to prevent these. Ultimately, the expression profiles of candidate promoters are closely aligned between B73 and LH244, demonstrating its utility for this research's goals.

Leaf Promoter Performance: *lhcb10* and *ris2* and their variants

Light harvesting chlorophyll a/b binding protein 1 is a nuclear encoded component of the photosystem II (Bansal and Bogoradt 1993; Rocca et al. 2000). Given its central role in photosynthesis, it has been identified in many trait assessments including stalk biomass, anther development, and drought resistance (Mazaheri et al. 2019; Li et al. 2021; Han et al. 2022). Recent work on single-cell RNA-seq and ATAC-seq in maize leaves places its expression alongside other photosynthetic cells in the mesophyll (Bezrutczyk et al. 2021, Marand et al. 2021). Transgene-based expression studies of the promoter in the maize variety FR9cms x FR37 showed that the promoter provided mesophyll-specific activity which is specifically mediated by a 159bp region at -1026 to -868 upstream of the transcription start site, studied through particle bombardment of leaf material (Bansal and Bogoradt 1993). This corresponds with the GUS staining observations of *lhcb10*, which is specific to the mesophyll and absent in bundle sheath cells. Our observations, however, did not match this previous study with our minimized *lhcb10* sequence which excluded these indicated regions and thus should have shown a decrease in overall expression plus an appearance of bundle sheath expression, neither of which is shown. A possible explanation for this is a difference of approach, as the original study utilized particle bombardment-based transient expression assays while we worked with stably transformed events. Additionally, we were working in a different maize genotype whose expression network might be influencing the promoter's activity. Rieske iron-sulfur protein 2, *ris2*, is a much more straightforward success in terms of identifying a promoter with tissue-specific activity. In maize, the *RIS2* protein is a component of the cytochrome f/b6 complex, which participates in the electron transport chain (Barkan and Walker 1994; Klusch et al. 2022). It is known to be involved in photosynthesis, and its expression in leaves is expected (Simkin et al. 2017). Single cell RNA-seq and ATAC-seq data in maize leaves shows its expression tied to the mesophyll,

though not conclusively to a reproducible cluster of cells (Bezrutczyk et al. 2021). The main potential detraction of the promoter is its weak performance compared to *lhcb10* and *ZmUbi1*, but this might be useful in situations where low but present expression is necessary.

Utilization of the first 2kb of promoter sequences for our tests was a very conservative approach, as evidenced by the success of minimizing the promoter sequences of *lhcb10* and *ris2*. Their smaller sequence size can increase their utility while minimizing their impact on T-DNA size. A natural limitation of this approach is that it necessitates having chromatin availability data in the specific tissue targeted. However, with the decreasing costs of next-generation sequencing, the available tissue information will only expand. Neither core promoter demonstrated expression alone, which is ideal when creating synthetic promoters built on core promoters. Results from tests of the final *lhcb10*-synthetic promoters are difficult to interpret, given that the fragment isolated likely also includes the core promoter of *lhcb10*. This piece core piece however does not seem sufficient for expression on its own, as if that was the case then both promoters would be expressed. However, only the CamV35S core promoter combined with the *lhcb10* fragment was expressed. Additional research is necessary to clarify the roles of the various elements.

Interpretation of Seed-Specific Expression of Promoters from Maize genes *aa2m*, *clo2b*, and *expb14*

Expression of alanine transferases has been observed in maize in two tissues, seed and leaf, both of which were observed in tests of the alanine transferase promoter sequence *aa2m*. Alanine transferases have been found in scutellar tissue of the embryo, where various isozymes were described (Watson et al. 1992). Presence of the enzyme matches the expression pattern

observed for the GUS gene driven by the *aa2m* promoter, with strong GUS staining observed in scutellar tissues of transgenic maize embryos at 11 DAP and even into 24 DAP. Other studies have shown alanine transferase activity in the endosperm; however, we did not detect GUS expression in that tissue with the *aa2m* promoter (Faleiros et al. 1996). Within the leaf tissues, alanine transferase activity is noted to increase as the leaf matures, as it is involved in a secondary decarboxylation pathway (Pick et al. 2011; Lori Tausta et al. 2014). The activity of the *aa2m* promoter in our tests, based on our GUS expression assays did not fit that scheme, being expressed in the vasculature and not the bundle sheath or mesophyll cells. Further research may unveil the role and functions of *aa2m* in both embryo and vasculature, however, the characterized promoter may be useful for other purposes.

ZmClo2b, also known as *ZmPhp20719a*, is a member of the caleosin/peroxygenases family of proteins. Originally identified through desiccation studies, this family has been found to associate with lipid bodies as well as having catalytic functions (Yamaguchi-shinozaki et al. 1992; Chen et al. 1999; Næsted et al. 2000; Hernandez-Pinzon et al. 2001). This family is ubiquitous in the plant kingdom and, within maize, 12 caleosin genes have been identified (Lizong et al. 2014; Hanano et al. 2023). As a family, caleosin has been detected in the embryo of maize (Tnani et al. 2011). Specifically, for *ZmClo2b*, it has been detected through RT-PCR in seed material (Lizong et al. 2014). As a family, though, they are detected in many different tissues (Lizong et al. 2014). Furthermore, mutations in the gene have been linked to smaller and thinner kernels (Jia et al. 2016). The expression pattern observed in our tests of the promoter of this gene exclusively in the early embryo of maize aligns with each of these expectations. The expression seems to extend into 24 DAP but only in the aleurone, which makes sense given that embryo development has been slowed by that stage of seed growth.

The expansin family has an observed role in seed development and has been linked to seed yield in sunflowers, *Arabidopsis*, and other plants (Bae et al. 2014; Castillo et al. 2018). Specifically in maize, *expb14* and *expb15* have been found to be downstream of MIR164 and correlated with smaller seeds overall (Zheng et al. 2019). Furthermore, *ZmEXPB15* has been found to underlie a maize kernel QTL, activated by *ZmNAC11* and *ZmNAC29*, resulting in larger kernels through regulating nucellus elimination (Sun et al. 2022). In the 2019 study by Zheng et al., the promoter of *expb14* was analyzed and found to be active in the endosperm, which was not observed in this study. This may be a result of differences in maize genotypes as the 2019 study was conducted using maize inbred line Mo17 and not LH244. Additionally, Zheng et al. examined an earlier phase of seed development, 9 DAP, which means this expression might have been missed by our analysis. Not mentioned in any of these studies is the activity of *expb14* in developing vascular tissue, where it might be acting to activate expansion as well. Discoveries such as this may justify the additional work of checking many tissues for expression analyses.

Root Specific Expression of *Nas2*

Nas2 is a nicotianamine synthase gene which creates nicotianamine (NA) from three molecules of S-adenosyl-L-methionine (Higuchi et al. 1999). NA itself is a precursor of phytosiderophores, which are secreted by grass roots to bind with Fe3+ which is then taken up by plasma membrane transporters (Römheld and Marschner 1986). In addition, NA is a chelator of metals broadly and can bind Fe, Mn, Cu, and Zn directly (Reichman and Parker 2002; Takahashi et al. 2003; Rellán-Álvarez et al. 2008; Haydon et al. 2012). *Nas2* is a class I nicotianamine synthase and has been detected to be upregulated under Fe deficient conditions and downregulated under conditions of excess (Mizuno et al. 2003; Zhou et al. 2013). Li *et al.* in

2022 examined root tips for single-cell RNA-seq which did not detect *Nas2* in any of their expression clusters (Li et al. 2022). Unsurprisingly, the gene has been found to underlie a QTL conferring iron deficiency tolerance (Xu et al. 2022). The above observations regarding *Nas2* broadly correspond to our observations from tests of the *nas2* promoter, with the greatest expression seen in cortex tissues, where synthesis of NA is expected. The lack of expression in the primary crown root but presence of expression in the secondary roots is also not surprising given that secondary roots and root hairs are the primary site of nutrient uptake. The activity of the *nas2* gene within leaf vasculature has not yet been noted in the literature. However, NA is known to be involved in the process of long-distance transport of FE and might be the reason why expression can be found there (Curie et al. 2009).

Conclusions

There is a constant need for more and better plant biotechnology tools, both for functional genomics and epigenomics research and for direct application in the genetic improvement of crops. Through our research, 6 maize-derived promoters demonstrating tissue-specific expression were identified and characterized and can be immediately utilized to advance genetic studies and improvement of maize and, likely, other plant species. Not all promoters are exclusive to the tissue they were designed for, revealing the necessity of thorough characterization and testing for off-target expression to ensure that promoters work as expected in transgenic applications. The advancement of single-cell techniques in plants opens exciting possibilities in the prediction of tissue-specific promoters. Recently in maize, single-cell RNA sequencing has been done across a variety of tissues including anthers, apical meristems, leaves, and more, holding great potential for identifying tissue-specific genes (Nelms and Walbot 2019;

Satterlee et al. 2020; Xu et al. 2021; Bezrutczyk et al. 2021; Tao et al. 2022; Li et al. 2022). There is a balance to be struck between efforts in characterizing promoters and the quality of information received. Ultimately, scientific research and discoveries are built upon the work of others and thoroughness of the research ensures quality of results moving forward. It is envisioned that the promoter tools developed here will enable further advances in plant genomics research and genetic enhancement.

Materials and Methods

RNA Seq Data Collection, Analysis, and Comparison

RNA was extracted from maize plant tissues using the Monarch Total RNA kit (New England BioLabs catalog # T2010S) with DNase treatment. RNA QC was done on the Agilent PicoChip (Agilent Technologies item # 50671513). Each library was created with the TruSeq Stranded mRNA Kit (Illumina item # 20020595). Libraries were quantified in singlet on a Qubit with dsDNA HS reagent. Libraries were sequenced on a NovaSeq 6000. Paired-end, 150bp sequencing was performed. Data was processed with bcl2fastq. Sequencing was provided by the University of Wisconsin – Madison Biotechnology Center's DNA Sequencing Facility (Research Resource Identifier – RRID:SCR_017759). Alignment and TPM calculations were done using Salmon with default settings (Patro et al. 2017). Alignment was done against the B73v5 reference genome for all samples analyzed. Full-length annotated cDNA sequences were used as the annotation target. Principal component analysis and Pearson correlations were calculated with R (R-Core-Team 2023).

Identification of Tissue-Specific Promoter Candidates

Candidate promoter sequences were identified from version 4 of the maize genome sequence (Jiao et al. 2017). Sequences were filtered to remove promoters with long-range activating cis-regulatory elements, which can be tens of thousands of bases away and are not captured readily by cloning (Lu et al. 2018; Ricci et al. 2019). Additionally, 50% of analyzed promoters were taken among those that had at least one associated study in MaizeGDB to ensure some level of documentation about the gene. RNA-seq expression data from the maize gene atlas, which contains expression data from all major tissues in maize, was utilized to identify genes with a desired expression profile (Sekhon et al. 2013; Stelpflug et al. 2016). Expression was normalized to "transcripts per million" (TPM). *ZmUbi1* expression in tissues was used as the standard of high, constitutive expression to create an idealized expression level within the tissues targeted. Specifically, maize zygotic embryos, leaves, and root tissues were targeted for promoter identification. Candidates were ranked by their Euclidean distance to the idealized expression level. High ranking candidates were manually appraised for their utility and genes that seemed influenced by stress or environmental conditions were eliminated from consideration.

Cloning of Candidate Promoter Sequences

Sequences for identified promoters were isolated from the version 4 sequence of the B73 maize genome (Jiao et al. 2017). The first 2kb of promoter sequence, including any annotated 5'-UTR, was synthesized as a level 0 compatible part for Golden Gate cloning which was used to drive GUSPlus expression (Broothaerts et al. 2005; Engler et al. 2014). The annotated primary transcript was used to determine the 3' end of the promoter sequence. The *OsUbq2* promoter was used to drive the bar gene (Thompson et al. 1987; Wang et al. 2000b). The red fluorescent protein reporter gene, tdTomato, was expressed with either the *ZmUbi1* promoter or the YLCV

promoter (Christensen et al. 1992; Stavolone et al. 2003). The Bar and tdTomato expression cassettes were separated from the tested promoter by the TBS insulator (Hily et al. 2009). Two versions of the transformation vector were created, with the only difference being in the promoter utilized to drive tdTomato expression (Shaner et al. 2004). In version 1, the promoter used was *ZmUbi1* and in version 2 the Yellow Leaf Curling Virus promoter was used (Christensen et al. 1992; Stavolone et al. 2003). All enzymes used were sourced from New England Biolabs. A linear map of the T-DNA portion of the binary vector is provided in **Error! Reference source not found.**. Shortened versions of *ris2* and *lhcb10* promoters were created by visually analyzing ATAC-seq data collected from young leaves (Ricci et al. 2019). The promoter was shortened from the 5' end until reads of accessibility in leaf tissue rose above background levels.

Stereo Microscope Imaging

A Leica stereomicroscope (model M165 FC) with a DFC 7000T camera attached was used to take all sample photographs with the LAS X software suite.

Plant Materials

Carlin SVD-250 pots were arranged in Carlin 3 – 234 ST-I-0804 vacuum standard inserts. A custom potting mix based on "Jolly Gardener" brand Pro-Line C/B (Part No. 18-1010) was distributed to the pots. Seeds of maize inbred line, LH244, were sown by placing one seed per cell, pressing down ¾ inch to 1 inch, then covering with loose, dry soil medium. Flats were placed on benches and manually watered until day 7 to 10, when flood fertigation was begun. Seedlings were grown in flats in the nursery area of the greenhouse for 18 to 21 days.

Greenhouse rooms were illuminated with high pressure sodium lamp fixtures (model #HS 2000 from PL lighting) with 600 W Phillips E39 HPS bulbs and supplemented with natural sunlight. The light fixtures were mounted about 2.7 m above the floor. Seeds were germinated on a bench top approximately 1.7 m below the lights. Custom LED fixtures were used for supplemental lighting. The LED fixtures were 600 W and 1.3 m from the bench top. Average photosynthetic photon flux density (PPFD) was 500 µmol m-2 s-1 on the bench top but varied between 370 and 660 µmol m-2 s-1 depending on the age of fixtures, the time of day and time of year.

After 18 to 21 days, plants were transferred to size 1200 Elite Blow-molded Nursery Containers (Carlin, Inc. catalog # 4-2022). The potting mix was the same as above. Pots were placed into carts, with the cart surface approximately 1.7 m below the light fixtures. After 7 to 9 days, pots were moved to the floor and placed on plastic pallets.

The temperature set points for the greenhouse were 26.7°C during the day and 21.1°C during the night. The photoperiod consisted of a 16-hr light: 8-hr dark with supplemental sunlight during the light period. The fertigation mix utilized was Peters Excel 15-5-15 Cal Mag Special (Carlin item number 20-235). Plants were on timed fertigation cycles based on growth stage and demand.

Sampling of Plant Materials

V1 Plant Sampling: Primary Root: A 2cm piece of primary root was collected, measuring from the base of the roots. Stem & Shoot: Plants were cut just above the root growth point and 2cm up the stem. A twisting motion with the fingertips was used to remove the other leaves resulting in the meristem being exposed. Pooled Leaves: All leaf material 2cm above the base of

the plant was collected. V5 Sampling: Only plants where the ligule of the 5th leaf had emerged less than an inch was sampled. 6th Leaf Base: The stem was cut from the roots and the outer 5 leaves were peeled away. A 1cm section just above the faint ligule of the 6th leaf was sampled. 6th Leaf Tip: A 2cm long segment of the 6th leaf was sampled.

V9 Sampling: 8th Leaf: The 8th leaf was isolated a cut 2cm above the ligule and 1cm below it was made to create the sample. 4th Internode: The sample was taken by cutting just below the 5th node and just above the 4th node. Crown Roots: Three crown roots from the 3rd node were collected from the plant and a 2cm section was sampled 10cm away from the where the root grew from the node.

R1 Sampling: Brace Roots: A 2cm section was cut from the tip of brace roots from the 5th node of the plant. 11th Leaf: 30cm from the tip of the 11th leaf, a 2cm wide segment of the leaf was sampled. Anthers: Four sets of anthers were taken from the main tassel just below where pollen shed had begun.

Genetic Transformation of Maize

Test vectors were transformed into maize inbred LH244 using Type I embryonic callus induction. The protocol utilized was originally developed for maize inbred B104, but, with minor modifications also worked for LH244 (Kang et al. 2022). *Agrobacterium tumefaciens* strain AGL1 was used for all transformations (Lazo et al. 1991). Single copy events were identified by examination of segregation ratios of tdTomato expression in T1 seeds. Sixty seeds were assayed per event and those with 33 positive seeds or less were deemed single copy for falling within the expected 1:1 segregation ratio for transgenic to non-transgenic progeny from a transgenic T0 by null plant cross.

Gus Staining Protocol

A buffered solution of X-gluc was used to assess expression and localization of transgene-encoded beta-glucuronidase (GUS) at a qualitative level (Hackett 1993). The buffer was a 0.1M solution of K2HPO4 dipotassium phosphate, brought to a pH of 7 with HCl / NaOH. Triton x-100 was then added to create a final concentration of 0.5%. The buffer was stored at 4°C until used. To create the final GUS assay solution, X-gluc dissolved in DMSO (at 50mg/mL) was added to the buffer solution to a final concentration of 100mg/L. Material to be assayed was dissected from plants as specified above and immediately submerged in the solution. Three rounds of applied vacuum were used to infiltrate the solution into the tissues. The tissues were stained overnight for 16 hours at 37°C. Samples were cleared of chlorophyll using 100% ethanol and stored in 70% ethanol.

MUG Assay Protocols

MUG assays were performed as standard procedure (Hackett 1993). Samples were stored at 2°C, never frozen, for up to 24 hours as needed. Fluorescence measurements were performed on a Tecan SPECTRAFluor Plus microplate reader.

Figures

Figure 2-1: Linear Maps of T-DNA Region within Promoter Test Vectors

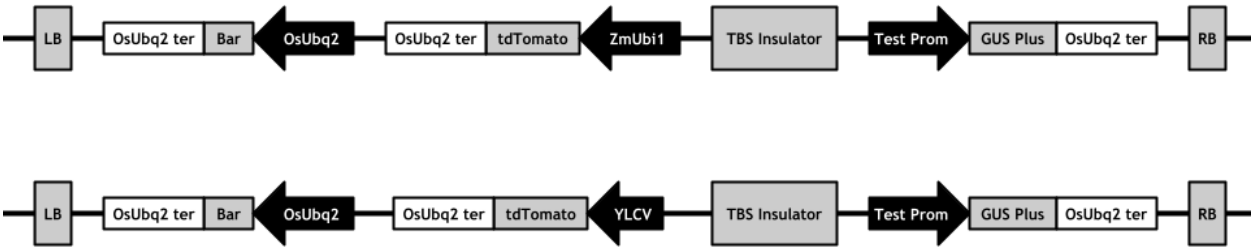

Figure 2-1. Linear maps of T-DNA region within promoter test vectors. Cestrum Yellow Lead Curling Virus (YLCV) promoter and ZmUbiquitin1 (ZmUbi1) drive tdTomato. The bialaphos resistance gene (Bar) is driven by the *Oryza sativa* ubiquitin 2 (OsUbq2) promoter. The tested promoter drives GUS Plus. The Petunia transformation boost sequence (TBS) insulator separates two sides of the T-DNA.

Figure 2-2: Expression Levels of RNA Transcripts of Candidate Tissue-Specific Genes in B73 and LH244 as Measured by RNA-seq

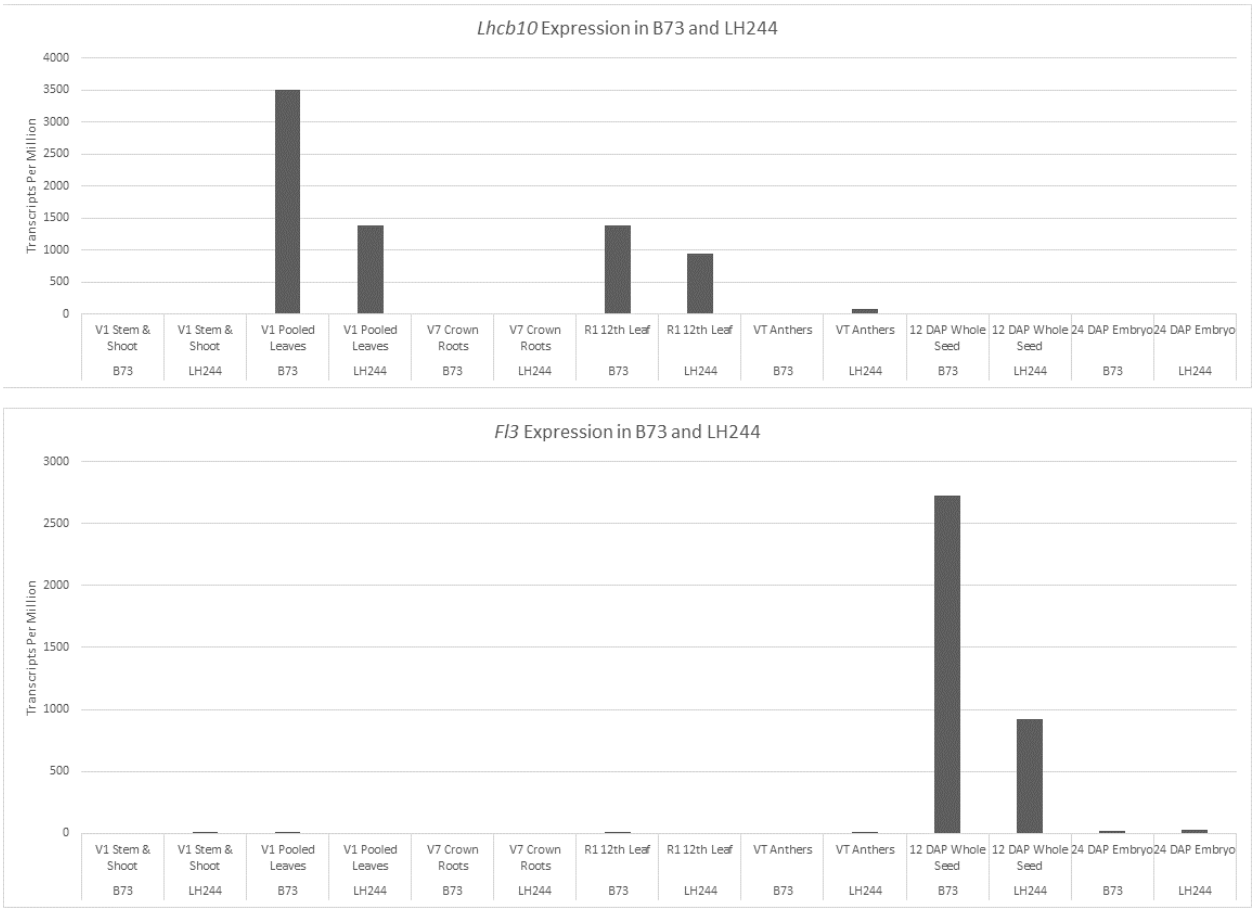

Figure 2-2. Expression levels of RNA transcripts of candidate tissue-specific genes in B73 and LH244 as measured by RNA-seq. The top chart features RNA transcript expression of the gene *lhcb10* and the bottom features *fl3*. Each shows expression in transcripts per million across V1 stem and shoot, V1 pooled leaves, V7 crown roots, R1 12th leaf, VT anthers, 12 days after pollination (DAP) whole seed, and 24 DAP embryo. Measurements are a mean of three replicates per tissue.

Figure 2-3: Chromatin Accessibility of *lhcb10* Promoter Sequence as Measured by ATAC-seq

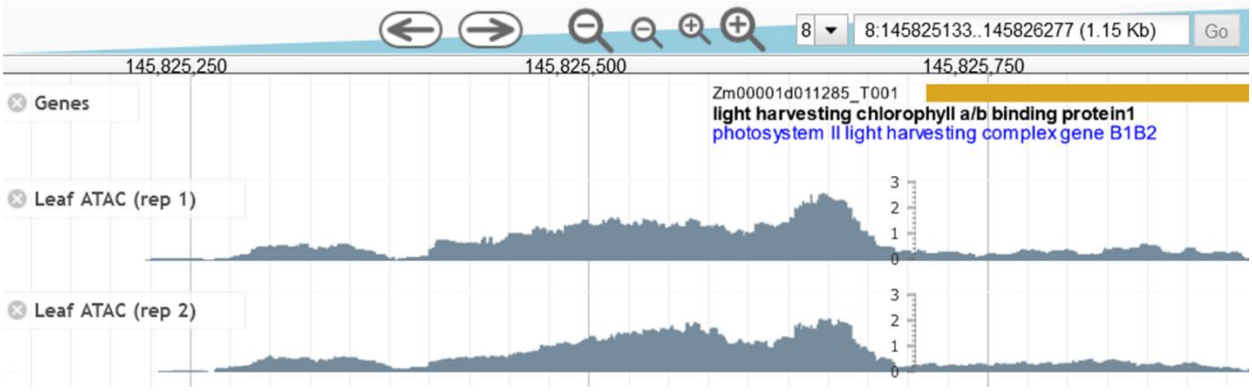

Figure 2-3. Chromatin accessibility of *lhcb10* promoter sequence as measured by ATAC-seq. Displayed is two reps from leaf-tissue data which shows the two regions of increased accessibility between -499 and -99 base pairs upstream of the transcription start site.

Figure 2-4: Results from GUS Staining Assay of β-Glucuronidase Expressed by the *lhcb10* Promoter in Various Maize Tissues

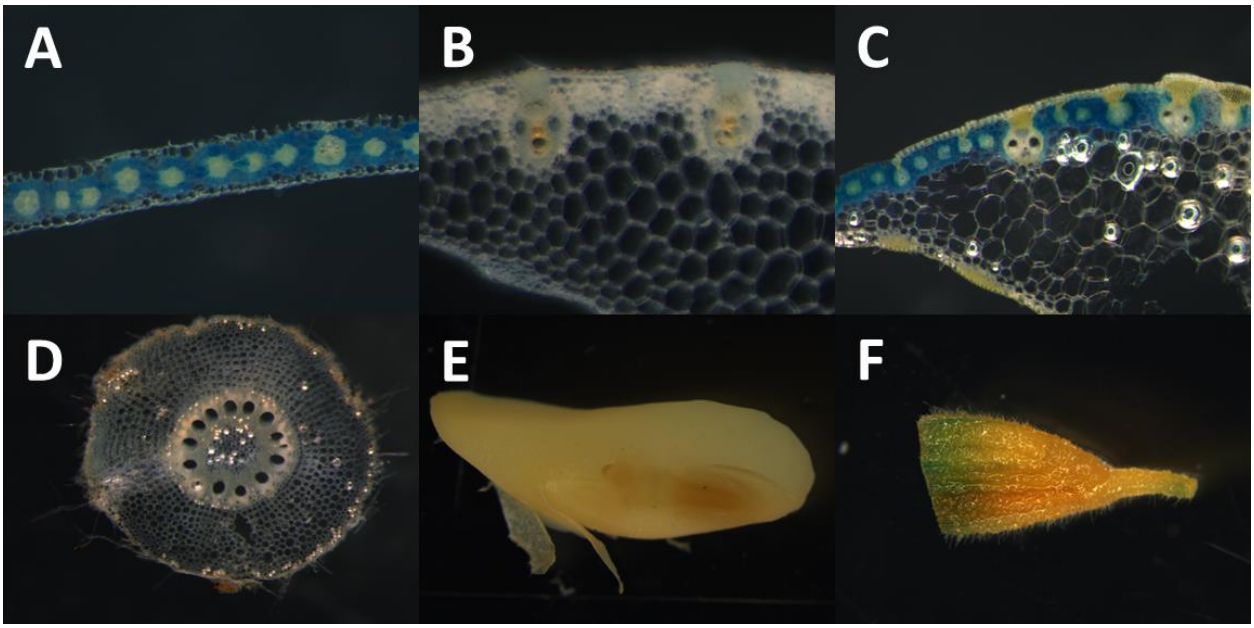

Figure 2-4. Results from GUS staining assay of β-Glucuronidase expressed under the *lhcb10* promoter in various maize tissues. Blue cells indicate presence of β-glucuronidase activity. A: V1 Leaf B: V5 Leaf Base C: R1 11th Leaf D: V9 Crown Roots E: 22 DAP Embryo F: Anther

Figure 2-5: Results from GUS Staining Assay of β-Glucuronidase Expressed by the *ris2* Promoter in Various Maize Tissues

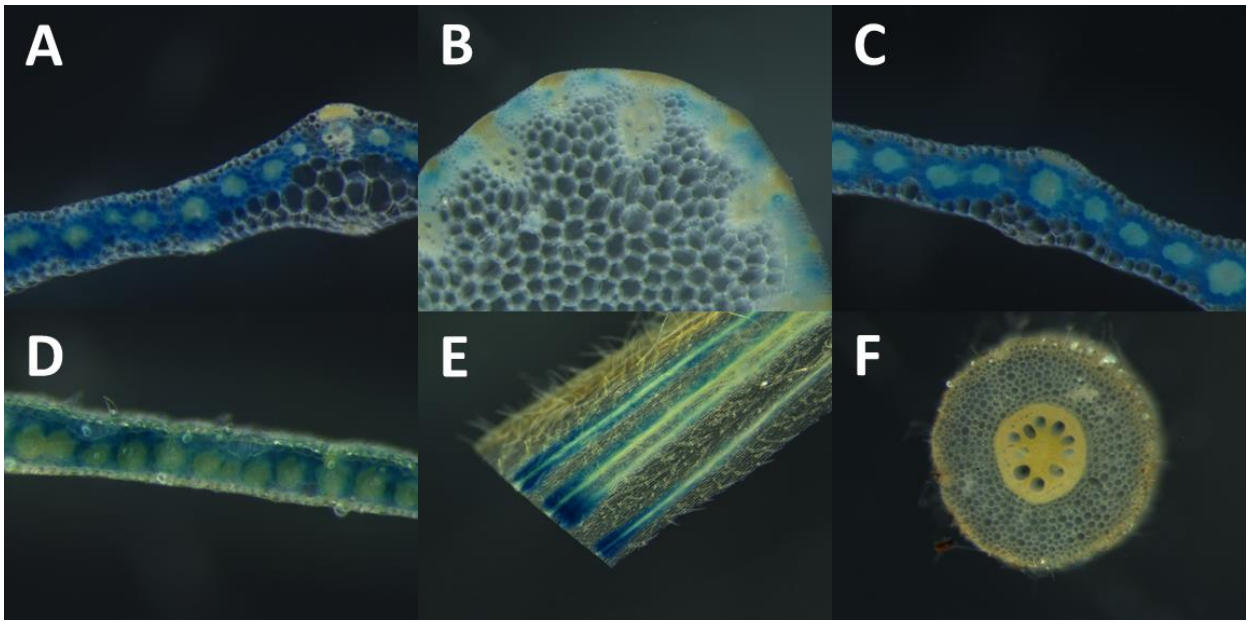

Figure 2-5. Results from GUS staining assay of β-Glucuronidase expressed under the *ris2* promoter in various maize tissues. Blue cells indicate presence of β-glucorinidase activity. A: V1 Leaf B: V5 Leaf Base C: V5 Leaf Tip D: R1 11th Leaf E: Glume F: V1 Primary Root

Figure 2-6: Results from GUS Staining Assay of β-Glucuronidase Expressed by the *aa2m* Promoter in Various Maize Tissues

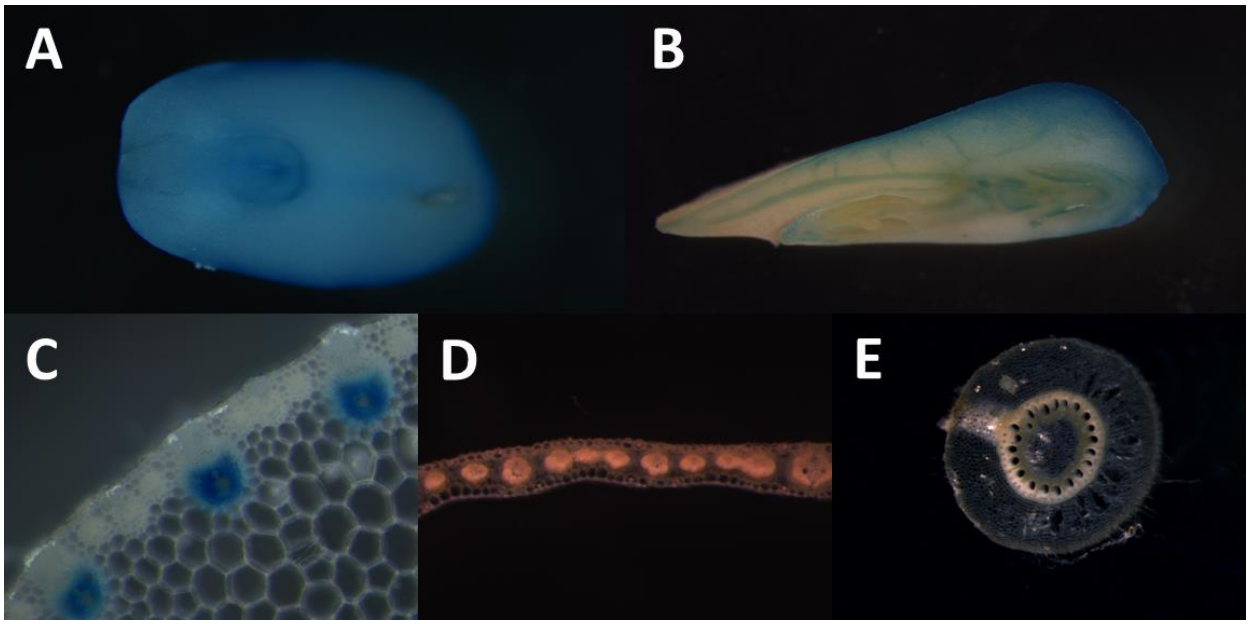

Figure 2-6. Results from GUS staining assay of β-Glucuronidase expressed under the *aa2m* promoter in various maize tissues. Blue cells indicate presence of β-glucorinidase activity. A: 11 DAP Embryo B: 22 DAP Embryo C: V5 Leaf Base D: V5 Leaf Tip E: V9 Crown Roots

Figure 2-7: Results from GUS Staining Assay of β-Glucuronidase Expressed by the *clo2b* Promoter in Various Maize Tissues

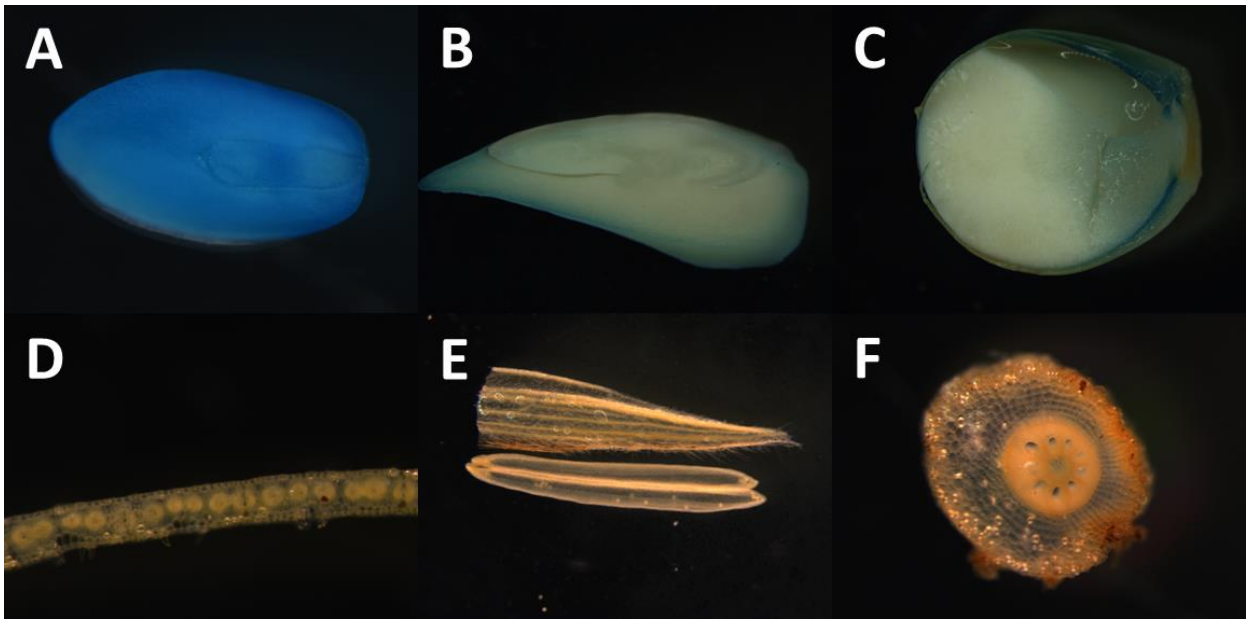

Figure 2-7. Results from GUS staining assay of β-Glucuronidase expressed under the *clo2b* promoter in various maize tissues. Blue cells indicate presence of β-glucorinidase activity. A: 11 DAP Embryo B: 22 DAP Embryo C: 22 DAP Endosperm D: R1 11th Leaf E: Anther F: V9 Crown Root

Figure 2-8: Results from GUS Staining Assay of β-Glucuronidase Expressed by the expb14 Promoter in Various Maize Tissues

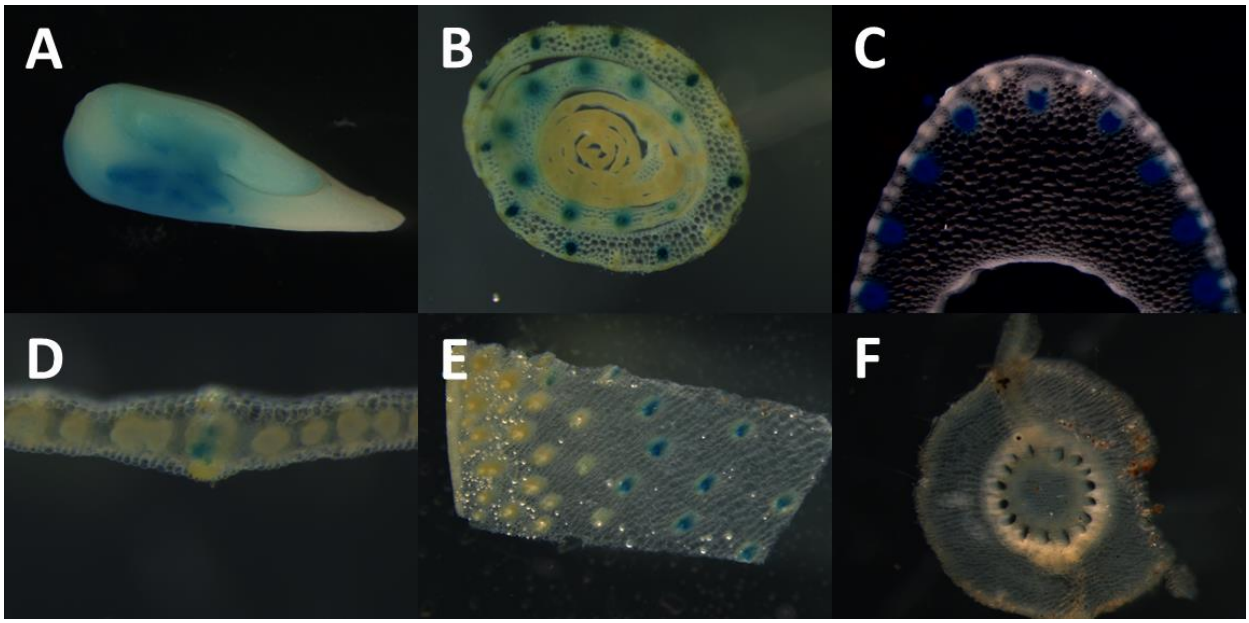

Figure 2-8. Results from GUS staining assay of β-Glucuronidase expressed under the *expb14* promoter in various maize tissues. Blue cells indicate presence of β-glucorinidase activity. A: 22 DAP Embryo B: V1 Leaves Base C: V5 Leaf Base D: V5 Leaf Tip E: V9 4th Internode F: V9 Crown Root

Figure 2-9: Results from GUS Staining Assay of β-Glucuronidase Expressed by the *nas2* Promoter in Various Maize Tissues

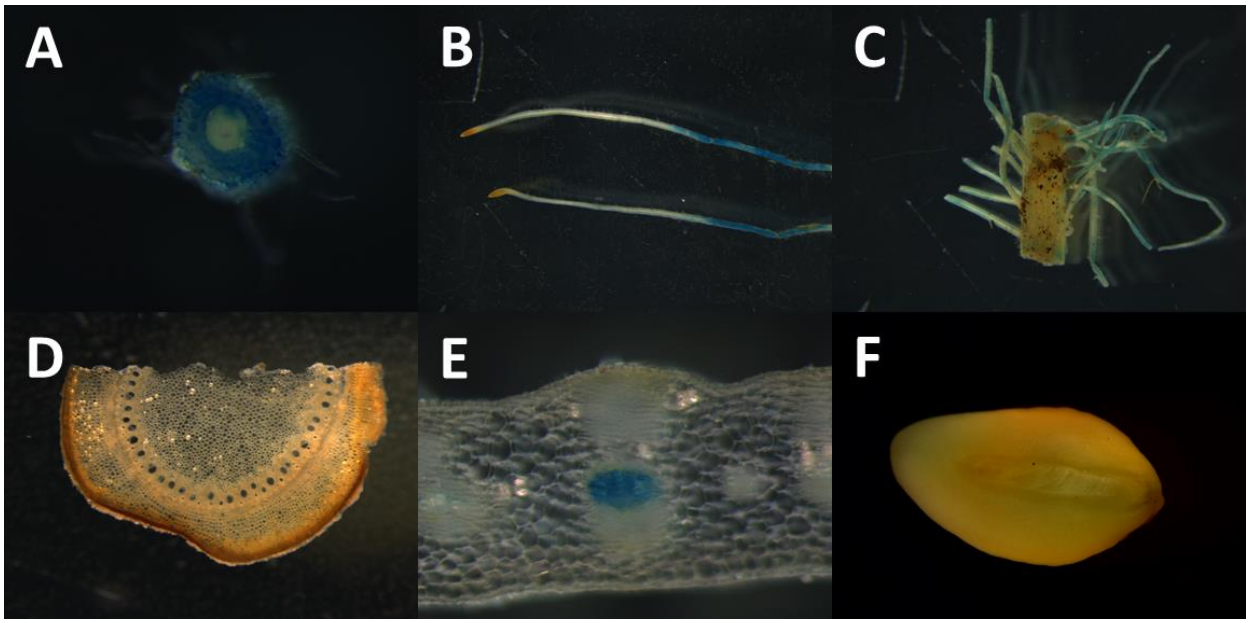

Figure 2-9. Results from GUS staining assay of β-Glucuronidase expressed under the *nas2* promoter in various maize tissues. Blue cells indicate presence of β-glucorinidase activity. A: V1 Primary Root B: V1 Primary Roots C: V9 Crown Roots D: Brace Root E: V9 8th Leaf Base F: 22 DAP Embryo

Figure 2-10: Results from GUS Staining Assay of β-Glucuronidase Expressed by the Version 2 Promoters in Various Maize Tissues

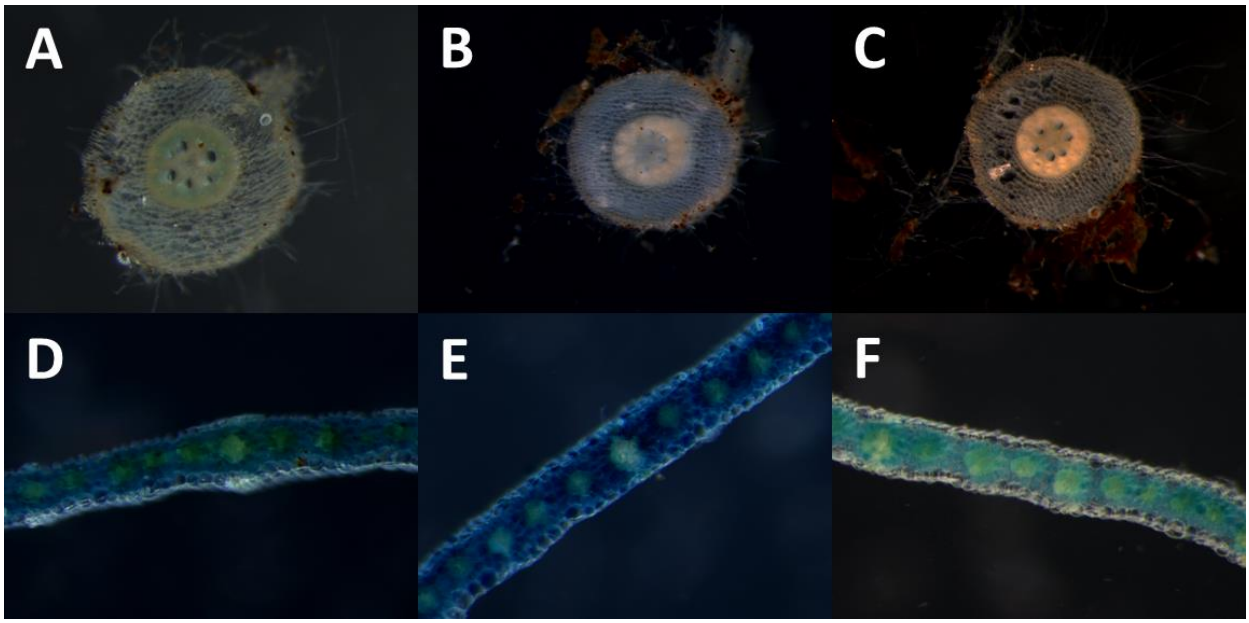

Figure 2-10. Results from GUS staining assay of β-Glucuronidase expressed under the version 2 promoters in various maize tissues. Blue cells indicate presence of β-glucorinidase activity. A: V1 Primary Root – *ris2* minimized promoter B: V1 Primary Root – *lhcb10* minimized promoter C: V1 Primary Root – *lhcb10* fragment and CaMV 35S core promoter D: V1 Pooled Leaves – *ris2* minimized promoter E: V1 Pooled Leaves – *lhcb10* minimized promoter F: V1 Pooled Leaves – *lhcb10* fragment and CaMV 35S core promoter

Aknowledgement

We wish to acknowledge scientists and technical staff who assisted with this research, including R. Collier, N. Walter, and B. Martinell.

Tables

Table 2-1: Stages of Maize Growth Tested for Promoter Activity as Assayed by β-Glucorinidase Activity

Stage	Tissues Assayed
V1	Stem and shoot, 1st leaf, primary root
V5	Base of 6th leaf, Tip of 6th leaf
V9	8th leaf, 4th internode, crown roots from 3rd node
Reproductive Stage	11th leaf, brace roots, anthers, 11 days-after-pollination embryos, 22 days-after-pollination embryos, 22 days-after-pollination endosperm

Table 2-1. Stages of maize growth tested for promoter activity as assayed by β-glucorinidase activity. Stages were assessed by the number of visible leaf collars in the case of V1, V5, and V9. Reproduction was defined as the beginning of pollen shed.

Table 2-2: Maize Genes Assayed for Tissue-Specific Expression of Promoter Region through β-Glucorinidase Activity

V4 Gene Name	Common Name	Tissue Target
Zm00001d011467	*csu31a*	Leaf
Zm00001d044666	*fha9*	Leaf
Zm00001d037513	*glp1*	Leaf
Zm00001d011285	*lhcb10*	Leaf
Zm00001d053432	*ris2*	Leaf
Zm00001d020590	*alt5*	Seed
Zm00001d037436	*dzs18*	Seed
Zm00001d045792	*expb14*	Seed
Zm00001d009292	*fl3*	Seed
Zm00001d034413	*glb2*	Seed
Zm00001d002768	*ole*	Seed
Zm00001d025833	*clo2b*	Seed
Zm00001d028888	*nas2*	Root
Zm00001d026177	*por2*	Root

Table 2-2. Maize genes assayed for tissue-specific expression of promoter region through β-glucorinidase activity. The gene name from the 4th version of the maize genome sequence as well as the common name are provided. The tissue target was determined from RNA sequencing data from the maize Gene Atlas.

Table 2-3: MUG Assay Results to Determine Expression Levels of Candidate Promoters in Maize Leaf and Root Tissues

V4 Gene Name	Tissue	Event 1	Event 2	Event 3
lhcb10	Leaf	33.31*	105.86*	78.34*
	Root	3.75	3.95	4.44
ris2	Leaf	4.58	6.73*	---
	Root	2.77	3.98	---
lhcb10-min	Leaf	30.64*	73.64*	4.03*
	Root	3.21	3.11	3.91
ris2-min	Leaf	7.91*	8.76*	7.16*
	Root	2.72*	2.85	4.07
lhcb10-syn	Leaf	3.77*	5.98	6.24*
	Root	2.48*	1.72	3.16
ubi1	Leaf	87.24*		
	Root	48.26*		
LH244 Wild Type	Leaf	1.51*		
	Root	4.38*		

Table 2-3. MUG assay results to determine expression levels of candidate promoters in maize leaf and root tissues. "*lhcb10-min*": shortened *lhcb10* promoter. "*ris2-min*": shortened *ris2* promoter. "*lhcb10-syn*": conjugate promoter created with the *lhcb10* fragment with the CaMV35S promoter. All measurements are in units of nmol 4-MU min-1 mg-1. All measurements shown are an average of three replicates per event. *: The sample is significantly different in MUG activity compared to the LH244 control samples, which have a mean of 4.38 for root tissue and 1.51 for leaf tissue.

References

Abendroth L, Elmore R, Boyer M, Marlay S (2011) Corn Growth and Development. Iowa State University Extension, Ames, Iowa

Bae JM, Kwak MS, Noh SA, et al (2014) Overexpression of sweetpotato expansin cDNA (IbEXP1) increases seed yield in Arabidopsis. Transgenic Res 23:657–667. https://doi.org/10.1007/S11248-014-9804-1/FIGURES/5

Bansal KC, Bogoradt L (1993) Cell type-preferred expression of maize cab-ml: Repression in bundle sheath cells and enhancement in mesophyll cells (C4 photosynthesis/AT-I sequence/light-harvesting chlorophyll a/b-binding proteins of photosystem Hl/Zea maysl mesophyll-specifying region). Plant Biol 90:4057–4061

Barkan A, Walker M (1994) Isolation and genomic mapping of cDNAs encoding the chloroplast Rieske Fe/S protein: two unlinked genes are transcribed in maize. MAIZE Genet. Coop. Newsl. 41–41

Bezrutczyk M, Zöllner NR, Kruse CPS, et al (2021) Evidence for phloem loading via the abaxial bundle sheath cells in maize leaves. Plant Cell 33:531. https://doi.org/10.1093/PLCELL/KOAA055

Broothaerts W, Mitchell HJ, Weir B, et al (2005) Gene transfer to plants by diverse species of bacteria. Nature 433:629–633. https://doi.org/10.1038/nature03309

Brothaerts W, Mitchell HJ, Weir B, et al (2005) Gene transfer to plants by diverse species of bacteria. Nat 2005 4337026 433:629–633. https://doi.org/10.1038/NATURE03309

Castillo FM, Canales J, Claude A, Calderini DF (2018) Expansin genes expression in growing ovaries and grains of sunflower are tissue-specific and associate with final grain weight. BMC Plant Biol 18:1–14. https://doi.org/10.1186/S12870-018-1535-7/TABLES/3

Chen JCF, Tsai CCY, Tzen JTC (1999) Cloning and Secondary Structure Analysis of Caleosin, a Unique Calcium-Binding Protein in Oil Bodies of Plant Seeds. Plant Cell Physiol 40:1079–1086. https://doi.org/10.1093/OXFORDJOURNALS.PCP.A029490

Chow CN, Lee TY, Hung YC, et al (2019) PlantPAN3.0: a new and updated resource for reconstructing transcriptional regulatory networks from ChIP-seq experiments in plants. Nucleic Acids Res 47:D1155–D1163. https://doi.org/10.1093/NAR/GKY1081

Christensen AH, Sharrock RA, Quail PH (1992) Maize polyubiquitin genes: structure, thermal perturbation of expression and transcript splicing, and promoter activity following transfer to protoplasts by electroporation. Plant Mol Biol 18:675–689. https://doi.org/10.1007/BF00020010/METRICS

Curie C, Cassin G, Couch D, et al (2009) Metal movement within the plant: contribution of nicotianamine and yellow stripe 1-like transporters. Ann Bot 103:1–11. https://doi.org/10.1093/AOB/MCN207

David Freedman, Robert Pisani RP (2007) Statistics, 4th edn. W.W. Northon and Company, Inc, New York, New York

Dietz-Pfeilstetter A, Arndt N, Manske U (2016) Effects of a petunia scaffold/matrix attachment region on copy number dependency and stability of transgene expression in Nicotiana

tabacum. Transgenic Res 25:149–162. https://doi.org/10.1007/S11248-015-9924-2/FIGURES/4

Engler C, Youles M, Gruetzner R, et al (2014) A Golden Gate modular cloning toolbox for plants. ACS Synth Biol 3:839–843. https://doi.org/10.1021/sb4001504

Erenstein O, Jaleta M, Sonder K, et al (2022) Global maize production, consumption and trade: trends and R&D implications. Food Secur 2022 145 14:1295–1319. https://doi.org/10.1007/S12571-022-01288-7

Ermakova M, Arrivault S, Giuliani R, et al (2021) Installation of C4 photosynthetic pathway enzymes in rice using a single construct. Plant Biotechnol J 19:575–588. https://doi.org/10.1111/PBI.13487

Faleiros RRS, Seebauer JR, Below FE (1996) Nutritionally Induced Changes in Endosperm of shrunken-1 and brittle-2 Maize Kernels Grown In Vitro. Crop Sci 36:947–954. https://doi.org/10.2135/CROPSCI1996.0011183X003600040022X

Gupta PK, Kulwal PL, Jaiswal V (2019) Association mapping in plants in the post-GWAS genomics era. Adv Genet 104:75–154. https://doi.org/10.1016/BS.ADGEN.2018.12.001

Han Y, Hu M, Ma X, et al (2022) Exploring key developmental phases and phase-specific genes across the entirety of anther development in maize. J Integr Plant Biol 64:1394–1410. https://doi.org/10.1111/JIPB.13276/SUPPINFO

Hanano A, Blée E, Murphy DJ (2023) Caleosin/peroxygenases: multifunctional proteins in plants. Ann Bot 131:387–409. https://doi.org/10.1093/AOB/MCAD001

Haydon MJ, Kawachi M, Wirtz M, et al (2012) Vacuolar Nicotianamine Has Critical and Distinct Roles under Iron Deficiency and for Zinc Sequestration in Arabidopsis. Plant Cell 24:724–737. https://doi.org/10.1105/TPC.111.095042

Hernandez-Pinzon I, Patel K, Murphy DJ (2001) The Brassica napus calcium-binding protein, caleosin, has distinct endoplasmic reticulum- and lipid body-associated isoforms. Plant Physiol Biochem 39:615–622. https://doi.org/10.1016/S0981-9428(01)01274-8

Higuchi K, Suzuki K, Nakanishi H, et al (1999) Cloning of Nicotianamine Synthase Genes, Novel Genes Involved in the Biosynthesis of Phytosiderophores. Plant Physiol 119:471–480. https://doi.org/10.1104/PP.119.2.471

Hily J-M, Singer SD, Yang Y, Liu Z (2009) A transformation booster sequence (TBS) from Petunia hybrida functions as an enhancer-blocking insulator in Arabidopsis thaliana. Plant Cell Rep 28:1095–104. https://doi.org/10.1007/s00299-009-0700-8

Hufford MB, Seetharam AS, Woodhouse MR, et al (2021) De novo assembly, annotation, and comparative analysis of 26 diverse maize genomes. Science 373:655–662. https://doi.org/10.1126/SCIENCE.ABG5289

ISAAA GM Approval Database | ISAAA.org. https://www.isaaa.org/gmapprovaldatabase/default.asp. Accessed 2 May 2023

Jia S, Li A, Morton K, et al (2016) A population of deletion mutants and an integrated mapping and exome-seq pipeline for gene discovery in maize. G3 Genes, Genomes, Genet 6:2385–2395. https://doi.org/10.1534/G3.116.030528/-/DC1

Jiao Y, Peluso P, Shi J, et al (2017) Improved maize reference genome with single-molecule technologies. Nat 2017 5467659 546:524–527. https://doi.org/10.1038/nature22971

Kang M, Lee K, Finley T, et al (2022) An Improved Agrobacterium-Mediated Transformation and Genome-Editing Method for Maize Inbred B104 Using a Ternary Vector System and Immature Embryos. Front Plant Sci 13:860971. https://doi.org/10.3389/FPLS.2022.860971/BIBTEX

Klusch N, Dreimann M, Senkler J, et al (2022) Cryo-EM structure of the respiratory I + III2 supercomplex from Arabidopsis thaliana at 2 Å resolution. Nat Plants 2022 91 9:142–156. https://doi.org/10.1038/S41477-022-01308-6

Kumar K, Gambhir G, Dass A, et al (2020) Genetically modified crops: current status and future prospects. Planta 2020 2514 251:1–27. https://doi.org/10.1007/S00425-020-03372-8

Lazo GR, Stein PA, Ludwig RA (1991) A DNA Transformation–Competent Arabidopsis Genomic Library in Agrobacterium. Bio/Technology 1991 910 9:963–967. https://doi.org/10.1038/NBT1091-963Li H, Yang M, Zhao C, et al (2021) Physiological and proteomic analyses revealed the response mechanisms of two different drought-resistant maize varieties. BMC Plant Biol 21:. https://doi.org/10.1186/S12870-021-03295-W

Li X, Zhang X, Gao S, et al (2022) Single-cell RNA sequencing reveals the landscape of maize root tips and assists in identification of cell type-specific nitrate-response genes. Crop J 10:1589–1600. https://doi.org/10.1016/J.CJ.2022.02.004

Lizong H, Jinjie G, Yongfeng Z, et al (2014) Molecular Characterization and Evolutionary Analysis of the Putative Caleosin Gene Family in Maize (Zea mays). Int J Agric Biol

Lori Tausta S, Li P, Si Y, et al (2014) Developmental dynamics of Kranz cell transcriptional specificity in maize leaf reveals early onset of C4-related processes. J Exp Bot 65:3543–3555. https://doi.org/10.1093/JXB/ERU152

Lowe K, Wu E, Wang N, et al (2016) Morphogenic Regulators Baby boom and Wuschel Improve Monocot Transformation. Plant Cell 28:1998–2015. https://doi.org/10.1105/tpc.16.00124

Lu Z, Ricci WA, Schmitz RJ, Zhang X (2018) Identification of cis-regulatory elements by chromatin structure. Curr Opin Plant Biol 42:90–94. https://doi.org/10.1016/j.pbi.2018.04.004

Mahmud K, Makaju S, Ibrahim R, Missaoui A (2020) Current Progress in Nitrogen Fixing Plants and Microbiome Research. Plants 2020, Vol 9, Page 97 9:97. https://doi.org/10.3390/PLANTS9010097

Marand AP, Chen Z, Gallavotti A, Schmitz RJ (2021) A cis-regulatory atlas in maize at single-cell resolution. Cell 184:3041-3055.e21. https://doi.org/10.1016/J.CELL.2021.04.014

Mazaheri M, Heckwolf M, Vaillancourt B, et al (2019) Genome-wide association analysis of stalk biomass and anatomical traits in maize. BMC Plant Biol 19:1–17. https://doi.org/10.1186/s12870-019-1653-x

Mizuno D, Higuchi K, Sakamoto T, et al (2003) Three Nicotianamine Synthase Genes Isolated from Maize Are Differentially Regulated by Iron Nutritional Status. Plant Physiol 132:1989–1997. https://doi.org/10.1104/PP.102.019869

Næsted H, Frandsen GI, Jauh GY, et al (2000) Caleosins: Ca2+-binding proteins associated with lipid bodies. Plant Mol Biol 44:463–476. https://doi.org/10.1023/A:1026564411918

Nelms B, Walbot V (2019) Defining the developmental program leading to meiosis in maize. Science (80-) 364:52–56. https://doi.org/10.1126/SCIENCE.AAV6428

Nguyen K Le, Grondin A, Courtois B, Gantet P (2019) Next-Generation Sequencing Accelerates Crop Gene Discovery. Trends Plant Sci 24:263–274. https://doi.org/10.1016/J.TPLANTS.2018.11.008

Nuccio ML (2018) A Brief History of Promoter Development for Use in Transgenic Maize Applications. Methods Mol Biol 1676:61–93. https://doi.org/10.1007/978-1-4939-7315-6_4

Patro R, Duggal G, Love MI, et al (2017) Salmon provides fast and bias-aware quantification of transcript expression. Nat Methods 2017 144 14:417–419. https://doi.org/10.1038/NMETH.4197

Pérez-González A, Caro E (2019) Benefits of using genomic insulators flanking transgenes to increase expression and avoid positional effects. Sci Reports 2019 91 9:1–11. https://doi.org/10.1038/S41598-019-44836-6

Pick TR, Bräutigam A, Schlüter U, et al (2011) Systems Analysis of a Maize Leaf Developmental Gradient Redefines the Current C4 Model and Provides Candidates for Regulation. Plant Cell 23:4208–4220. https://doi.org/10.1105/TPC.111.090324

Pörtner HO, Roberts DC, Tignor M, et al (2022) IPCC, 2022: Climate Change 2022: Impacts, Adaptation and Vulnerability. Cambridge, UK and New York, NY, USA

R-Core-Team (2023) R: A language and environment for statistical

Rajeevkumar S, Anunanthini P, Ramalingam S (2015) Epigenetic silencing in transgenic plants. Front Plant Sci 6:693. https://doi.org/10.3389/FPLS.2015.00693

Raji JA, Frame B, Little D, et al (2018) Agrobacterium and Biolistic-Mediated Transformation of Maize B104 Inbred

Reichman SM, Parker DR (2002) Revisiting the Metal-Binding Chemistry of Nicotianamine and 2′-Deoxymugineic Acid. Implications for Iron Nutrition in Strategy II Plants. Plant Physiol 129:1435–1438. https://doi.org/10.1104/PP.005009

Rellán-Álvarez R, Abadía J, Álvarez-Fernández A (2008) Formation of metal-nicotianamine complexes as affected by pH, ligand exchange with citrate and metal exchange. A study by electrospray ionization time-of-flight mass spectrometry. Rapid Commun Mass Spectrom 22:1553–1562. https://doi.org/10.1002/RCM.3523

Ricci WA, Lu Z, Ji L, et al (2019) Widespread long-range cis-regulatory elements in the maize genome. Nat Plants 5:1237–1249. https://doi.org/10.1038/s41477-019-0547-0

Rocca N La, Barbato R, Vecchia F dalla, Rascio N (2000) Cab gene expression in bleached leaves of carotenoid-deficient maize. Photosynth Res 64:119–126. https://doi.org/10.1023/A:1006477215572/METRICS

Römheld V, Marschner H (1986) Evidence for a Specific Uptake System for Iron Phytosiderophores in Roots of Grasses. Plant Physiol 80:175–180. https://doi.org/10.1104/PP.80.1.175

Satterlee JW, Strable J, Scanlon MJ (2020) Plant stem-cell organization and differentiation at single-cell resolution. Proc Natl Acad Sci U S A 117:33689–33699. https://doi.org/10.1073/PNAS.2018788117/SUPPL_FILE/PNAS.2018788117.SD12.TXT

Sekhon RS, Briskine R, Hirsch CN, et al (2013) Maize Gene Atlas Developed by RNA Sequencing and Comparative Evaluation of Transcriptomes Based on RNA Sequencing and Microarrays. PLoS One 8:. https://doi.org/10.1371/journal.pone.0061005

Sekhon RS, Lin H, Childs KL, et al (2011) Genome-wide atlas of transcription during maize development. Plant J 66:553–563. https://doi.org/10.1111/j.1365-313X.2011.04527.x

Shaner NC, Campbell RE, Steinbach PA, et al (2004) Improved monomeric red, orange and yellow fluorescent proteins derived from Discosoma sp. red fluorescent protein. Nat Biotechnol 22:1567–1572. https://doi.org/10.1038/nbt1037

Simkin AJ, McAusland L, Lawson T, Raines CA (2017) Overexpression of the RieskeFeS Protein Increases Electron Transport Rates and Biomass Yield. Plant Physiol 175:134–145. https://doi.org/10.1104/PP.17.00622

Singer SD, Hily JM, Cox KD (2011) Analysis of the enhancer-blocking function of the TBS element from Petunia hybrida in transgenic Arabidopsis thaliana and Nicotiana tabacum. Plant Cell Rep 30:2013–2025. https://doi.org/10.1007/S00299-011-1109-8/FIGURES/5

Stavolone L, Kononova M, Pauli S, et al (2003) Cestrum yellow leaf curling virus (CmYLCV) promoter: A new strong constitutive promoter for heterologous gene expression in a wide variety of crops. Plant Mol Biol 53:663–673. https://doi.org/10.1023/b:plan.0000019110.95420.bb

Stelpflug SC, Sekhon RS, Vaillancourt B, et al (2016) An Expanded Maize Gene Expression Atlas based on RNA Sequencing and its Use to Explore Root Development. Plant Genome 9:0. https://doi.org/10.3835/plantgenome2015.04.0025

Sun Q, Li Y, Gong D, et al (2022) A NAC-EXPANSIN module enhances maize kernel size by controlling nucellus elimination. Nat Commun 2022 131 13:1–14. https://doi.org/10.1038/S41467-022-33513-4

Takahashi M, Terada Y, Nakai I, et al (2003) Role of Nicotianamine in the Intracellular Delivery of Metals and Plant Reproductive Development. Plant Cell 15:1263–1280. https://doi.org/10.1105/TPC.15.010256

Tao S, Liu P, Shi Y, et al (2022) Single-cell transcriptome and network analyses unveil key transcription factors regulating mesophyll cell development in maize. Genes (Basel) 13:374. https://doi.org/10.3390/GENES13020374/S1

Thompson CJ, Movva NR, Tizard R, et al (1987) Characterization of the herbicide-resistance gene bar from Streptomyces hygroscopicus . EMBO J 6:2519–2523. https://doi.org/10.1002/j.1460-2075.1987.tb02538.x

Tibbs Cortes L, Zhang Z, Yu J (2021) Status and prospects of genome-wide association studies in plants. Plant Genome 14:e20077. https://doi.org/10.1002/TPG2.20077

Tnani H, López I, Jouenne T, Vicient CM (2011) Protein composition analysis of oil bodies from maize embryos during germination. J Plant Physiol 168:510–513. https://doi.org/10.1016/J.JPLPH.2010.08.020

Wang J, Jiang J, Oard JH (2000a) Structure, expression and promoter activity of two polyubiquitin genes from rice (Oryza sativa L.). Plant Sci 156:201–211. https://doi.org/10.1016/S0168-9452(00)00255-7

Wang J, Jiang J, Oard JH (2000b) Structure, expression and promoter activity of two polyubiquitin genes from rice (Oryza sativa L.). Plant Sci 156:201–211. https://doi.org/10.1016/S0168-9452(00)00255-7

Watson NR, Peschke VM, Russell DA, Sachs MM (1992) Analysis of l-alanine:2-oxoglutarate aminotransferase isozymes in maize. Biochem Genet 30:371–383. https://doi.org/10.1007/BF00569328/METRICS

Xu J, Zhu X, Yan F, et al (2022) Identification of Quantitative Trait Loci Associated With Iron Deficiency Tolerance in Maize. Front Plant Sci 13:858. https://doi.org/10.3389/FPLS.2022.805247/BIBTEX

Xu X, Crow M, Rice BR, et al (2021) Single-cell RNA sequencing of developing maize ears facilitates functional analysis and trait candidate gene discovery. Dev Cell 56:557. https://doi.org/10.1016/J.DEVCEL.2020.12.015

Yamaguchi-shinozaki K, Koizumi M, Urao S, Shinozaki K (1992) Molecular Cloning and Characterization of 9 cDNAs for Genes That Are Responsive to Desiccation in Arabidopsis thaliana: SequenceAnalysis of One cDNA Clone That Encodes a Putative Transmembrane Channel Protein. Plant Cell Physiol 33:217–224. https://doi.org/10.1093/OXFORDJOURNALS.PCP.A078243

Zheng L, Zhang X, Zhang H, et al (2019) The miR164-dependent regulatory pathway in developing maize seed. Mol Genet Genomics 294:501–517. https://doi.org/10.1007/S00438-018-1524-4/TABLES/3

Zhou X, Li S, Zhao Q, et al (2013) Genome-wide identification, classification and expression profiling of nicotianamine synthase (NAS) gene family in maize. BMC Genomics 14:238. https://doi.org/10.1186/1471-2164-14-238

CHAPTER 3: *MADS69* ACTS ACROSS MAIZE DEVELOPMENTAL PHASES TO INFLUENCE FLOWERING TIME

Abstract

Flowering time is a vital trait for agriculture and thus a target of genetic manipulation for crop improvement. The genetic control of flowering time, however, is complex and in need of further investigation to decipher the various mechanisms underlying this trait. Many plant genetic systems intersect at the developmental juncture of floral transition and separating the underlying causes represents a fundamental challenge in plant genetics. Within the globally important crop, maize (*Zea mays* L.), an important flowering time gene that has been identified is *mads69*, which was identified repeatedly in genome-wide association studies and quantitative trait loci studies on flowering time. Within maize inbred line, B73, *mads69* is broadly expressed in many tissues, raising the question of its function in those tissues and the impacts of variable gene expression. We have identified several traits affected by *mads69* expression including decreased root mass, increased leaf and tassel angles, and, surprisingly, an increase in the number of embryonic leaves within the maize embryo. Tissue-specific ectopic expression of *mads69* was used to demonstrate that both leaf- and embryo-specific expression induces earlier flowering in maize. Additionally, we have characterized the DNA-binding behavior of *mads69* through DNA affinity purification and sequencing, revealing binding sites at the *vgt* locus of *ZmRap2.7*. This study expands understanding of the role that *mads69* plays in the control of maize flowering time specifically while also revealing the complex traits that underlie flowering time in maize.

Significance Statement

The important maize flowering-time gene, *ZmMads69*, has been characterized in its DNA-binding, in its impact on flowering-time through expression in different tissues, and its pleiotropic effects on multiple traits. The spread of commercial maize production northward from its initial site of domestication required manipulation of its flowering time by plant breeders, but some of the underlying genes selected through the breeding process have been shown to be involved in many pleiotropic effects. *Mads69* is one such gene, which has been demonstrated to impact traits such as plant height and stalk diameter in addition to its impact on flowering time. We have characterized the DNA-binding profile of *mads69* and shown it has sites in the *vgt* locus which regulates *rap2.7*, in addition to identifying *bbx33* as another potential target. Through tissue-specific ectopic expression of *mads69*, we have shown leaf and embryo specific expression reciprocates the early-flowering phenotype. Finally, we have demonstrated additional impacts of *mads69* expression on increasing tassel and leaf angle, decreasing root mass, and increasing the number of leaves in the maize embryo.

Introduction

The domestication of wild plants into cultivated crops is one of the most successful and impactful innovations of humanity. We have not only increased the product yield of these species but have vastly increased their range of cultivation. Maize stands as a prominent example, with initial domestication occurring in Mexico and subsequently spreading northward, with cultivation now extending into Canada (Wang et al. 2017). Adaptation of flowering time has been pivotal in enabling this spread to new environments, evident by the ubiquity of flowering time genes as sites for domestication selection. This adaptation comes with a price, however: early flowering syndrome. Earlier flowering maize genotypes are smaller overall as

they have had less time to develop prior to maturity, with many physiological repercussions. Plant height and stalk diameter both decrease with earlier flowering, decreasing vegetative biomass (Salvi et al. 2007; Peiffer et al. 2014; Mazaheri et al. 2019). Early flowering plants tend to have less leaves below the primary ear (Li et al. 2015). Below ground traits mirror this, with root area, depth, and other traits have been found correlated with flowering time (Zheng et al. 2020). Furthermore, the period of "stay green" is shorter in earlier flowering varieties which limit the capacity for grain filling (Khanal et al. 2011; Sekhon et al. 2019). The cumulative effect of these traits can lead to decreased yield for the crop. Genes that underlie flowering time control, such as the *CCT* family, reflect this linkage of traits ranging from plant architecture to abiotic stress tolerance, with maize being no exception (Li and Xu 2017; Stephenson et al. 2019a; Liu et al. 2020). Dissecting these complex gene activities is vital to optimizing crop growth, made more important in a world under the threat of climate change. Several flowering time-related pathways have been delineated in maize: photoperiod, aging, sugar, and autonomous (Hill and Li 2016; Minow et al. 2018; Stephenson et al. 2019b). In addition, there are many abiotic stress pathways which can subsequently influence flowering time (Song et al. 2017; Leng et al. 2022; Gallo et al. 2023). Each of these pathways are independently sufficient to influence the flowering time of maize but converge on the florigen molecules, with *ZCN8* being the primarily identified gene. However, others, such as *ZCN7*, have been shown to have potential impacts on the plant (Meng et al. 2011; Lazakis et al. 2011; Mascheretti et al. 2015). In both Arabidopsis and rice, florigen expression has been localized to phloem companion cells specifically (Tamaki et al. 2007; Mathieu et al. 2007). The robustness of the flowering time decision apparatus also creates its phenotypic complexity. Almost every system within the plant can influence flowering time, which makes interpretation of experimental results challenging.

For example, the carbohydrate status of a maize plant can influence the choice to flower or not, with removal of leaves sufficient to stimulate this effect (Khangura et al. 2020). Resultantly, changes in the photosynthetic capacity of a plant can manifest as a "flowering time gene" despite the underlying cause being something much different. Delineating these possibilities is vital with genes that appear to be pleiotropic in nature.

Within maize, the MADS-box family of proteins are frequently linked to flowering time and in a connected manner are a common target of domestication (Schilling et al. 2018). They have been directly utilized to improve maize by enhancing grain yield or modifying plant height (Wu et al. 2019; Song et al. 2021). In particular, *mads69* (also known as *Zmm22*) has been identified repeatedly in QTL and GWAS studies, primarily as a flowering time regulator but linked to many other traits such as plant height, stalk diameter, tassel branch number, leaf length, and leaf number (Liang et al. 2019; Chen et al. 2020; Michel et al. 2022). Its role in flowering time has been investigated and it is theorized that at least part of its activity is through repression of *rap2.7*, which normally suppresses *zcn8* (Liang et al. 2019). One of the most curious features of *mads69* is its broad expression across many tissues (Figure 3-1) which, in combination with its pleiotropic phenotypic impacts, raises the natural question of whether this expression is spurious or impactful. In this investigation, research was conducted to further characterize the role of *mads69* in flowering time in maize, and its effects on pleiotropic traits when expressed in a tissue-specific manner. Furthermore, the characteristics of *MADS69* were defined through bioinformatic analysis, analysis of transgenics, and DNA affinity purification sequencing (DAP-seq) (O'Malley et al. 2016; Bartlett et al. 2017).

Results

Characterization of the *mads69* Gene

The maize *MADS69* protein was analyzed for domains using InterPro (Jones et al. 2014; Paysan-Lafosse et al. 2023), which identified a MEF2-like MADS-box domain, a coiled coil region, and a disordered region (Figure 3-2). Notably, no intervening domain (I), keratin-like domains (K), or C-terminal domains (C) were detected, which are typical of type II MADS-box genes (Gramzow and Theissen 2010). The structure as predicted by AlphaFold was visually analyzed (Figure 3-2), and showed the predicted MADS-box, coiled coil, and unstructured domains (Jumper et al. 2021; Varadi et al. 2022). The MADS-box domain appears to have two intact β-strands, necessary for MADS-box homo- and hetero- dimer formation (Pellegrini et al. 1995).

Present within the cDNA of *MADS69* is a 24kb intron which separates the first and second exons. This intron was analyzed across the maize NAM population for variation and the most prevalent version matched maize inbred line, B73, with a length of roughly 24kb and was contained in 17 genotypes (Hufford et al. 2021). The other major variant was a shorter, roughly 11kb, segment contained in maize inbred lines B97, Il14H, Ky21, P39, M37W, CML322, Ms71, and Oh7b. Finally, maize inbred line, Ki3, possessed an even longer version of the intron that was 53kb. None of these variants were significantly associated with changes in expression or flowering time.

The first exon of *mads69* contains the entirety of the MADS-box domain, making it a prime target for study. To elucidate the importance of proper intron excision as well as the domains of the protein, two overexpression cassettes were made. The first contained the full-length cDNA of *mads69* and the second had only the first exon. Both were expressed with the maize ubi1 promoter (Figure 3-3) Each of these constructs were subsequently delivered into

target explant tissues of maize inbred line, LH244, stable, T0 transgenic plants regenerated, and plants/lines containing single copy transgene inserts identified through segregation analysis in T1 seed/progeny. Lines representing three independent events for each vector were grown in the greenhouse and assessed for flowering time, measured as days to anthesis (DTA), and were compared to non-transgenic (null) progeny derived from seed from the same T0 cob. Sixteen plants were grown in the greenhouse for each event, 8 positives for the transgene and 8 negatives. Events containing the full-length cDNA flowered at 56.8 DTA on average, compared to 61.4 DTA for transgenic-negative plants, a statistically significant difference. First-exon expressing events flowered at 61.5 and 61.4 days for positive and negative plants respectively, which is not a significant difference. Node number was counted at anthesis. This trend was the same for the number of nodes, with full-length cDNA expressing events having 15 nodes on average compared to negative plants with 19.5. First-exon expressing plants had node numbers that were not significantly different from null plants.

The binding profile of *MADS69* was determined through DAP-seq, resulting in 5,447 putative binding sites identified (O'Malley et al. 2016; Bartlett et al. 2017). The top 600 peaks were analyzed with MEME for motifs and revealed the CArg-box binding site, which is typical of MADS-box proteins (Gramzow and Theissen 2010; Bailey et al. 2015). Potential targets of *mads69* in leaf tissue were identified in Liang et al. 2018, that indicated that *rap2.7* and *zcn8* may be part of a pathway through which *MADS69* might influence flowering time. Similar results were found following the DAP-seq analysis, with binding sites identified within 3kb upstream of *rap2.7* but none found associated with *zcn8* or any other florigen genes. Additionally, a binding site was identified within the *vgt*1 locus (Figure 3-4) which also overlaps with an unmethylated chromatin region and accessible chromatin from ATAC-seq data (Salvi et

al. 2007; Ricci et al. 2019; Hufford et al. 2021). The zinc-finger transcription factor *bbx33* was shown to contain *mads69* binding sites in its promoter region as well as being positively upregulated in *mads69* overexpression plants. These binding sites coincide with an unmethylated chromatin region and borders accessible chromatin (Figure 3-4). In total, 42 unique genes were found to overlap between the DAP-seq and over-expression studies.

Tissue-Specific Expression of *mads69* and its Pleiotropic Effects

To better understand what phenotypic impacts *mads69* expression has within different tissues, vectors in which *mads69* expression was driven by different tissue-specific promoters were constructed and transformed into maize line LH244. Specifically, tissue-specific promoters derived from the maize genes *lhcb10*, *clo2b*, and *nas2* were utilized which are leaf, embryo and primary root -specific, respectively (described in Chapter 2). Vectors design is shown in Figure 3-3. The constructs were each assessed for effects on flowering time in greenhouse and field. Eight rows of each of three events expressing *lhcb10*-promoter-driven *mads69* were measured in the field in 2022, with four of those rows being positive for the transgene and four were negative. Transgene-expressing plants for the 3 events had a DTA of 66.8 and a DTS of 64.3. Transgene-negative plants for the events respectively had a DTA of 67.8 and DTS of 65.3. This one-day difference translates to a 20 growing degree day (GDD) difference. The two other promoters, *clo2b* and *nas2*, were assessed in the greenhouse, initially in one event. *Clo2b*-promoter expressed *mads69* resulted in detection of significant differences in DTA, DTS and number of nodes. The differences were greater than the *lhcb10*-promoter driven *mads69* expression vectors, with positive vs negative events measured at 56.5 : 58.7 for DTA, 58.3 : 60.7 for DTS, and 17.8 : 18.8 for node number. Finally, for the *nas2*-promoter driven *mads69* events, there was not a

significant difference in flowering time, however node number significantly differed with transgene positive events having 17.8 nodes on average and null controls having 18.8. Subsequently, the early flowering phenotype of *clo2b*-promoter driven *mads69* was confirmed in a full set of three single-copy events, which were found to have a DTA of 54.7 : 56.8 for positive vs negative events.

In addition to investigation of the effects of tissue-specific expression of *mads69*, other traits were analyzed for the *mads69* overexpressing plants outside of typical flowering time measurements. Three events overexpressing the full-length mad69 transcript with the *ZmUbi1* promoter were assessed as seeds, with 10 seeds from transgene-positive and transgene-negative each. Neither difference in seed weight nor in embryo length were found between the two groups. However, the number of visible developed leaves were significantly different, with transgenic *mads69*-overexpressing seeds having 7.4 leaves on average and null seeds having 6.8. Those same events/lines were also assessed for differences in the angle of leaves and tassel branches in the greenhouse. Leaf and tassel angle was measured with a protractor at flowering time, with the bottom-most tassel measured and the leaf immediately above the ear measured. Sixteen plants per event were assessed, 8 transgene positive plants and 8 transgene negative ones. Significant differences were found in both traits between transgenic lines and controls, with mean leaf angle being 34.5° and tassel angle being 17.4° in positive events and mean leaf angle being 29.3° and tassel angle being 7.9° in negative events. Analysis of the effect of tissue-specific expression of *mads69* on root traits was conducted through the DIRT platform using image analysis with a variety of traits identified (Bucksch et al. 2014a; Das et al. 2015a). This analysis was done on plants from three independent events overexpressing *mads69*, again with the ubi1 promoter but a different construct as previously described (Mazaheri et al. 2019). The

study was done in the field in 2022 with 16 rows of plants for each event, 8 positive and 8 negative, and 3 plants from each row being analyzed. Transgene positive plants were found to have a 28% smaller root area compared to negative control plants, a significant difference. The number of days from planting to harvest was found to be significant in the ANOVA and the interaction of that term with transgenic status was nearly significant with a p-value of 0.0559. There was not an observed difference in the density of the roots or their angle.

Discussion

The Properties and Activity of *MADS69*

Results from our investigation indicate that *mads69* belongs to the MIKC-type MADS-box proteins. While a K domain was not detected, the coiled coil structure found shares many similarities with a K domain. Additionally, it clusters with MIKC-type proteins in phylogenetic analysis and shares the MEF2-type MADS domain of those proteins (Gramzow and Theissen 2010; Zhao et al. 2021). Resultantly, calling it a MIKC protein seems appropriate. Specifically, its shorter I domain would make it a MIKCc subtype (Zobell et al. 2010). Research into MIKC proteins has revealed many of their properties, with one of the most prominent being its dimerization. Both homo- and hetero- dimers have been observed in MIKC genes, which together create the CArG binding that is broadly observed, as well as specifically observed in the DAP-seq study. Several domains have been implicated in being necessary for dimerization to occur: The original crystallography work pointed to the β-strands as being a dimerization mediating domain which is contained within the mads-box domain itself. Subsequent work has indicated that the K-domain and its alpha-helices can be the controlling factor in some dimer pairings (Yang et al. 2003; Kaufmann et al. 2005). These dimers go on to interact with other

dimers to form multimeric complexes, which is the basis of the floral quartet model of flowers (Theißen et al. 2016). In addition to these within-family interactions, MADS-box proteins have been shown to interact with a wide variety of other proteins to enable their activating and repressive functions (Bemer et al. 2017; Herrera-Ubaldo et al. 2023). In the case of *AtAp1* and *AtSep3*, the C-domain seems to be a transcriptional activator specifically (Honma and Goto 2001).

Given this information, the lack of a change in flowering time in the transcript lacking the non-MADS-box domains is consistent with that these are vital to the functioning of *mads69*, though their exact functionality requires additional research. That study also showed that excision of the 25kb intron is necessary for *mads69* to function. While long introns are not uncommon in the maize genome, the 25kb intron within the *mads69* gene places it among <1% of introns (Schnable et al. 2009). The first intron of *mads69* possesses the typical GT and AG sequences at 5’ and 3’ boundaries, indicating it is likely a spliceosomally regulated intron (Poverennaya and Roytberg 2020). Functionally, this intron could influence *mads69* through many mechanisms as introns have been noted to impact transcripts throughout their cycle beginning at initiation, through transcript stability, and ending in translational efficiency (Gallegos and Rose 2015; Laxa 2017; Shaul 2017; Dwyer et al. 2021). Separating if and how this intron impacts *mads69* function requires additional research.

A CArG-box binding motif was identified by the DAP-seq analysis, which has the consensus sequence of 5’-CC[A/T]6GG-3’. This motif is plentiful within the maize genome alone but *mads69* seems selective in its sites. Available information indicates that dimerization is necessary for CArG-box recognition, and since the DAP-seq only utilized the *MADS69* protein, it seems likely that *mads69* forms homodimers. This does not rule out that heterodimers are also

formed by *MADS69* which may have altered binding specificity and act on different genes. The appearance of binding sites at *rap2.7*, both upstream of the transcription start site and within the *vgt* locus illustrates a potential explanation for how *rap2.7* is directly regulated by *mads69*. The *vgt* locus was identified through QTL mapping for flowering time and identified as a region 70kb upstream of the *rap2.7* gene and has been subsequently found to be important to adaptation to elevation and latitude (Salvi et al. 2002, 2007; Ducrocq et al. 2009). This region is conserved across maize, rice, and sorghum (Freeling and Subramaniam 2009). Despite this, the orthologous gene to *mads69* in rice is *OsMads51*, which has been found to likewise impact flowering time, though does not appear to act on the *rap2.7* orthologue (Song et al. 2007; Sun et al. 2012). The *vgt* locus is thought to positively regulate *rap2.7* expression, as MITE transposon insertion into the region results in methylation of *vgt* and downregulation of *rap2.7* (Castelletti et al. 2014). Studies of *mads69* indicate that it negatively regulates *rap2.7* expression (Liang et al. 2019). Within the context of the identified DAP-seq site, this means that *mads69* may act antagonistically to other transcription factors which form at *vgt* and would normally enhance expression of *rap2.7*.

Many genes were identified in the overlap between *mads69* binding sites and differentially regulated genes but among them only *bbx33* has documentation of being associated with flowering time. Broadly, the B-box proteins are a subgroup of zinc-finger transcription factors which are present in most plants and many members, such as CONSTANS in *Arabidopsis*, have been associated with flowering time control (Cheng and Wang 2005; Miller et al. 2008; Tripathi et al. 2017). Within maize, *bbx33* has been found associated with altered flowering time (Jamann et al. 2017). Regulation-wise, B-box proteins commonly act on one another and other transcription factors and within maize *bbx33* is especially connected to other

BBX proteins (Xu et al. 2023). *Bbx33* expression occurs in leaf and stalk tissue, especially in mature leaf tissues, which makes sense given its putative role in flowering (Sekhon et al. 2013; Stelpflug et al. 2016). Its activities beyond these details are unclear but the presence of *mads69* binding sites places it within the broader flowering network of maize.

Separating the Pleiotropic Effects of *MADS69*

This study revealed the manifestations of pleiotropy of *mads69*. Increased embryo leaf number, decreased root volume, and increased leaf and tassel angle were each detected in transgenic maize plants overexpressing *mads69*. The connection between root phenotypes and flowering time has been observed in maize before, with a recent study in elite inbred lines finding that 21% of the genetic variation in crown root traits was linkable to flowering time variation (Sha et al. 2023). MADS-box proteins have been identified as influencing root development in *Arabidopsis* such as *AtXal1* and *AtAgl21*, both of which shape root development (Yu et al. 2014; Alvarez-Buylla et al. 2019). More directly, *OsMads50* has been found to reduce the number of crown roots when overexpressed (Shao et al. 2019). Other genes involved in the flowering time process have been linked to root traits, such as *rap2.7* which has been shown to impact brace root number (Li et al. 2019). *Mads69* overexpressing plants were found to have a 28% smaller root area compared to null plants. This might be due to the gene acting in the roots itself or might just be due to lack of time spent developing. There was not an observed difference in the density of the roots or their angle. While our experimental design did not allow us to separate the time of harvest as a term in the model, the decrease in size is quite large and worthy of continued investigation.

The analysis of transgenic maize seed overexpressing *mads69* revealed surprising correlations between flowering time and seed morphology. While no difference was found between seed weight and embryo size, a significant, detectable difference was found in the number of visible embryonic leaves. Within maize, not many MADS-box genes are active in the embryo, and none have been linked with embryo-specific phenotypes (Zhao et al. 2021). In *Arabidopsis*, it is Type I MADS-box genes such as AtAgl23 that have been found to be the most active in embryo development (Colombo et al. 2008; Gramzow and Theissen 2010; Masiero et al. 2011). Flowering time has been found to be linked with seed phenotypes as in the case of MtSoc1c which, when overexpressed, results in smaller seeds and much earlier flowering, but separating the reduction in biomass from this result is very challenging (Yuan et al. 2023). The detected increase in leaf number is in the opposite direction of what would be expected from a merely weak plant and thus represents a novel direction of inquiry into flowering time.

Tissue-specific expression of *mads69* revealed that the overall effect of faster flowering in *mads69*-overexpressing plants is the cumulative effect across several tissues. *Lhcb10*-driven expression confirmed previous work showing that *mads69* is active and impactful in leaf-based flowering pathways (Liang et al. 2019). The *lhcb10* promoter provides expression solely in the mesophyll and is absent from bundle-sheath and vascular cells. The pathway from mesophyll tissues to the phloem, where florigen is active (Tamaki et al. 2007; Mathieu et al. 2007), is still unclear. Single-cell sequencing could potentially elucidate the sub-tissues which genes are active within, but current studies have examined leaves too young to demonstrate the flowering pathway (Bezrutczyk et al. 2021; Tao et al. 2022; Marand et al. 2022). Mesophyll-expression being sufficient for early flowering is consistent with communication occurring between the mesophyll and vasculature.

The results of *clo2b*-driven *mads69* expression were surprising given the developmental time that separates the embryo from anthesis. A potential explanation is that signaling for the floral transition has already begun at the embryo stage, inducing an acceleration of leaf formation in the embryo. This induction is then carried through to anthesis. Finally, root specific expression of *mads69* did not significantly impact flowering time but did reduce node number, demonstrating that node number may not be wholly linked to flowering time.

There are several potential explanations for the pleiotropy observed with *mads69* expression. First, *mads69* could be acting on different genes in various tissues due to differences in chromatin status and availability of cofactors in those tissues. Second, differences could be arising from heterodimer formation with other MADS-box proteins which are specific to the tissue, which influences the DNA binding preferences as well as types of multimeric complexes that are formed. Determination of the exact causative agent(s) requires additional research to separate.

This research has demonstrated the important and multifaceted role that *mads69* plays in flowering time as well as characterized some of the protein's features. DAP-seq results show that *MADS69* seems to directly act on *rap2.7* through the *vgt* locus, as well as on other genes such as *bbx33*. *MADS69* impacts flowering time through non-leaf tissues, with embryo-specific expression being sufficient to change anthesis. Many traits are impacted by *MADS69* overexpression ranging from leaf angle to root mass. We also demonstrated that the large first exon (25kb) must be excised for proper functionality of the gene, pointing to the important role that the later exons have. Subsequent research will focus on other potential targets of *MADS69*, the role the embryo plays in flowering time, and the specific functionality of the different domains of *MADS69*.

Materials and Methods

Greenhouse-Grown Plant Materials

Transgenic plants grown in the greenhouse were arranged in a completely randomized design. Carlin SVD-250 pots were arranged in Carlin 3 – 234 ST-I-0804 vacuum standard inserts. A custom potting mix based on “Jolly Gardener” brand Pro-Line C/B (Part No. 18-1010) was distributed to the pots. Seeds of maize inbred line, LH244, were sown by placing one seed per cell, pressing down 0.75 inch to 1 inch, then covering with loose, dry soil medium. Flats were placed on benches and manually watered until day 7 to 10, when flood fertigation was begun. Seedlings were grown in flats in the nursery area of the greenhouse for 18 to 21 days.

Greenhouse rooms were illuminated with high pressure sodium lamp fixtures (model #HS 2000 from PL lighting) with 600 W Phillips E39 HPS bulbs and supplemented with natural sunlight. The light fixtures were mounted about 2.7 m above the floor. Seeds were germinated on a bench top approximately 1.7 m below the lights. Custom LED fixtures were used for supplemental lighting. The LED fixtures were 600 W and 1.3 m from the bench top. Average photosynthetic photon flux density (PPFD) was 500 µmol m-2 s-1 on the bench top but varied between 370 and 660 µmol m-2 s-1 depending on the age of fixtures, the time of day and time of year.

After 18 to 21 days, plants were transferred to size 1200 Elite Blow-molded Nursery Containers (Carlin, Inc. catalog # 4-2022). The potting mix was the same as above. Pots were placed into carts, with the cart surface approximately 1.7 m below the light fixtures. After 7 to 9 days, pots were moved to the floor and placed on plastic pallets.

The temperature set points for the greenhouse were 26.7°C during the day and 21.1°C during the night. The photoperiod consisted of a 16-hr light: 8-hr dark with supplemental sunlight during the light period. The fertigation mix utilized was Peters Excel 15-5-15 Cal Mag Special (Carlin item number 20-235). Plants were on timed fertigation cycles based on growth stage and demand.

Field Plant Materials

Phenotypic data was collected on transgenic and null plants at the West Madison Agricultural Research Station in Madison, WI in 2022. Plants were grown in one-row plots (3.8 m long and 0.76 m apart) in a randomized complete block design at a density of 20 plants per row. Temperature data was collected by a nearby weather station.

Plant Phenotyping

Node Number: All leaf-bearing nodes were counted at anthesis.

Tassel Angle: The angle between the lowest branch of the tassel and the main stalk was considered the tassel angle.

Leaf Angle: Measurements were taken at flowering time on the leaf just above the top-most ear. A protractor was used to measure the angle between the leaf and the internode just above the collar of the leaf.

Days to Anthesis: Plants were counted as at anthesis if at least 50% of the tassel was shedding pollen. Plants were checked in the afternoon for pollen shed.

Days to Silking: Plants were considered as having emerged silks if more than 5 silk strands were visible from the ear.

Embryo Length: Embryos were measured using a fine-scale ruler on seeds that had been soaked in water overnight. Embryos were measured laterally from the base to the tip.

Embryo Leaf Number: Seeds to be dissected were soaked overnight in water to make material easier to dissect. For each seed, with the embryo-side facing outward, a razor blade was used to make a vertical incision through the center of the embryo, bisecting it. Subsequently, parallel incisions were made to the left and right of the primary incision to create thin slices. The slices were freed from the seed and stained for one minute in toluidine blue and then rinsed with 70% ethanol. Slices were stored in 70% ethanol until evaluation under a microscope. Leaves were counted manually in the slices using a Leica stereomicroscope (model M165 FC). Any primordia were counted regardless of size if they were visually identifiable. Scalpel blades were used to manipulate tissue as necessary to count the leaves.

Root Phenotypes: Roots were assessed using the "shovelomics" process and scored using the Digital imaging of root traits (DIRT) analysis pipeline (Trachsel et al. 2011; Bucksch et al. 2014b; Colombi et al. 2015; Das et al. 2015b). Roots were harvested within a two-week window after they had reached anthesis. Images were taken on a Sony A6000 camera.

Characterization of *MADS69*

The canonical full-length protein sequence of *mads69* was submitted to InterPro and scanned across all databases at the default cutoffs (Jones et al. 2014; Paysan-Lafosse et al. 2023). Analysis of the intron sequence was done on the first intron of *mads69* homologues across the maize NAM population which were aligned using ClustalOmega hosted on the European Bioinformatics Institute website (Goujon et al. 2010; Sievers et al. 2011; Hufford et al. 2021). Default settings were used for the alignment.

Cloning of Plasmids

Vectors were all created using the Golden-Gate cloning process (Engler et al. 2014). The cDNAs for *MADS69*, both the full-length and first-exon bearing, were synthesized by Genscript as a level 0 part and paired with the maize *ZmUbi1* promoter. In addition, *MADS69*-expressing constructs were made with the full-length cDNA paired with tissue-specific promoters derived from the maize genes *lhcb10*, *clo2b*, and *nas2* (described in Chapter 2). Each of these were assembled into level 2 vectors along with the parts necessary for maize transformation. The selectable marker gene, bar, was driven by the rice *OsUbq2* promoter (Thompson et al. 1987; Wang et al. 2000). Expression of the screenable marker gene, tdTomato, was driven by the yellow-leaf curling virus promoter (Stavolone et al. 2003). Bar and tdTomato expression cassettes were separated from the *MADS69*-expression cassette by the TBS insulator (Hily et al. 2009). Vectors maps are shown in Figure 3-3. Finally, a previously described *MADS69* overexpression vector and material derived from it was used in the characterization of *MADS69*'s effect on root morphology (Mazaheri et al. 2019).

Transformation of maize inbred line LH244

Test vectors were transformed into maize inbred LH244 using via a Type I embryonic callus-based system. The protocol utilized was originally developed for inbred B104 but worked for LH244 as well with minimal modifications (Kang et al. 2022). *Agrobacterium tumefaciens* strain AGL1 was used for all transformations treatments (Lazo et al. 1991). Single copy events were identified by examination of segregation ratios of tdTomato expression in T1 seeds. Sixty seeds were assayed per event and those with 33 positive seeds or less were deemed single copy

for falling within the expected 1:1 segregation ratio for transgenic to non-transgenic progeny derived from backcrossing the transgenic T0 plants to null LH244 plants.

DAP-seq

A DAP-seq genomic DNA library was generated as described previously (Bartlett et al. 2017). Briefly, 7.5 micrograms of phenol:chloroform:IAA-extracted DNA from maize B73 14-day old seedlings was diluted in EB (10mM Tris-HCl pH 8.5) and fragmented to 200bp fragments using a Covaris S2 sonicator. DNA was purified using AmpureXP beads with a 2:1 bead to DNA ratio, and then end-repaired using the End-It kit according to the manufacturer's recommendations. Samples were cleaned using a Qiaquick PCR purification kit (Qiagen), A-tailed using Klenow (3' to 5' exo-), cleaned with Qiaquick and ligated overnight with a truncated Illumina Y-adapter. The library was purified using AmpureXP beads with a 1:1 bead to DNA ratio and eluted in 40ul of EB. DNA concentration was determined using the Qubit HS kit.

A *MADS69* pENTR-ORF clone was recombined into the pIX-HALO vector using LR clonase II (ThermoFisher) and purified using a Qiagen miniprep kit. Protein expression was carried out using the TNT SP6 wheatgerm expression kit (Promega) using 1ug of plasmid DNA according to the manufacturer's recommendations. The sample was incubated for 2 hours at 30 degrees. Fifty microliters of expression reaction were then diluted in 50ul of wash buffer (1xPBS with 0.005% NP40) containing 10ul of Magne-HALO beads (Promega). The sample was rotated for 1 hour at room temperature and then washed three times with wash buffer. Beads were suspended in 50ul wash buffer and 1ug of B73 genomic DNA library diluted in wash buffer was added. The sample was then rotated at room temperature for 1 hour and washed seven times with wash buffer. To elute DNA, 30ul of EB was added and the sample was heated at 98C for 10

minutes. DNA was transferred to a new tube prior to 19 cycles of PCR enrichment and barcoding (Bartlett et al. 2017). DNA was sequenced on a HiSeqX with 150bp PE reads.

Mapping and Peak Calling

DAP-seq reads were trimmed using Trimmomatic (Bolger et al. 2014) with the following parameters: ILLUMINACLIP:TruSeq3-SE:2:30:10 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:50. Trimmed reads were mapped to the B73v5 reference genome using bowtie2.3.3 (Langmead and Salzberg 2012). Mapped reads were filtered for reads containing > MAPQ30 using samtools (samtools view -b -q 30). Peaks were called using GEM3 (Guo et al. 2012) with a q-value threshold of 0.00001 and the following parameters: --d Read_Distribution_default.txt --k_min 6 --k_max 20 –outNP –q 5. Background peaks were subtracted using a HALO-GST negative control sample and a list of blacklisted regions from (Galli et al. 2018).

Target Gene Assignment and Motif Analysis

Putative target genes were assigned to peaks using ChIPseeker (Yu et al. 2015). For motif analysis, sequences corresponding to the top 600 peaks (ranked by adjusted p-value) were extracted using the bedtools utility getfasta and submitted to meme-chip using default parameters (Machanick and Bailey 2011).

Figures

Figure 3-1: RNA Expression of *mads69* Across Maize Tissues as Measured by RNA-seq

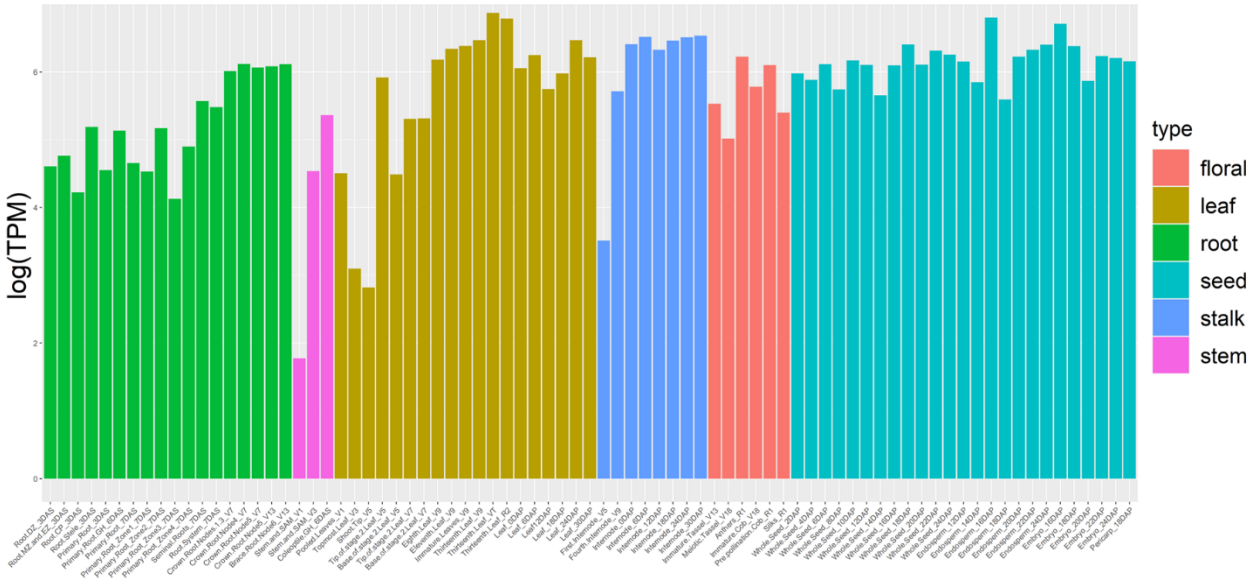

Figure 3-1. RNA expression of *mads69* across maize tissues as measured by RNA-seq. RNA-seq data was taken from the Maize Gene Atlas and normalized to transcripts per million (TPM). The major tissue groups are floral (coral), leaf (orange), root (green), seed (light blue), stalk (dark blue), and stem (pink). The vertical axis is in terms of log(TPM).

Figure 3-2: Protein Domains and Structure of ZmMADS69

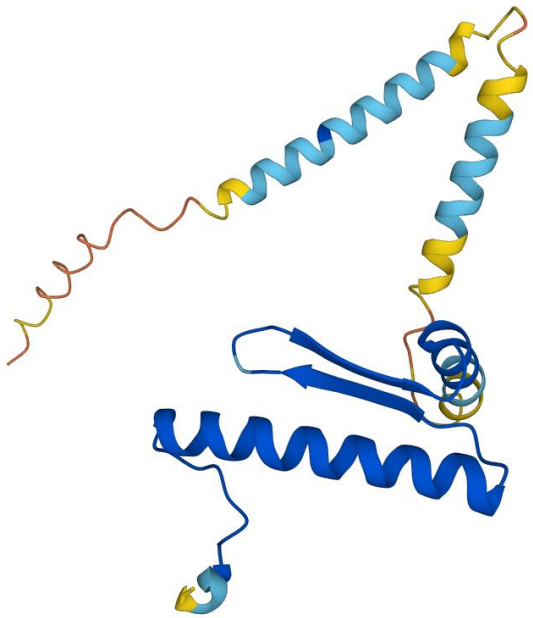

Figure 3-2. Above: Domains detected from InterPro domain database. Notable identified domains include the MEF2-like MADS-box, a coiled coil, and a disordered section. Left: Predicted folded structure of MADS69 from AlphaFold predictions accessed from the AlphaFold Protein Structure Database. Notable is the two beta strands, a conserved feature of MADS-box proteins and necessary for dimer formation.

Figure 3-3: Linear Map of MADS69 Expression Constructs

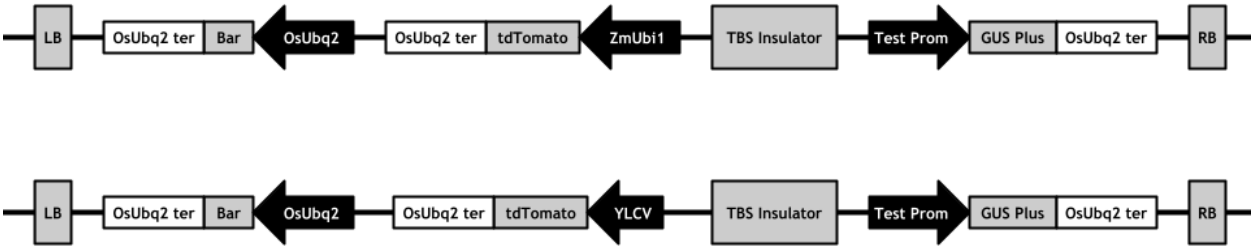

Figure 3-3. Linear map of MADS69 expression constructs. The Cestrum Yellow Leaf Curling Virus promoter (YLCV) promoter drives tdTomato expression. The bialaphos resistance gene (Bar) is driven by the Oryza sativa ubiquitin 2 promoter. Separating the expression unit from the rest of the construct is the Petunia hybrida transformation booster sequence (TBS) which insulates the promoter from nearby components. The tested promoter drives either the full length or shortened version of *mads69*.

Figure 3-4: Alignment of a ZmMADS69 Binding Site with the VGT Locus from DAP-seq Data

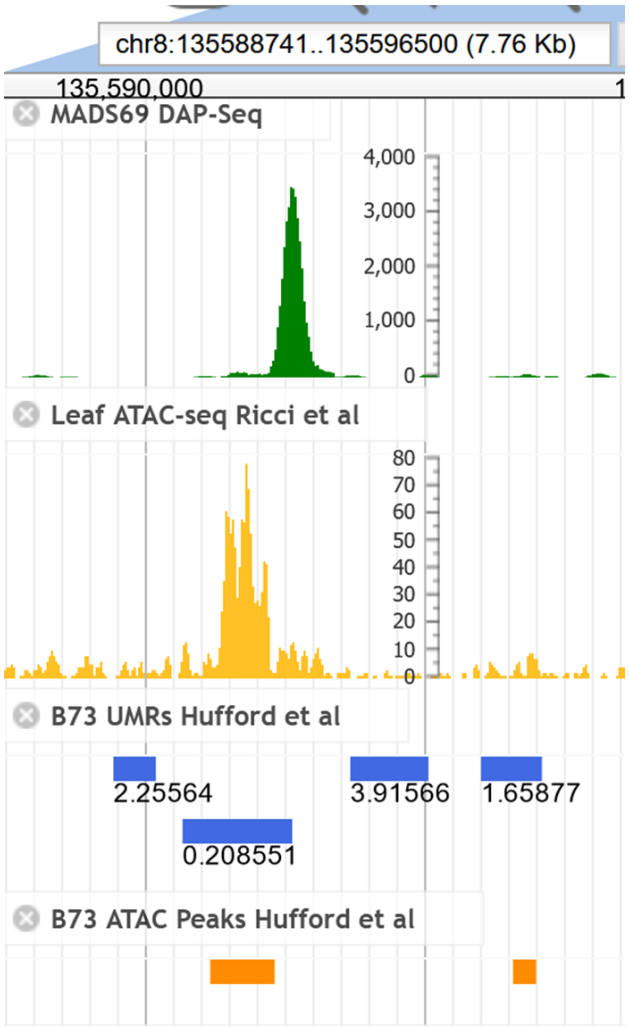

Figure 3-4. Alignment of the MADS69 binding site with the VGT locus from DAP-seq data. The binding site of MADS69 coincides with the identified VGT locus which controls *rap2.7*. This region also coincides with unmethylated regions and accessible chromatin.

Aknowledgement

We wish to acknowledge scientists and technical staff who assisted with this research, including R. Collier, N. Walter, and B. Martinell at the Wisconsin Crop Innovation Center and Mary Galli from the Gallavotti lab at Rutgers University.

References

Alvarez-Buylla ER, García-Ponce B, Sánchez M de la P, et al (2019) MADS-box genes underground becoming mainstream: plant root developmental mechanisms. New Phytol 223:1143–1158. https://doi.org/10.1111/NPH.15793

Bailey TL, Johnson J, Grant CE, Noble WS (2015) The MEME Suite. Nucleic Acids Res 43:W39–W49. https://doi.org/10.1093/nar/gkv416

Bartlett A, O'Malley RC, Huang SSC, et al (2017) Mapping genome-wide transcription-factor binding sites using DAP-seq. Nat Protoc 2017 128 12:1659–1672. https://doi.org/10.1038/nprot.2017.055

Bemer M, van Dijk ADJ, Immink RGH, Angenent GC (2017) Cross-Family Transcription Factor Interactions: An Additional Layer of Gene Regulation. Trends Plant Sci 22:66–80. https://doi.org/10.1016/J.TPLANTS.2016.10.007

Bezrutczyk M, Zöllner NR, Kruse CPS, et al (2021) Evidence for phloem loading via the abaxial bundle sheath cells in maize leaves. Plant Cell 33:531. https://doi.org/10.1093/PLCELL/KOAA055

Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120. https://doi.org/10.1093/BIOINFORMATICS/BTU170

Bucksch A, Burridge J, York LM, et al (2014a) Image-Based High-Throughput Field Phenotyping of Crop Roots. Plant Physiol 166:470–486. https://doi.org/10.1104/PP.114.243519

Bucksch A, Burridge J, York LM, et al (2014b) Image-Based High-Throughput Field Phenotyping of Crop Roots. Plant Physiol 166:470–486. https://doi.org/10.1104/PP.114.243519

Castelletti S, Tuberosa R, Pindo M, Salvi S (2014) A MITE Transposon Insertion Is Associated with Differential Methylation at the Maize Flowering Time QTL Vgt1 . G3: Genes|Genomes|Genetics 4:805–812. https://doi.org/10.1534/g3.114.010686

Chen Q, Samayoa LF, Yang CJ, et al (2020) The genetic architecture of the maize progenitor, teosinte, and how it was altered during maize domestication. PLOS Genet 16:e1008791. https://doi.org/10.1371/JOURNAL.PGEN.1008791

Cheng XF, Wang ZY (2005) Overexpression of COL9, a CONSTANS-LIKE gene, delays flowering by reducing expression of CO and FT in Arabidopsis thaliana. Plant J 43:758–768. https://doi.org/10.1111/J.1365-313X.2005.02491.X

Colombi T, Kirchgessner N, Le Marié CA, et al (2015) Next generation shovelomics: set up a tent and REST. Plant Soil 388:1–20. https://doi.org/10.1007/s11104-015-2379-7

Colombo M, Masiero S, Vanzulli S, et al (2008) AGL23, a type I MADS-box gene that controls female gametophyte and embryo development in Arabidopsis. Plant J 54:1037–1048. https://doi.org/10.1111/J.1365-313X.2008.03485.X

Das A, Schneider H, Burridge J, et al (2015a) Digital imaging of root traits (DIRT): A high-throughput computing and collaboration platform for field-based root phenomics. Plant Methods 11:1–12. https://doi.org/10.1186/S13007-015-0093-3/FIGURES/6

Das A, Schneider H, Burridge J, et al (2015b) Digital imaging of root traits (DIRT): A high-throughput computing and collaboration platform for field-based root phenomics. Plant Methods 11:1–12. https://doi.org/10.1186/S13007-015-0093-3/FIGURES/6

Ducrocq S, Giauffret C, Madur D, et al (2009) Fine mapping and haplotype structure analysis of a major flowering time quantitative trait locus on maize chromosome 10. Genetics 183:1555–1563. https://doi.org/10.1534/genetics.109.106922

Dwyer K, Agarwal N, Pile L, Ansari A (2021) Gene Architecture Facilitates Intron-Mediated Enhancement of Transcription. Front Mol Biosci 8:276. https://doi.org/10.3389/FMOLB.2021.669004/BIBTEX

Engler C, Youles M, Gruetzner R, et al (2014) A Golden Gate modular cloning toolbox for plants. ACS Synth Biol 3:839–843. https://doi.org/10.1021/sb4001504

Freeling M, Subramaniam S (2009) Conserved noncoding sequences (CNSs) in higher plants. Curr Opin Plant Biol 12:126–132. https://doi.org/10.1016/J.PBI.2009.01.005

Gallegos JE, Rose AB (2015) The enduring mystery of intron-mediated enhancement. Plant Sci 237:8–15. https://doi.org/10.1016/J.PLANTSCI.2015.04.017

Galli M, Khakhar A, Lu Z, et al (2018) The DNA binding landscape of the maize AUXIN RESPONSE FACTOR family. Nat Commun 2018 91 9:1–14. https://doi.org/10.1038/s41467-018-06977-6

Gallo M Del, De T, Santos O, et al (2023) Maize Breeding for Low Nitrogen Inputs in Agriculture: Mechanisms Underlying the Tolerance to the Abiotic Stress. Stress 2023, Vol 3, Pages 136-152 3:136–152. https://doi.org/10.3390/STRESSES3010011

Goujon M, McWilliam H, Li W, et al (2010) A new bioinformatics analysis tools framework at EMBL–EBI. Nucleic Acids Res 38:W695–W699. https://doi.org/10.1093/NAR/GKQ313

Gramzow L, Theissen G (2010) A hitchhiker's guide to the MADS world of plants. Genome Biol 11:1–11. https://doi.org/10.1186/GB-2010-11-6-214/FIGURES/5

Guo Y, Mahony S, Gifford DK (2012) High Resolution Genome Wide Binding Event Finding and Motif Discovery Reveals Transcription Factor Spatial Binding Constraints. PLOS Comput Biol 8:e1002638. https://doi.org/10.1371/JOURNAL.PCBI.1002638

Herrera-Ubaldo H, Campos SE, López-Gómez P, et al (2023) The protein–protein interaction landscape of transcription factors during gynoecium development in Arabidopsis. Mol Plant 16:260–278. https://doi.org/10.1016/J.MOLP.2022.09.004

Hill CB, Li C (2016) Genetic Architecture of Flowering Phenology in Cereals and Opportunities for Crop Improvement. Front Plant Sci 7:1–23. https://doi.org/10.3389/fpls.2016.01906

Hily J-M, Singer SD, Yang Y, Liu Z (2009) A transformation booster sequence (TBS) from Petunia hybrida functions as an enhancer-blocking insulator in Arabidopsis thaliana. Plant Cell Rep 28:1095–104. https://doi.org/10.1007/s00299-009-0700-8

Honma T, Goto K (2001) Complexes of MADS-box proteins are sufficient to convert leaves into floral organs. Nat 2001 4096819 409:525–529. https://doi.org/10.1038/35054083

Hufford MB, Seetharam AS, Woodhouse MR, et al (2021) De novo assembly, annotation, and comparative analysis of 26 diverse maize genomes. Science 373:655–662. https://doi.org/10.1126/SCIENCE.ABG5289

Jamann TM, Sood S, Wisser RJ, Holland JB (2017) High-Throughput Resequencing of Maize Landraces at Genomic Regions Associated with Flowering Time. PLoS One 12:e0168910. https://doi.org/10.1371/JOURNAL.PONE.0168910

Jones P, Binns D, Chang HY, et al (2014) InterProScan 5: genome-scale protein function classification. Bioinformatics 30:1236–1240. https://doi.org/10.1093/BIOINFORMATICS/BTU031

Jumper J, Evans R, Pritzel A, et al (2021) Highly accurate protein structure prediction with AlphaFold. Nat 2021 5967873 596:583–589. https://doi.org/10.1038/s41586-021-03819-2

Kang M, Lee K, Finley T, et al (2022) An Improved Agrobacterium-Mediated Transformation and Genome-Editing Method for Maize Inbred B104 Using a Ternary Vector System and Immature Embryos. Front Plant Sci 13:860971. https://doi.org/10.3389/FPLS.2022.860971/BIBTEX

Kaufmann K, Melzer R, Theißen G (2005) MIKC-type MADS-domain proteins: structural modularity, protein interactions and network evolution in land plants. Gene 347:183–198. https://doi.org/10.1016/J.GENE.2004.12.014

Khanal R, Earl H, Lee EA, Lukens L (2011) The genetic architecture of flowering time and related traits in two early flowering maize lines. Crop Sci 51:146–156. https://doi.org/10.2135/CROPSCI2010.03.0177

Khangura RS, Venkata BP, Marla SR, et al (2020) Interaction Between Induced and Natural Variation at oil yellow1 Delays Reproductive Maturity in Maize. G3 Genes|Genomes|Genetics 10:797–810. https://doi.org/10.1534/G3.119.400838

Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. Nat Methods 2012 94 9:357–359. https://doi.org/10.1038/NMETH.1923

Laxa M (2017) Intron-mediated enhancement: A tool for heterologous gene expression in plants? Front Plant Sci 7:1977. https://doi.org/10.3389/FPLS.2016.01977/BIBTEX

Lazakis CM, Coneva V, Colasanti J (2011) ZCN8 encodes a potential orthologue of Arabidopsis FT florigen that integrates both endogenous and photoperiod flowering signals in maize. J Exp Bot 62:4833–4842. https://doi.org/10.1093/JXB/ERR129

Lazo GR, Stein PA, Ludwig RA (1991) A DNA Transformation–Competent Arabidopsis Genomic Library in Agrobacterium. Bio/Technology 1991 910 9:963–967. https://doi.org/10.1038/NBT1091-963

Leng P;, Khan SU;, Zhang D;, et al (2022) Linkage Mapping Reveals QTL for Flowering Time-Related Traits under Multiple Abiotic Stress Conditions in Maize. Int J Mol Sci 2022, Vol 23, Page 8410 23:8410. https://doi.org/10.3390/IJMS23158410

Li D, Wang X, Zhang X, et al (2015) The genetic architecture of leaf number and its genetic relationship to flowering time in maize. New Phytol 210:256–268. https://doi.org/10.1111/nph.13765

Li J, Chen F, Li Y, et al (2019) ZmRAP2.7, an AP2 transcription factor, is involved in maize brace roots development. Front Plant Sci 10:820. https://doi.org/10.3389/FPLS.2019.00820/BIBTEX

Li Y, Xu M (2017) CCT family genes in cereal crops: A current overview. Crop J 5:449–458. https://doi.org/10.1016/J.CJ.2017.07.001

Liang Y, Liu Q, Wang X, et al (2019) ZmMADS69 functions as a flowering activator through the ZmRap2.7-ZCN8 regulatory module and contributes to maize flowering time adaptation. New Phytol 221:2335–2347. https://doi.org/10.1111/nph.15512

Liu H, Zhou X, Li Q, et al (2020) CCT domain-containing genes in cereal crops: flowering time and beyond. Theor Appl Genet 2020 1335 133:1385–1396. https://doi.org/10.1007/S00122-020-03554-8

Machanick P, Bailey TL (2011) MEME-ChIP: motif analysis of large DNA datasets. Bioinformatics 27:1696–1697. https://doi.org/10.1093/BIOINFORMATICS/BTR189

Marand AP, Eveland AL, Kaufmann K, Springer NM (2022) cis-Regulatory Elements in Plant Development, Adaptation, and Evolution. Annu Rev Plant Biol 74:. https://doi.org/10.1146/ANNUREV-ARPLANT-070122-030236

Mascheretti I, Turner K, Brivio RS, et al (2015) Florigen-Encoding Genes of Day-Neutral and Photoperiod-Sensitive Maize Are Regulated by Different Chromatin Modifications at the Floral Transition. Plant Physiol 168:1351–1363. https://doi.org/10.1104/PP.15.00535

Masiero S, Colombo L, Grini PE, et al (2011) The Emerging Importance of Type I MADS Box Transcription Factors for Plant Reproduction. Plant Cell 23:865–872. https://doi.org/10.1105/TPC.110.081737

Mathieu J, Warthmann N, Küttner F, Schmid M (2007) Export of FT protein from phloem companion cells is sufficient for floral induction in Arabidopsis. Curr Biol 17:1055–1060. https://doi.org/10.1016/j.cub.2007.05.009

Mazaheri M, Heckwolf M, Vaillancourt B, et al (2019) Genome-wide association analysis of stalk biomass and anatomical traits in maize. BMC Plant Biol 19:1–17. https://doi.org/10.1186/s12870-019-1653-x

Meng X, Muszynski MG, Danilevskaya ON (2011) The FT-Like ZCN8 Gene Functions as a Floral Activator and Is Involved in Photoperiod Sensitivity in Maize. Plant Cell 23:942–960. https://doi.org/10.1105/tpc.110.081406

Michel KJ, Lima DC, Hundley H, et al (2022) Genetic mapping and prediction of flowering time and plant height in a maize Stiff Stalk MAGIC population. Genetics 221:. https://doi.org/10.1093/GENETICS/IYAC063

Miller TA, Muslin EH, Dorweiler JE (2008) A maize CONSTANS-like gene, conz1, exhibits distinct diurnal expression patterns in varied photoperiods. Planta 227:1377–1388. https://doi.org/10.1007/s00425-008-0709-1

Minow MAA, Ávila LM, Turner K, et al (2018) Distinct gene networks modulate floral induction of autonomous maize and photoperiod-dependent teosinte. J Exp Bot 69:2937–2952. https://doi.org/10.1093/jxb/ery110

O'Malley RC, Huang S-SC, Song L, et al (2016) Cistrome and Epicistrome Features Shape the Regulatory DNA Landscape. Cell 165:1280–1292. https://doi.org/10.1016/j.cell.2016.04.038

Paysan-Lafosse T, Blum M, Chuguransky S, et al (2023) InterPro in 2022. Nucleic Acids Res 51:D418–D427. https://doi.org/10.1093/NAR/GKAC993

Peiffer JA, Romay MC, Gore MA, et al (2014) The genetic architecture of maize height. Genetics 196:1337–56. https://doi.org/10.1534/genetics.113.159152

Pellegrini L, Tan S, Richmond TJ (1995) Structure of serum response factor core bound to DNA. Nat 1995 3766540 376:490–498. https://doi.org/10.1038/376490a0

Poverennaya I V., Roytberg MA (2020) Spliceosomal Introns: Features, Functions, and Evolution. Biochem 2020 857 85:725–734. https://doi.org/10.1134/S0006297920070019

Ricci WA, Lu Z, Ji L, et al (2019) Widespread long-range cis-regulatory elements in the maize genome. Nat Plants 5:1237–1249. https://doi.org/10.1038/s41477-019-0547-0

Salvi S, Sponza G, Morgante M, et al (2007) Conserved noncoding genomic sequences associated with a flowering-time quantitative trait locus in maize. Proc Natl Acad Sci 104:11376–11381. https://doi.org/10.1073/pnas.0704145104

Salvi S, Tuberosa R, Chiapparino E, et al (2002) Toward positional cloning of Vgt1, a QTL controlling the transition from the vegetative to the reproductive phase in maize. Plant Mol Biol 48:601–613. https://doi.org/10.1023/A:1014838024509

Schilling S, Pan S, Kennedy A, Melzer R (2018) MADS-box genes and crop domestication: the jack of all traits. J Exp Bot 69:1447–1469. https://doi.org/10.1093/JXB/ERX479

Schnable PS, Ware D, Fulton RS, et al (2009) The B73 maize genome: Complexity, diversity, and dynamics. Science (80-) 326:1112–1115. https://doi.org/10.1126/SCIENCE.1178534/SUPPL_FILE/SCHNABLE.SOM.PDF

Sekhon RS, Briskine R, Hirsch CN, et al (2013) Maize Gene Atlas Developed by RNA Sequencing and Comparative Evaluation of Transcriptomes Based on RNA Sequencing and Microarrays. PLoS One 8:. https://doi.org/10.1371/journal.pone.0061005

Sekhon RS, Saski C, Kumar R, et al (2019) Integrated genome-scale analysis identifies novel genes and networks underlying senescence in maize. Plant Cell 31:1968–1989. https://doi.org/10.1105/TPC.18.00930

Sha X, Guan H, Zhou Y, et al (2023) Genetic dissection of crown root traits and their relationship to aboveground agronomic traits in maize. J Integr Agric. https://doi.org/10.1016/J.JIA.2023.04.022

Shao Y, Zhou HZ, Wu Y, et al (2019) OsSPL3, an SBP-Domain Protein, Regulates Crown Root Development in Rice. Plant Cell 31:1257–1275. https://doi.org/10.1105/TPC.19.00038

Shaul O (2017) How introns enhance gene expression. Int J Biochem Cell Biol 91:145–155. https://doi.org/10.1016/J.BIOCEL.2017.06.016

Sievers F, Wilm A, Dineen D, et al (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol 7:539. https://doi.org/10.1038/MSB.2011.75

Song G qing, Han X, Ryner JT, et al (2021) Utilizing MIKC-type MADS-box protein SOC1 for yield potential enhancement in maize. Plant Cell Rep 40:1679–1693. https://doi.org/10.1007/S00299-021-02722-4/FIGURES/5

Song K, Kim HC, Shin S, et al (2017) Transcriptome analysis of flowering time genes under drought stress in maize leaves. Front Plant Sci 8:267. https://doi.org/10.3389/FPLS.2017.00267/BIBTEX

Song LK, Lee S, Hyo JK, et al (2007) OsMADS51 Is a Short-Day Flowering Promoter That Functions Upstream of Ehd1, OsMADS14, and Hd3a. Plant Physiol 145:1484. https://doi.org/10.1104/PP.107.103291

Stavolone L, Kononova M, Pauli S, et al (2003) Cestrum yellow leaf curling virus (CmYLCV) promoter: A new strong constitutive promoter for heterologous gene expression in a wide variety of crops. Plant Mol Biol 53:663–673. https://doi.org/10.1023/b:plan.0000019110.95420.bb

Stelpflug SC, Sekhon RS, Vaillancourt B, et al (2016) An Expanded Maize Gene Expression Atlas based on RNA Sequencing and its Use to Explore Root Development. Plant Genome 9:0. https://doi.org/10.3835/plantgenome2015.04.0025

Stephenson E, Estrada S, Meng X, et al (2019a) Over-expression of the photoperiod response regulator ZmCCT10 modifies plant architecture, flowering time and inflorescence morphology in maize. PLoS One 14:e0203728. https://doi.org/10.1371/JOURNAL.PONE.0203728

Stephenson E, Estrada S, Meng X, et al (2019b) Over-expression of the photoperiod response regulator ZmCCT10 modifies plant architecture, flowering time and inflorescence morphology in maize

Sun C, Fang J, Zhao T, et al (2012) The Histone Methyltransferase SDG724 Mediates H3K36me2/3 Deposition at MADS50 and RFT1 and Promotes Flowering in Rice. Plant Cell 24:3235–3247. https://doi.org/10.1105/TPC.112.101436

Tamaki S, Matsuo S, Hann LW, et al (2007) Hd3a protein is a mobile flowering signal in rice. Science (80-) 316:1033–1036. https://doi.org/10.1126/SCIENCE.1141753

Tao S, Liu P, Shi Y, et al (2022) Single-cell transcriptome and network analyses unveil key transcription factors regulating mesophyll cell development in maize. Genes (Basel) 13:374. https://doi.org/10.3390/GENES13020374/S1

Theißen G, Melzer R, Ruümpler F (2016) MADS-domain transcription factors and the floral quartet model of flower development: linking plant development and evolution. Development 143:3259–3271. https://doi.org/10.1242/DEV.134080

Thompson CJ, Movva NR, Tizard R, et al (1987) Characterization of the herbicide-resistance gene bar from Streptomyces hygroscopicus . EMBO J 6:2519–2523. https://doi.org/10.1002/j.1460-2075.1987.tb02538.x

Trachsel S, Kaeppler SM, Brown KM, Lynch JP (2011) Shovelomics: High throughput phenotyping of maize (Zea mays L.) root architecture in the field. Plant Soil 341:75–87. https://doi.org/10.1007/s11104-010-0623-8

Tripathi P, Carvallo M, Hamilton EE, et al (2017) Arabidopsis B-BOX32 interacts with CONSTANS-LIKE3 to regulate flowering. Proc Natl Acad Sci U S A 114:172–177. https://doi.org/10.1073/PNAS.1616459114/SUPPL_FILE/PNAS.1616459114.SD01.XLSX

Varadi M, Anyango S, Deshpande M, et al (2022) AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. Nucleic Acids Res 50:D439–D444. https://doi.org/10.1093/NAR/GKAB1061

Wang J, Jiang J, Oard JH (2000) Structure, expression and promoter activity of two polyubiquitin genes from rice (Oryza sativa L.). Plant Sci 156:201–211. https://doi.org/10.1016/S0168-9452(00)00255-7

Wang L, Beissinger TM, Lorant A, et al (2017) The interplay of demography and selection during maize domestication and expansion. Genome Biol 18:1–13. https://doi.org/10.1186/S13059-017-1346-4/FIGURES/3

Wu J, Lawit SJ, Weers B, et al (2019) Overexpression of zmm28 increases maize grain yield in the field. Proc Natl Acad Sci U S A 116:23850–23858. https://doi.org/10.1073/pnas.1902593116

Xu X, Li W, Yang S, et al (2023) Identification, evolution, expression and protein interaction analysis of genes encoding B-box zinc-finger proteins in maize. J Integr Agric 22:371–388. https://doi.org/10.1016/J.JIA.2022.08.091

Yang Y, Fanning L, Jack T (2003) The K domain mediates heterodimerization of the Arabidopsis floral organ identity proteins, APETALA3 and PISTILLATA. Plant J 33:47–59. https://doi.org/10.1046/J.0960-7412.2003.01473.X

Yu G, Wang LG, He QY (2015) ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. Bioinformatics 31:2382–2383. https://doi.org/10.1093/BIOINFORMATICS/BTV145

Yu LH, Miao ZQ, Qi GF, et al (2014) MADS-Box Transcription Factor AGL21 Regulates Lateral Root Development and Responds to Multiple External and Physiological Signals. Mol Plant 7:1653–1669. https://doi.org/10.1093/MP/SSU088

Yuan J, Long H, Qiu F, et al (2023) MADS-box protein MtSOC1c regulates flowering and seed development in Medicago truncatula. Ind Crops Prod 193:116125. https://doi.org/10.1016/J.INDCROP.2022.116125

Zhao D, Chen Z, Xu L, et al (2021) Genome-wide analysis of the mads-box gene family in maize: Gene structure, evolution, and relationships. Genes (Basel) 12:1956. https://doi.org/10.3390/GENES12121956/S1

Zheng Z, Hey S, Jubery T, et al (2020) Shared genetic control of root system architecture between Zea mays and Sorghum bicolor. Plant Physiol 182:977–991. https://doi.org/10.1104/PP.19.00752

Zobell O, Faigl W, Saedler H, Münster T (2010) MIKC* MADS-Box Proteins: Conserved Regulators of the Gametophytic Generation of Land Plants. Mol Biol Evol 27:1201–1211. https://doi.org/10.1093/MOLBEV/MSQ005

APPENDIX

Sequences

Aa2m Promoter Sequence:
CGCGAAGTCTGACTGGCCAGCCCAAACTCAGCCTTTGTAGTAGAGAGAGCCGACCT
TGGCGGCCTGCACGACATTTTTGAACTTTTTTTATCGGTTTGGACTTCATTACTTTAT
ATTCTTTGTCAACCAAATACTCCCTTCCTCTTAAATTATAATTCGTTTGATTTTTTTTA
TGTAAAGTTTTATGTATAGTCCACCTAGTTAACTTTAAAAGTTGGTTAGTCTAGATTT
TATGGAGGCCGACTACAATAGCAAATCTAACAATAAAAAAGTTAGCTAGATCAGAT
TTCAAGAACAGTCTGACCGATTTTGAAGATTTGATTCAGGTTAATTCTACGATTTTGA
AGATCCGACCTAACCTAATTTTACGTTCAGACTAAAGACATGTTTGTTTAAGATTAT
GATATGTCTAGTTATATAATTTAACTTATTTTAAACTAACACTTAATTTAAAATAAGT
TAGATTATATGATCAGAATATTATATTTATATAATTCCAAACAAATAACTCTAATAT
CGATTCTGGTTTATATTGTATAATTTTTATGTTGAATATTACGGCGCGCCAACAAATT
GCGATCGTTATGTAATCATGTATACGACCGACTGTACTATTCATAAAATATAATTCA
AAATAGGAGATGCTCGAGATAGCCTGGCAAGGGAGGAGGTATCGGAGGCGGCGTC
GCCTGCCGCGAGCGCACCATGCGTCAACGCTCAGGACCTCTCGCAGGCGCGCTGTTG
GATCGTGAGCAAGCACGCGTGGCAAATGGCATGCATTCCTCTCGTGCAAGTGCCGA
GCTGAGGGTCCCGACGCGGGGGGCAGGACGTGGCAGAGCGCCGCCCAGGCGGACG
CCAAGTGCCGAGCTACCAGGGCACCTTCTTTATTCCGTCCTTGCTCAGTCACACCTC
GCTCTCGCTCACTCTCGCCGTCCGCACAGCCGCTCATCGTCACCCACTGCCTGCCCTC
TCCCTGCGCCGGCCGGTGCCCCGACGCGC

Expb14 Promoter Sequence:
CGGCAAAGAAGTCTTTGCCGATGTACTGTTTGCCGAGCCTTCTTTGCCGAGTGTTAC
ACTCGGCAAAGCCTTTGCCGAGTGTTTTTAAGGCTTCGCCGAGTGCTTGCGGCACTC
GGCGAAGAGGTTGATTCCGGTAGTGATGCTGACACACACACACTACTGCACTATTAT
ATTATAAATAAATAAAAAGATATTGCATTCCCATCAGACGAGACTCTTATATATTTG
CGCACCTATTGTTGCTACAAATTAAGGTTAGCAAGAATAGAATCCATGCATATCAGG
CATGCAAGAAAATTACATGAATAGTTAGGATAAATTTAATATTTTAAAATAGATATA
TATGAAATTTAATGTTAATATTTTCTTACGTGTTATCAAGCACATTGTTATCGACAAG
AATAATTTTTTGCATAACTGTTTTTACATTATGTGCTCCAAAAAATTAAATGAAAATA
TCCGTATCTGGTTTTTATATCTATCATTTAAGAAGATATACGTTAGATTTGAAGTTTA
TATTTTATGAATCTTAACAAGCTCAATACTAAAAACAAGAATATGAATTTGCACATA
TCTTATTTAATTATTCACGGTCAAAGGAAAAGAGAACAAAATAACAGGCACGCGGA
TGAGTATTTGTTTTCATCCCTAACCAAGAGGATGGATTAGTTAGAGAAATATTTATG
GACGTGTAAGGAAAGAAAAAAGATTAACAACCTAACATGTGTAAGTAAAGGAAAT
ATATTCATATAGAAGAAAAGGAAAATTCAATCAGCTGTGGAAGAATTAATTAGAAA
CCAGGATTAATGAAACTCGAGGAACGTCAATATAATGATGTCCTTCTGTAAATTTCT
GTGCTATGGAAGGTGCATAGGCCGGCTATAAATAGATGCCATGACACCATGCACCA
CACCCCGCATCTCCTAGCCCCAGCTAGATCTTTGGACACATATCGAAACAGCTCTAG
CTAGCGCGCTCGTCGTCTAACCTGGCTTCCGG

Lhcb10 Promoter Sequence:

AAGTCGGTTCCCTTTTCTGCGAACATGCCGAAGCTTTCCTGCTTCACAGCTTCGGCGC
TGTGTCAACCTTTGTAGCTTTTCCCTTTCTGATACGAGTCCGAAGATACCTGCTCACA
CATTTTAAGCCGCTTCTCCCTTGTAGACCTTCGGCGACGAAGCATAGACCCAACAGA
TACCTGTTCACACATTATACTCCAGAAATACTGTTAAATCCTGTTTTTGAGGACCTTC
GGATGCCGAAGGTCCCCAACAATCGGTATTCTAAAATATCAGAAAATAAGTTTGTTA
TAGAAAATACATCGACTATGATCGAAATCTAAAATATGATTGATAAAAATCGTAAC
ATGTACCCTAAATTGATTTAAAAAGTCACATAAGTTAGTAATAATATAAATATTCAC
AAATTTATGAAGGATTAAAGCATATCTCATAAAAGCAATAAAACAATATCTCACAA
AATACAAGTGGCAAACATTATACAAACATACACATAGTCAGAAAGTCACAACTCAG
GACCTTAAAAAATGAAACTATCCGATTGAAAATACATTGATAACAATTGAACACTA
GAAAATAATATCACAAATCAAACTATGGAGCATATAACTAGCCATATAACTCTTATA
ATACAATAATAAAATCATCATATATTTAAATAAAACACTAGCAAGTCTAATAACATA
TGACTATAGAATCAAGATGTGTATGATGACATGACACTTGCAATTTTATCATCTCCT
ACTACTCGACATAGTCAATATAATTGATGTCCTCCTTATCTTTAAAGTTTCCATGCGA
ATTATAAATATATGTATGAAGAGTAATGATTGATAAGAAACTATAAATAAGAGTCA
CAATAGTTCAAACAACTCTAAACTATATATCATTAGATAGATCTTGATTTTAGAAAA
ATAACGAAATCAGTTTCATAATTTTCTAAGTTAAGATGAATTTACAAAGATTAGTTT
AGATTTAATATTTTTTCTGAAAAAATACCGATTTCGGAAACGGGCAAAAGAGATCCA
AACTATTTCTGTTTTTTTTTACCGATTTCATTTCCGTATTTTCGGTAACGGTTTCCGGT
TTCGTATGACCCTAAATTTTGGTAAAGTTTCGAAAAAAAATATTTTAAGAACTGAAA
ATTAACGTTCCTGTTTTCATCCATACTAATGGCTCTTTACCGCTAAAATGTTGCCCAC
AATCATTGAGTAGGTTTAGACGTGAGAGCAAACAGTACAACATTACGATTCGCCCTT
GCCCAAATTTACATGCCTTTTCCCTACGGAAACAACATAGAATCAAGTTGACGGGGT
TACTTACATTGAAGTGGCCAAACTGATGGTAGCTGTAGATTTGGATGTATGTTTTCT
ATAAATTAGTCAAAATTGAGACAAAATAAACTGCAATTTAAAACTGAGGAAATAGT
AAAAAAAAGGTGAAGAAGGGAGGAAGAGGAAATCAGAAGCAAAAAATGGGCAAC
TTTAGGCCCATTATCTCGATGGTCACGTCGGAGTCCAGATATGTGATTGACGGATTG
GATTGGGCCGTACATCTTGCATGAGAGTTCGCCAAGATTTCATTGTTTAACAAGAAG
CGCGTGACAACAAAACCAAGCCTATCTCATCCACTCTTTTTTTCCCTTCCCACAATGG
CAAGTGGCAGCTCCTGATTCGCTCTGGCCATTCCTACGTGGCACACACCAGGATTCT
TGTGTGATAGGCCACTGGGTCCCACCCACCAGGTGCCACATCAGACGCCAAGCCAT
CCCGGCAGAACCAATCCCAGCCCAGCAACAGATGGTCTGCTATCCAGTTCCAACTGT
ATAAAAGCAGCTGCTGTGTTCTGTTAATGGCACAGCCATCACACGCACGCATACACA
GCACAGAGTGAGGTAAGCATCCGAAAAAAGCTGTGATCTGATCGAA

Nas2 Promoter Sequence:
GTGCTATGGTTGTGGACCTGCCTTCTGCTACGATGAAGGACACTGATAAAATTAGAA
GAGGGTTCCTTTGATATGGTTGTAAAAATGCCATGGGTGGACATTGTGCAGTGGCAT
GGGGAAGAGTGTGCAGACTCACGAGCTTGGTGGGCTGGGTATTTTCAGCATCACTG
AGCTGAGTTGGGCTCTTAAAATCCGATGGGCTTGGCTGCAAAAAACTGAGCCATGTC
AGCCCTGGTCATCCCTGCCCCTTCATATCCTGAGGAAAATCCGAGAGTTTTTACAAA
TGGCCATGTATTCTGAGATTGGAAATGGAGCCTCTACCCTCTTCTGGAGTGATCGAT
GACTGGGTGGGCAGCGGATAGTGGATATTGCTCCTCGTCTTTTAGGGACCATTCCTA
AGAAGCTGATTAATAAAAGGACAGTTCAGAAGGCGATCGGAGATAATTTATGGATT
AATGATATTCCTAGGAGTTTATCTGTTGGAGCACTTGCAGACTTTCTGCGCCTTTGGG
ATTTGGTGGCCCAAGTTGATCTTTGCCCTGATAAATAGGACAAGCATATTTTTCGGTT

GGCTGCTAATGGCAAATATCCGGCTAAAGCTGCTTATGAGGGGCTCTTCTTGGGTTT
AGTGGAGTTCGAACCTTATGAGAGGATTTGGAAAACTTGGGCTCCGCCCAAATGCC
AATTCTTCTTGTGGTTGGTGGCTCATAAGAAATGTTGGACGACAGATATACTTGAAA
AGCGAGGTTTGAATCATCTGAAAAGATGTCCTCTTTGTGATCAGAAAAGGGAGACA
ATAGACCATCTGCTTGTGACTTGTGTCTTTTCTAGAGAATGTTGGTTCCTCCTTCTTA
GGCAATTCAGACTCCATGGTCCGCAACCATCGAATATTAATATCATGGAGTGGTGGC
ATCAAGCAAATAAAGCTATTTAGGGGTCCTTCAGAGATGGCCTAAACTCTCTCATTA
TGCTAGGTGCTTGGGTCTTATGGAATCACCAAAACAGATGCATCTTTGATAGTCTTT
CCCCTAATATTGCTAATTTCTTAATTCAAATTGGGGACCATCACTGTCTGTGGGAGA
CTGCTGGGGCCAAGGGGCTGTCCTCCCTTGCTGCCTTCCTCTCAGTTGTGGACTAAA
GCTTTATAGATGAGTCAGTTTATGTTCGTTACGTTTAGACAGCCGTAAGGCGTCAAT
TGTACCCTGGTCTATTTTGGACCTTTTTTCTTTCTATTTTTTATATAATGATATGTAAC
TCTCCTGCGTGTTCGAGAAAAAAAAGATAATAATGAGATTAAGGGAGAAGATCGGC
ATCTCAGGAGGAGATGGTGAGTTGCCTGACCAAGGTCATGAAGTGGACGTGGGCAC
GTGGCAAATCCACTCCAGGAACGGACCAAAAATAGGCTACCACTTTCAGGGCGCGA
CGTGGACGTATTGAAGATCTGGTGGTCACCGCGAAATGTACACATCAGAACGAGGC
ATGAAACTGGCCTTGGCCGTGGACGCGTGAAGCGCGCCATGCGCGTCCGACGTGGA
CGTAATGAGCTACTACTTCAGAGCCACATGTGTCCACGCTTCATCAGAAAGAAAACA
CCTGCACACGTCAACTGGAGCTGCCATCAGTTTAGTTTACATGCTGCATCGTCTGAA
GGCCAGATCAGTGATGAGCGAAACATTCTATATATACCCTCAGTCCCTGTTCCAGTT
CTCGTCAAGCTAGCAGCACGAGTTGTCGAACACTTGCCTGCTCTTGAGCTCAAGCTA
TCATCAGCTGCGTCTTGTGCACAGCAACAGCTTCCCAACCGCGCGCAACTGTACCAG
CGAAACTACCAACTGCCATA

Php20719a Promoter Sequence:
AGGCTATGTCCAGTAACGCTATCTCTATCATCTCTCCAATCATATCATTTATCCTATT
TTTTATTTTAAATTTATCATCTATATATTTTTACACGATCATATAAACATTATAATGA
AACGTATAGAAAAAAGAATTTTAGTCACTATATATAAAAACACAGATATGAAGATA
TTAACCTCATGTACTAGCAACTCCAAAAGGCTCTATATATTCTTCCTAATCTATATAA
ATATATTTTTTAAAATATATATATATATATCTTTCAATAACTTATCTAAATGACTATC
CATAAATATAGACATCTTATATTTTAGGTTTCTCGCTATTCAAATATAGAAAACATG
AATGATTCTTAGAATACAAACAAGATATAGACTGTTAGAGATTAAAAAGATGTAGA
AAATGGTTTTTATACAAATAATTATTTAAATAATGATTAAATGATAACTTAAAGAGT
GAAATTTAGAATGACTCAAGACGTTCTAAATTTAGGTTCTCTCCTCTCTGTATCGACG
AAATTTTGGCTTTTGTTCTTTCTCTCTACCTTGGCCGCGTTCGCACGTGCTCGCTGCCC
TAGATGATTTTTTTTTGTTTGTGACATGATTAAGGGCTTGAGGATTAGTATGGATTGG
AAATTATTTTGGTTTACTAAAAGTATCTTAAGAGACTCTTTATTTTTAACTCTCAATT
TAACTCTTTATTTGTTTATTTCTCTATAAATATTAAAATTTATATATAACATCTCATTA
CAACAGACTCTCTAACATACAAGTTAACTTGGCCAGCGAGAGTTGCTAGCTAAATTT
GGTTAGTGAGAAAGTTAATTTATAGAACCAGATAACTTAGAAAGTTAAATGACTAG
TCTGTTGAAGAGCTATTATGTTATTAAGTGGCTAAAATTTAGTTTGACAAACTATTTA
ACTAGTTTATTGGAGATACTCTAAAGATTTAAAGTCATCCGATCTCTAACGATCCCA
TGAATTAAGATTCTAAAGGGTCTAGGTGTCGATACCATCCGTATTCAGGCGTCCGTT
GAAAATAGTCTTACAGCTAGGGTTGGATATCGAGTAAATTTGACTCGCCTCGTCCGA
GTTTAAACTGGCTCGTTAAACTAACGAACCACATAGTCAGCTTGACTCGACTCATTT
ACTAGCTCGTTTAGGTTGCGATCCTAAGTAAAAAAACAATATGTATATAAAATAATC

CATTTCTAGTTAATTTTAACTAATCTAATAATATAAAAATTATGCTTATACCCATAGT
TTCACATACCACATCAATGTAACATTAAACTGATACAATTCATCACTTGTCAATTTAC
AATTTATAATTCACATATATATTCGTCAGTTTAATAATAGTTATATTCGTATAGACCA
ATTTAACATGTAGACATAGTTCATCAATAATTAGTTTAACTCATGAGTAATAAACAC
TACACATAAATATAACATATTCATCAATTATTACTATATCAATTATTACTATATATAT
ATATAGTTTTGTTTTGCTAAAATATAATACCTCCTTGTTTAGCTCACAAAATAGCTCC
TTACAAACCAAGCTTGAATATCAGTTTAACTTATAAAAAAATTCAACGAGTCAGAGT
TAAATTGAGCTCACGAGCATTTCGTTCTGTCTACTTCTACCATCAAGCACCAAAACC
AGCGGAGCGGGCCGGACGAGGGACGACAAGAAACCATTTCTGGTACGGCTCGCGAC
ACGTTACAGCCTAGCAGGGACGGCCGCCATGCACGCGCCCACGTCGCGTCCGCTGA
CGCGCCACGTCCCGCGCCACCGCCGAGCCGCTTTATATATCTGTACTACACCGCAGC
CGGCGCAGCGGAATACACGCCAGAAACTAATCCAGATAGCTAGCCTTTGATTTCTCA
CTTCAATCTGACGAAGTGCACACCTAGGCAACAAACTCGAACGCGCGCGGTAGTGG
AGCCGAA

Ris2 Promoter Sequence:
AAGGCGGACTATGGCATTTGGGGTCAGTTGATGCATGAATACCTGGTGCTTTTGCAA
TATCATTCCAATCAATTTATACATGGGCAGCTGAAGTCCAGCCTTGAAAAAGCTCCT
GTAGACGACAACTTTGTCTTTCTTGGGCTTCAGAGTTGTCTCTTCTCCAGGGAGTCAT
ACGTTGACCTTGTTGCTAAAGTAACCAAGCTTTTTCATCGTTTGCAAATCTTTCTCCT
TCAGTGTGTACCGGCCCATCTCAATGTGACTTGGTTTCGATGGCCTATCCTTCGGCAA
CTCCTCCACAGGCTATGCTAAGGACGAAGCCTTCATGGCTTCGGAATTCAGCAAAGT
ATCCTAAAGTGCCTCCGAAGAAGTCAGCGTGGTTAGTCTAGCAGCTGAGTGGATCCT
AGCCATCTAGACAATTTCTACCCGAAGGTGTTTTCCCTAAACTAGTCATCGAGGCAG
TGGTGGAGAAAAAGCTTCAGAGTCGCAGAAACTTTTGGCAGAAAGTGGGTGAGGGC
AGTAAATGTGACTCACTCACACCCGCTCCGCTGCTTTTATACTGAGAATTGGGTGCG
AAAGGGCGCGAGCAGCGAAAAAAAATCTGCACACGTGGACGTTGGCCTTCCCACCT
TCTCCCTGTCTGGTGGAACACGAGCTTGGCTTCATTTGTACCAAAGTATTCATTTAAC
CTTTTCCATAACGAGCTTTGGGAAGGTGTTCTTCGGACATTTAGCTCAAGGACTACT
AGATATATTTTTGTGTTTCGAGCTCGTTATGAAAAAATGAATTTATCCCGCAAGAGG
CTACTGTTGGGGGCCTAGTTGCCCGAAGATCCTCAAAATAGCTAACCTTCGACATAA
AAAGATGTATTTCAGGCACATAAGGAAGGGTGGCGAAGATGAAGTCTGAAGTCACC
AGCTTCGGCGTGCCAAAGGCTTCGTGTGCGGAAGCGACGAAGGAGAAAAGGTTGAA
GTCTTGTGCAGCAAGAAAGAGAGAAAAAACTTTGATGCCCCATGATATTTGTAAATC
ATAATTGTTAACTTTGTGGGCATACTTATAATTTCATACAAAGCTGTACCTCTCCCCT
ATAAATAGATGAACAGTGTCATGCATAAATGCTCCTTTTGGAGGGCTTCAAGCTTCA
TCTCTCTTACCCTTCGAAGTACATTCTAGTTGTTTTTTTGTGTAAAACCGAATGTATG
CTTGTACATTTAATTCATAAGAAGGGTGAAATAAGAGAATGCCAAGGCATACAAGG
TGACTAGCTTGCTTCGTCCTTAATCACACTTTTATACATCCTTACTTCGCCTGTGTCCT
TCGTTTATTCCTTCTCCCTTAGACGAAGAAAGTTTCAATCCTCTAACTTGTGTTGCCT
TGTTTTTGATATCATATAACAATTTAAAACAAGTGACCAACAATTATATATTCCATCC
TATCCAAAGTATAATTCGTTACACTAATTACACATTCTTTAAGTAATATATAAATTTA
GTTTGTATGTATGTCTATATTCATTATCATTTAAATGAGTGTGTACAGAAAAAATAA
GCTAGAAAGAACTATATTTTAGGAGGCAGGAAGTATATTTTTTTATAAGAACTGAAT
GAATAGTTTACCGATAAACACCGCGTCATAAGCAGAGCGCGGGCGCGTGCCACAGC
CACTGCCCGCAGCTTCGACATGTGCATGTGGTGGACGCAGTCCGGTGGGCGGTGGCT

GCGCATCAACCAATCACCGGCCGCCTTCCTCGCCACACGGGGCAGAGATATAGTTA
GGCTAGGCCCCGGGCGGCTCATCCGCAAGCCAGTGGACGCCAGTGAGTGAACACCA
GCCGTCTTGGCCCGCTCCTTCCCTGCTGCGGCA

Lhcb10minimal:
TAAAAAAAAGGTGAAGAAGGGAGGAAGAGGAAATCAGAAGCAAAAAATGGGCAA
CTTTAGGCCCATTATCTCGATGGTCACGTCGGAGTCCAGATATGTGATTGACGGATT
GGATTGGGCCGTACATCTTGCATGAGAGTTCGCCAAGATTTCATTGTTTAACAAGAA
GCGCGTGACAACAAAACCAAGCCTATCTCATCCACTCTTTTTTTCCCTTCCCACAATG
GCAAGTGGCAGCTCCTGATTCGCTCTGGCCATTCCTACGTGGCACACACCAGGATTC
TTGTGTGATAGGCCACTGGGTCCCACCCACCAGGTGCCACATCAGACGCCAAGCCAT
CCCGGCAGAACCAATCCCAGCCCAGCAACAGATGGTCTGCTATCCAGTTCCAACTGT
ATAAAAGCAGCTGCTGTGTTCTGTTAATGGCACAGCCATCACACGCACGCATACACA
GCACAGAGTGAGGTAAGCATCCGAAAAAAGCTGTGATCTGATCGA

Ris2minimal:
CACATTCTTTAAGTAATATATAAATTTAGTTTGTATGTATGTCTATATTCATTATCAT
TTAAATGAGTGTGTACAGAAAAAATAAGCTAGAAAGAACTATATTTTAGGAGGCAG
GAAGTATATTTTTTTATAAGAACTGAATGAATAGTTTACCGATAAACACCGCGTCAT
AAGCAGAGCGCGGGCGCGTGCCACAGCCACTGCCCGCAGCTTCGACATGTGCATGT
GGTGGACGCAGTCCGGTGGGCGGTGGCTGCGCATCAACCAATCACCGGCCGCCTTC
CTCGCCACACGGGGCAGAGATATAGTTAGGCTAGGCCCCGGGCGGCTCATCCGCAA
GCCAGTGGACGCCAGTGAGTGAACACCAGCCGTCTTGGCCCGCTCCTTCCCTGCTGC
GGCC

Snakemake RNA-seq Processing Pipeline

Pipeline Setup:
# Installing SRA toolkit
sudo bash setup-apt.sh
source /etc/profile.d/sra-tools.sh
Downloading data
prefetch SRR950078 -O /shared/data/seq/2021-05-11_ERCC_control
Monitor all jobs with "jobs -l"
nohup fastq-dump --split-files --gzip SRR950078.sra &
cp /mnt/c/Users/natha/Downloads/FastQC .
chmod 755 fastqc
nohup ./FastQC/fastqc SRR950078_*.fastq.gz &
cp /mnt/c/Users/natha/Downloads/Homo_sapiens.GRCh38.* .
nohup fastp -i SRR950078_1.fastq.gz -I SRR950078_2.fastq.gz -o trim_SRR950078_1.fastq.gz -O trim_SRR950078_2.fastq.gz &
Get latest STAR source from releases
wget https://github.com/alexdobin/STAR/archive/2.7.8a.tar.gz
tar -xzf 2.7.8a.tar.gz
cd STAR-2.7.8a
Compile

```
cd STAR/source
make STAR
nohup gunzip -k Homo_sapiens.GRCh38.* &
mkdir indexes
mkdir indexes/full
nohup STAR --runMode genomeGenerate --genomeDir
/shared/data/genomes/homo_sapiens/hg38/index/2021-05-11_star.v2.7.8a/full \
        --runThreadN 4 \
        --genomeFastaFiles /shared/data/genomes/homo_sapiens/hg38/fa/hg38.fa \
        --sjdbGTFfile /shared/data/genomes/homo_sapiens/hg38/annotation/hg38_ucsc_2021-05-
11.gtf \
        --sjdbOverhang 50 --outFileNamePrefix full &
mkdir alignments
$RUN STAR --genomeDir indexes/full \
     --readFilesIn trim_SRR950078_1.fastq.gz trim_SRR950078_2.fastq.gz \
     --readFilesCommand zcat \
     --outSAMtype BAM SortedByCoordinate \
     --quantMode GeneCounts \
     --outFileNamePrefix alignments/full
cp /mnt/c/Users/natha/Downloads/qualimap_v2.2.1 .
chmod 755 qualimap
./qualimap_v2.2.1/qualimap rnaseq -pe -bam alignments/fullAligned.sortedByCoord.out.bam \
        -gtf Homo_sapiens.GRCh38.96.chr.gtf \
        -outdir QC -p strand-specific-reverse

cluster.yaml:
__default__:
   jobname: 'snakemake'
   partition: "htc"
   cpus: 4
   mem: "16g"
   output: "results/slurm/{rule}.{wildcards}.out"
seqtk:
   jobname: "seqtk"
fastp_pe:
   jobname: "fastp"
   cpus: "{threads}"
   mem: "16g"
fastqc:
   jobname: "fastqc"
   cpus: 2
   mem: "6g"
star_pe_multi:
   jobname: "star_pe_multi"
   cpus: "{threads}"
   mem: "30g"
```

```
picard_collect_gc_bias_metrics:
   jobname: 'picardgcbias'
   cpus: 6
   mem: "24g"

config.yaml:
samples: resources/samples.tsv
ref:
   starindex: /data/nschleif/Indexes/B73
   ref: /data/nschleif/Genomes/Zm-B73-REFERENCE-NAM-5.0.fa
   refFlat: /data/nschleif/Annotation/Zm-B73-REFERENCE-NAM-5.0_Zm00001eb.1.flatref
   gtf: /data/nschleif/Annotation/Zm-B73-REFERENCE-NAM-5.0_Zm00001eb.1.gtf

multiqc_config.yaml:
read_count_multiplier: 1
read_count_desc: ''
decimalPoint_format: '.'
thousandsSep_format: ','
flash:
   use_output_name: true
extra_fn_clean_trim:
   - _L001_R1_001
   - _L001_R2_001
   - _L002_R1_001
   - _L002_R2_001
   - _L003_R1_001
   - _L003_R2_001
   - _L004_R1_001
   - _L004_R2_001
   - _L005_R1_001
   - _L005_R2_001
   - _L006_R1_001
   - _L006_R2_001
   - _L007_R1_001
   - _L007_R2_001
   - _L008_R1_001
   - _L008_R2_001
   - .subsampled
module_order:
   - general_stats
   - fastqc
   - samtools
   - picard
remove_sections:
   - fastqc_sequence_duplication_levels
   - fastqc_status_checks
```

```
    - fastqc_overrepresented_sequences
table_columns_visible:
    FastQC:
        total_sequences: False
        percent_duplicates: False
        percent_gc: False
    gatk3-depth-of-coverage:
        granular_first_quartile: False
        granular_third_quartile: False
        total: False
table_columns_placement:
    fastqc-custom-metrics:
        Total Read Pairs: 10
    FastQC:
        avg_sequence_length: 20
    Picard:
        PERCENT_DUPLICATION: 30
        summed_median: 40
    gatk3-depth-of-coverage:
        Median: 81
        Mean: 89
        Percent bases >= 1: 200
        Percent bases >= 10: 220
        Percent bases >= 20: 240
        Percent bases >= 30: 260
        Percent bases >= 40: 270
        Percent bases >= 50: 280
        Percent bases >= 75: 290
        Percent bases >= 100: 300
        Percent bases >= 250: 320
        Percent bases >= 500: 340
        Percent bases >= 1000: 360
        Percent bases >= 5000: 380

Snakemake:
import pandas as pd
import os

#### helper functions ####
def get_fastq(wildcards):
    """Get fastq files for given sample"""
    fastqs = samples.loc[(wildcards.sample), ["fq1", "fq2"]].dropna()
    if len(fastqs) == 2:
        return {"r1": fastqs.fq1, "r2": fastqs.fq2}
    return {"r1": fastqs.fq1}
```

```
def get_subsample_amount(wildcards):
    """Get the subsampling amount from the samples input file"""
    try:
        subsamp = unpack(samples.loc[(wildcards.sample), ["subsample"]].dropna())
    except KeyError:
        subsamp = 1
    return subsamp

#### setup config and sample sheet ####
configfile: "config/config.yaml"
samples = pd.read_table(config["samples"]).set_index("sample", drop=False)

rule all:
    input:
        expand("results/star/{sample}/{sample}.Aligned.out.sam",sample=samples.index),
        "results/SUMMARIZEDRESULTS/multiqc_report.html",
        expand("results/picard/{sample}.rna_metrics",sample=samples.index),
        expand("results/tpm_calculator/{sample}.Aligned.out_genes.out",sample=samples.index)

rule seqtk:
    input:
        unpack(get_fastq)
    output:
        r1="results/reads/{sample}_1.subsampled.fastq",
        r2="results/reads/{sample}_2.subsampled.fastq"
    params:
        n=get_subsample_amount,
        seed=12345
    log:
        r1="results/logs/{sample}.seqtk_r1.log",
        r2="results/logs/{sample}.seqtk_r2.log"
    conda:
         "envs/seqtk.yaml"
    shell:
        "(seqtk sample -s {params.seed} {input.r1} {params.n} > {output.r1} 2> {log.r1}) &&
(seqtk sample -s {params.seed} {input.r2} {params.n} > {output.r2} 2> {log.r2})"

rule fastp_pe:
    input:
        sample=[rules.seqtk.output.r1, rules.seqtk.output.r2]
    output:
        trimmed=["results/trimmed/{sample}_1.fastq", "results/trimmed/{sample}_2.fastq"],
        html="results/report/{sample}.fastp.html",
        json="results/report/{sample}.fastp.json"
    log:
        "results/logs/{sample}.fastp.log"
```

```
    # params:
    #     adapters="--adapter_sequence ACGGCTAGCTA --adapter_sequence_r2
AGATCGGAAGAGCACACGTCTGAACTCCAGTCAC",
    #     extra=""
    threads:
        6
    wrapper:
        "0.74.0/bio/fastp"

rule fastqc:
    input:
        "results/trimmed/{sample}_1.fastq",
        "results/trimmed/{sample}_2.fastq"
    output:
        "results/fastqc/{sample}_1_fastqc.html",
        "results/fastqc/{sample}_2_fastqc.html",
        "results/fastqc/{sample}_1_fastqc.zip", # the suffix _fastqc.zip is necessary for multiqc to
find the file. If not using multiqc, you are free to choose an arbitrary filename
        "results/fastqc/{sample}_2_fastqc.zip"
    params:
        "--nogroup -t 2"
    log:
        "results/logs/fastqc/{sample}.log"
    conda:
        "envs/fastqc.yaml"
    shell:
        "fastqc -o results/fastqc/ {params} {input}"

rule star_pe_multi:
    input:
        # use a list for multiple fastq files for one sample
        # usually technical replicates across lanes/flowcells
        fq1 = "results/trimmed/{sample}_1.fastq",
        # paired end reads needs to be ordered so each item in the two lists match
        fq2 = "results/trimmed/{sample}_2.fastq"
    output:
        "results/star/{sample}/{sample}.Aligned.out.sam",
        "results/star/{sample}/{sample}.Log.final.out"
    log:
        "results/logs/{sample}.star.log"
    params:
        # path to STAR reference genome index
        index=config["ref"]["starindex"],
        threads="8",
        prefix="results/star/{sample}/{sample}."
    conda:
```

```
        "envs/star.yaml"
    shell:
        "STAR --runThreadN {params.threads} --quantMode GeneCounts --genomeDir
{params.index} --readFilesIn {input.fq1} {input.fq2} --outFileNamePrefix {params.prefix} --
outStd Log {log}"

rule samtools_view:
    input:
        "results/star/{sample}/{sample}.Aligned.out.sam"
    output:
        "results/star/{sample}/{sample}.Aligned.out.bam"
    log:
        "results/logs/{sample}.samtoolsview.log"
    params:
        extra=""
    wrapper:
        "v0.75.0/bio/samtools/view"

rule samtools_sort:
    input:
        "results/star/{sample}/{sample}.Aligned.out.bam"
    output:
        "results/star/{sample}/{sample}.Aligned.out.sorted.bam"
    params:
        extra = "-m 3G",
        tmp_dir = "/tmp/"
    threads:  # Samtools takes additional threads through its option -@
        4     # This value - 1 will be sent to -@.
    wrapper:
        "0.77.0/bio/samtools/sort"

rule tpm_calculator:
    input:
        "results/star/{sample}/{sample}.Aligned.out.bam"
    output:
        "results/tpm_calculator/{sample}.Aligned.out_genes.out"
    params:
        # path to STAR reference genome index
        gtf=config["ref"]["gtf"]
    conda:
        "envs/tpmcalc.yaml"
    shell:
        "(cd results/tpm_calculator/; TPMCalculator -g {params.gtf} -b ../../{input})"

rule samtools_index:
    input:
```

```
        "results/star/{sample}/{sample}.Aligned.out.sorted.bam"
    output:
        "results/star/{sample}/{sample}.Aligned.out.bam.bai"
    params:
        "-@ 4" # threads
    wrapper:
        "0.49.0/bio/samtools/index"

rule idxstats:
    input:
        bam="results/star/{sample}/{sample}.Aligned.out.sorted.bam",
        idx="results/star/{sample}/{sample}.Aligned.out.bam.bai"
    output:
        "results/idxstats/{sample}.idxstats.txt"
    log:
        "results/logs/{sample}.idxstats.log"
    wrapper:
        "0.63.0/bio/samtools/idxstats"

rule picard_collect_gc_bias_metrics:
    input:
        bam=rules.samtools_sort.output,
        bai=rules.samtools_index.output
    output:
        metrics="results/picard/{sample}.gc_metrics.txt",
        chart="results/picard/{sample}.gc_chart.pdf",
        summary="results/picard/{sample}.gc_summary.txt"
    log:
        "results/logs/picard/{sample}.gc_metrics.log"
    params:
        f"R={config['ref']['ref']}"
    conda:
        "envs/picard.yaml"
    shell:
        "picard -Xmx6g -Xms4g CollectGcBiasMetrics INPUT={input.bam}
OUTPUT={output.metrics} CHART_OUTPUT={output.chart}
SUMMARY_OUTPUT={output.summary} {params} &> {log}"

rule insert_size:
    input:
        rules.samtools_sort.output
    output:
        txt="results/picard/{sample}.insert_size.txt",
        pdf="results/picard/{sample}.insert_size.pdf"
    log:
        "results/logs/picard/{sample}.insert_size.log"
```

```
    params:
        # optional parameters (e.g. relax checks as below)
        "VALIDATION_STRINGENCY=LENIENT "
    wrapper:
        "0.62.0/bio/picard/collectinsertsizemetrics"

rule picard_collect_rna_seq_metrics:
    input:
        bam=rules.samtools_sort.output
    output:
        rna_metrics="results/picard/{sample}.rna_metrics"
    log:
        "results/logs/picard/{sample}.rna_metrics.log"
    params:
        refFlat=config["ref"]["refFlat"]
    conda:
        "envs/picard.yaml"
    shell:
        "picard -Xmx6g -Xms4g CollectRnaSeqMetrics INPUT={input.bam}
O={output.rna_metrics} REF_FLAT={params.refFlat} STRAND=NONE &> {log}"

rule multiqc:
    input:
        expand("results/fastqc/{sample}_1_fastqc.html", sample=samples.index),     # fastqc
        expand("results/picard/{sample}.gc_metrics.txt", sample=samples.index),   #
picard_collect_gc_bias_metrics
        expand("results/picard/{sample}.gc_chart.pdf", sample=samples.index),     #
picard_collect_gc_bias_metrics
        expand("results/picard/{sample}.gc_summary.txt", sample=samples.index),   #
picard_collect_gc_bias_metrics
        expand("results/picard/{sample}.insert_size.txt", sample=samples.index),  # insert_size
        expand("results/idxstats/{sample}.idxstats.txt", sample=samples.index),   # samtools
idxstats
        expand("results/picard/{sample}.rna_metrics", sample=samples.index),   # picard
rna_seq_metrics
        expand("results/star/{sample}/{sample}.Log.final.out", sample=samples.index),   # star
alignment information
        expand("results/report/{sample}.fastp.json", sample=samples.index) # fastp results
    output:
        "results/SUMMARIZEDRESULTS/multiqc_report.html"
    params:
        "--config config/multiqc_config.yaml "
    log:
        "results/logs/multiqc.log"
    wrapper:
        "0.63.0/bio/multiqc"
```

Field and Greenhouse Statistical Analysis

```
setwd("G:/My Drive/Research/Flowering.Time.Project/data")
GDDdf <- read.csv("GDDplot.csv")
GDDdf$Date <- as.Date(GDDdf$Date,format="%m/%d/%Y")

# V3 vs V4
# ANOVA of the greenhouse traits for the V3 sequence
gh_mads69 <- read.table(file = "GH_mads69.csv", sep = ",", header = TRUE, fill = TRUE,
comment.char = "")
gh_1626_aov <- aov(Days ~ Transgenic +
Event,data=gh_mads69[gh_mads69$Construct==1626,])
gh_1626_aov <- aov(Number.of.Nodes ~ Transgenic +
Event,data=gh_mads69[gh_mads69$Construct==1626,])
gh_1626_aov <- aov(Leaf.Angle ~ Transgenic +
Event,data=gh_mads69[gh_mads69$Construct==1626,])
gh_1626_aov <- aov(Tassel.Angle ~ Transgenic +
Event,data=gh_mads69[gh_mads69$Construct==1626,])
summary(gh_1626_aov)
# Change variables to get different means
mean(gh_mads69[gh_mads69$Construct==1626 &
gh_mads69$Transgenic=="Positive",]$Leaf.Angle, na.rm=TRUE)
# Check out the
gh_1627_aov <- aov(Days ~ Transgenic,data=gh_mads69[gh_mads69$Construct==1627,])
gh_1626_aov <- aov(Number.of.Nodes ~ Transgenic +
Event,data=gh_mads69[gh_mads69$Construct==1627,])
summary(gh_1627_aov)
# Change variables to get different means
mean(gh_mads69[gh_mads69$Construct==1627 & gh_mads69$Transgenic=="Positive",]$Days,
na.rm=TRUE)

# Embryo Specific Promoter Flowering
# Initial analysis with only 1 event
em_root_df <- read.table(file = "em_root_GH_notes.csv", sep = ",", header = TRUE, fill =
TRUE )
em_flowering_aov <- aov(Days.Flowering ~
Transgenic,data=em_root_df[em_root_df$Construct==1955,])
mean(em_root_df[em_root_df$Transgenic == "Positive" & em_root_df$Construct==1955,
]$Days.Flowering, na.rm=TRUE)
mean(em_root_df[em_root_df$Transgenic == "Negative" & em_root_df$Construct==1955,
]$Days.Flowering, na.rm=TRUE)
em_flowering_aov <- aov(Days.Silking ~
Transgenic,data=em_root_df[em_root_df$Construct==1955,])
```

```
mean(em_root_df[em_root_df$Transgenic == "Positive" & em_root_df$Construct==1955,
]$Days.Silking, na.rm=TRUE)
mean(em_root_df[em_root_df$Transgenic == "Negative" & em_root_df$Construct==1955,
]$Days.Silking, na.rm=TRUE)
em_flowering_aov <- aov(Node.Num ~
Transgenic,data=em_root_df[em_root_df$Construct==1955,])
mean(em_root_df[em_root_df$Transgenic == "Positive" & em_root_df$Construct==1955,
]$Node.Num, na.rm=TRUE)
mean(em_root_df[em_root_df$Transgenic == "Negative" & em_root_df$Construct==1955,
]$Node.Num, na.rm=TRUE)
em_flowering_aov <- aov(Ear.Node ~
Transgenic,data=em_root_df[em_root_df$Construct==1955,])
summary(em_flowering_aov)
em_flowering_aov <- aov(Days.Flowering ~
Transgenic,data=em_root_df[em_root_df$Construct==1956,])
em_flowering_aov <- aov(Node.Num ~
Transgenic,data=em_root_df[em_root_df$Construct==1956,])
mean(em_root_df[em_root_df$Transgenic == "Positive" & em_root_df$Construct==1956,
]$Node.Num, na.rm=TRUE)
mean(em_root_df[em_root_df$Transgenic == "Negative" & em_root_df$Construct==1956,
]$Node.Num, na.rm=TRUE)
summary(em_flowering_aov)
# Second analysis with 3 events
emb_full_phenos <- read.table(file = "big_embryo_prom.csv", sep = ",", header = TRUE,
na.strings = "")
mean(emb_full_phenos[emb_full_phenos$Transgenic == "pos", ]$Days, na.rm=TRUE)
mean(emb_full_phenos[emb_full_phenos$Transgenic == "neg", ]$Days, na.rm=TRUE)
em_flowering_aov <- aov(Days ~ Transgenic + Event,data=emb_full_phenos)
summary(em_flowering_aov)

# Analysis of embryo-specific data
emb_phenos <- read.table(file = "emb_phenos.csv", sep = ",", header = TRUE, na.strings = "")
mean(emb_phenos[emb_phenos$transgenic == "positive", ]$leaf.number, na.rm=TRUE)
mean(emb_phenos[emb_phenos$transgenic == "negative", ]$leaf.number, na.rm=TRUE)
emb_aov <- aov(leaf.number ~ transgenic + event,data=emb_phenos)
summary(emb_aov)
emb_length_df <- read.table(file = "Embryo_Lengths.csv", sep = ",", header = TRUE, na.strings
= "")
emb_length_aov <- aov(embryo.length..mm. ~ transgenic + event,data=emb_length_df)
summary(emb_length_aov)

# Root analysis
root_fun <- function(x, GDD, colu) {
  beginDate <- GDD$Date >= as.Date("6/29/2022",format="%m/%d/%Y")
  endDate <- GDD$Date <= x[colu]
  return(sum(GDD[beginDate & endDate, ]$GDDDaily))
```

```
}
root_phenos <- read.table(file = "roots_image_data.csv", sep = ",", header = TRUE, fill = TRUE)
root_phenos$Block <- as.character(root_phenos$Block)
root_phenos$Harvest.Date <- as.Date(root_phenos$Harvest.Date,format="%m/%d/%Y")
root_phenos$harvestGDD <- apply(root_phenos, 1, root_fun, GDD = GDDdf, colu =
"Harvest.Date")
root_aov <- aov(AREA ~ Transgenic + Event + Block + (Transgenic *
Event),data=root_phenos)
root_aov <- aov(AREA ~ Transgenic + Cob + Block,data=root_phenos)
root_aov <- aov(AREA ~ Transgenic + harvestGDD + (Transgenic * harvestGDD) +
Block,data=root_phenos)
summary(root_aov)
mean(root_phenos[root_phenos$Transgenic == "Positive", ]$AREA)

# 2022 Field Ubi1 Promoter
ubi_fun <- function(x, GDD, colu) {
  beginDate <- GDD$Date >= as.Date("6/29/2022",format="%m/%d/%Y")
  endDate <- GDD$Date <= x[colu]
  return(sum(GDD[beginDate & endDate, ]$GDDDaily))
}
field_ubi_df <- read.table(file = "2022_field_ubi1_mads69.csv", sep = ",", header = TRUE, fill =
TRUE)
field_ubi_df$Plot <- as.character(field_ubi_df$Plot)
field_ubi_df$Pollen.Shed.Date <-
as.Date(field_ubi_df$Pollen.Shed.Date,format="%m/%d/%Y")
field_ubi_df$pollenGDD <- apply(field_ubi_df, 1, ubi_fun, GDD = GDDdf, colu =
"Pollen.Shed.Date")
field_ubi_df$Pollen.Days <- difftime(field_ubi_df$Pollen.Shed.Date
,as.Date("6/29/2022",format="%m/%d/%Y") , units = c("days"))
mean(field_ubi_df[field_ubi_df$Positive == "Y", ]$Pollen.Days, na.rm=TRUE)
mean(field_ubi_df[field_ubi_df$Positive == "N", ]$Pollen.Days, na.rm=TRUE)
ubi1_field_aov <- aov(Node.Number ~ Positive + Plot + Event, data=field_ubi_df)
ubi1_field_aov <- aov(pollenGDD ~ Positive + Plot + Event, data=field_ubi_df)
ubi1_field_aov <- aov(Leaf.Angle ~ Positive + Plot + Event, data=field_ubi_df)
summary(ubi1_field_aov)

# 2022 Field Leaf Promoter Data
leaf_fun <- function(x, GDD, colu) {
  beginDate <- GDD$Date >= as.Date("6/28/2022",format="%m/%d/%Y")
  endDate <- GDD$Date <= x[colu]
  return(sum(GDD[beginDate & endDate, ]$GDDDaily))
}
field_leafprom_df <- read.table(file = "2022_field_leaf_prom.csv", sep = ",", header = TRUE,
fill = TRUE)
field_leafprom_df$Plot <- as.character(field_leafprom_df$Plot)
```

```
field_leafprom_df$Pollen.Date <-
as.Date(field_leafprom_df$Pollen.Date,format="%m/%d/%Y")
field_leafprom_df$Silking.Date <-
as.Date(field_leafprom_df$Silking.Date,format="%m/%d/%Y")
field_leafprom_df$pollenGDD <- apply(field_leafprom_df, 1, leaf_fun, GDD = GDDdf, colu =
"Pollen.Date")
mean(field_leafprom_df[field_leafprom_df$Positive == "Y", ]$pollenGDD, na.rm=TRUE)
mean(field_leafprom_df[field_leafprom_df$Positive == "N", ]$pollenGDD, na.rm=TRUE)
field_leafprom_df$silkingGDD <- apply(field_leafprom_df, 1, leaf_fun, GDD = GDDdf, colu =
"Silking.Date")
mean(field_leafprom_df[field_leafprom_df$Positive == "Y", ]$silkingGDD, na.rm=TRUE)
mean(field_leafprom_df[field_leafprom_df$Positive == "N", ]$silkingGDD, na.rm=TRUE)
field_leafprom_df$Pollen.Days <- difftime(field_leafprom_df$Pollen.Date
,as.Date("6/28/2022",format="%m/%d/%Y") , units = c("days"))
mean(field_leafprom_df[field_leafprom_df$Positive == "Y", ]$Pollen.Days, na.rm=TRUE)
mean(field_leafprom_df[field_leafprom_df$Positive == "N", ]$Pollen.Days, na.rm=TRUE)
field_leafprom_df$Silk.Days <- difftime(field_leafprom_df$Silking.Date
,as.Date("6/28/2022",format="%m/%d/%Y") , units = c("days"))
mean(field_leafprom_df[field_leafprom_df$Positive == "Y", ]$Silk.Days, na.rm=TRUE)
mean(field_leafprom_df[field_leafprom_df$Positive == "N", ]$Silk.Days, na.rm=TRUE)
leafprom_field_aov <- aov(pollenGDD ~ Positive + Plot, data=field_leafprom_df)
eafprom_field_aov <- aov(silkingGDD ~ Positive + Plot, data=field_leafprom_df)
summary(leafprom_field_aov)

# 2022 GH Leaf Promoter
GH_leaf_df <- read.table(file = "GH_leaf_promoter.csv", sep = ",", header = TRUE, fill =
TRUE)
GH_leaf_df <- GH_leaf_df[GH_leaf_df$Event == "1713-2", ]
mean(GH_leaf_df[GH_leaf_df$Transgenic == "Positive", ]$Flowering.Days)
mean(GH_leaf_df[GH_leaf_df$Transgenic == "Negative", ]$Flowering.Days)
GH_leaf_aov <- aov(Flowering.Days ~ Transgenic, data=GH_leaf_df)
summary(GH_leaf_aov)
```

LH244 and B73 RNA-seq Comparison

```
setwd("G:/My Drive/Research/Expression.Analysis/RNAseq")
# Expression dataset
# Start with LH244 data
files <- list.files(path="quants", pattern="*.sf", full.names=TRUE, recursive=FALSE)
express <- read.table(file = "quants/NS-S01_S96_L003_quant.sf", sep = "\t", header = TRUE,
na.strings = "")
express = subset(express, select = -c(Length, EffectiveLength, TPM, NumReads) )
for (x in files) {
  print(x)
  express[substr(x,8,13)] <- read.table(file = x, sep = "\t", header = TRUE, na.strings = "")[4]
}
```

```
# Then we add the B73 data on top
files <- list.files(path="quants_B73", pattern="*.sf", full.names=TRUE, recursive=FALSE)
for (x in files) {
  print(x)
  express[substr(x,12,23)] <- read.table(file = x, sep = "\t", header = TRUE, na.strings = "")[4]
}

# Examine data for individual genes
lhcb10_data <- express[grepl("Zm00001eb357740", express$Name),]
fl3_data <- express[grepl("Zm00001eb341550", express$Name),]
clo2b_data <- express[grepl("Zm00001eb426360", express$Name),]
write.csv(lhcb10_data, "lhcb10_data.csv")
write.csv(fl3_data, "fl3_data.csv")

# Rearrange the data a bit
express <-
express[,c(2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,31,32,33,34,35
,36,37,38,39,40,41,42,28,29,30)]
# Pull out the genes with 0 variability
express <- t(express)
express <- express[ , which(apply(express, 2, var) != 0)]
express <- t(express)
res <- cor(express)
round(res, 2)
write.csv(res, "pear.csv")

# Show the clustering of the various tissue samples
express <- t(express)
test.pca <- prcomp(express, center = TRUE,scale. = TRUE)
library(ggfortify)
library(cluster)
autoplot(test.pca,label=TRUE, label.size=4, shape = FALSE) + theme_bw()
```

Promoter Searching Code

```
library(ggplot2)
plotExpr <- function(masterDF,gene,name,path) {
  df <- as.matrix(masterDF[masterDF$gene==gene,][,-1])
  tiss <- colnames(df)
  #colnames(df) <- 1:length(colnames(df))
  rownames(df) <- "TPM"
  df=data.frame(t(df))
  df$tissue = factor(rownames(df),levels=rownames(df))
  df$type =
factor(c(rep("root",18),rep("stem",3),rep("leaf",19),rep("stalk",8),rep("floral",6),rep("seed",25)))
  #typeof(df)
```

```
  plt1 = ggplot(df,aes(tissue,TPM,fill=type)) +
    geom_bar(stat = "identity") +
    ggtitle(name) +
    ylab("log(TPM)") +
    theme(title = element_text(size = 30),
          axis.title.x = element_blank(),
          axis.text.x = element_text(angle=45,hjust=1),
          axis.title.y = element_text(size=30),
          plot.margin = margin(1,1,1,2),
          legend.title = element_text(size = 25),
          legend.text = element_text(size = 25),
          legend.key.size = unit(3,"line"))
  plt1
  ggsave(paste(paste(name,gene,sep="-"),".png",sep=""), plt1, width = 20, height = 10,
path=path)
}
findHighTiss <- function(df,cutoff,tissues,limit=NULL,ignore=NULL){
  tissues = tissues - 1
  # First need to create the fictional profile based off the to.find
  fict <- df[1,-1]
  fict[] <- 0
  fict[,tissues] <- 10

  # Creating a function to make sure a base level of expression is achieved
  meansToFind = function(x) {
    result <- mean(x[tissues])
    return(result)
  }
  sufficient <- apply(df[,-1],1,meansToFind)
  highEnough <- df[sufficient>4,]

  # Calculate the distance measure based on data we care about
  coToFind = function(x) {
    return(cosine(as.numeric(x),as.numeric(fict)))
  }
  if (!is.null(ignore)) {
    #print(dim(fict))
    #print(dim(highEnough))
    fict <- fict[-(ignore-1)]
    distance <- apply(highEnough[,-c(1,ignore)],1,coToFind)
  } else {
    distance <- apply(highEnough[,-1],1,coToFind)
  }

  # Combine everything in an easy to view package
  result <- cbind(as.character(highEnough[,1]),sufficient[sufficient>4],distance)
```

```
  # Want to trim the dataset based on the list in limit
  if (!is.null(limit)) {
    result <- result[(result[,1] %in% limit),]
  }

  colnames(result) <- c("KEY","MEAN","DISTANCE")
  return(result)
}
# Expression dataset
express <-
read.csv(file="../raw/Processed_gene_atlas_and_stress_log2_flt1_fpkm_matrix_Feb2018.csv",
header=TRUE, stringsAsFactors=FALSE)
default.trim.mat <- cbind(gene=express[,1],express[,c(6:(ncol(express)-16))])

# Want to trim the maizegdb list
curated <- read.csv(file="../raw/MaizeGDB_curated_locusXgene_models.csv", header=TRUE,
stringsAsFactors=FALSE)
curated <- subset(curated, !grepl('G',KEY) & !grepl('G',NAME))[,1:2]

# Tests and GOIs
plotExpr(default.trim.mat,"Zm00001d042212","LD","../results/")
plotExpr(default.trim.mat,"Zm00001d042315","MADS69")

### Embryo Specific
embryoCandidates <-
findHighTiss(default.trim.mat,4,c(75:79),limit=curated$KEY,ignore=c(56:67))
embryoCandidates <- as.data.frame(embryoCandidates,stringsAsFactors=FALSE)
embryoCandidates <- merge(embryoCandidates,curated,by="KEY")
write.csv(embryoCandidates[,c(4,1,3,2)],file="../results/tissues/embryo/highinembryo.csv")

# Whole seed specific
wholeCandidates <-
findHighTiss(default.trim.mat[default.trim.mat$Whole.Seed_2DAP>4,],4,c(56:80),limit=curated
$KEY)
wholeCandidates <- as.data.frame(wholeCandidates,stringsAsFactors=FALSE)
wholeCandidates <- merge(wholeCandidates,curated,by="KEY")
write.csv(wholeCandidates[,c(4,1,3,2)],file="../results/tissues/embryo/highinseed.csv")
write.csv(wholeCandidates[,c(4,1,3,2)],file="../results/tissues/embryo/highinseed.csv")

# Literature sources
plotExpr(default.trim.mat,"Zm00001d020590","Aa2m","../results/tissues/embryo/")
plotExpr(default.trim.mat,"Zm00001d034413","globulin2","../results/tissues/embryo/")
plotExpr(default.trim.mat,"Zm00001d002768","ole1","../results/tissues/embryo/")
plotExpr(default.trim.mat,"Zm00001d045192","hyp1","../results/tissues/embryo/")
```

```
# Found via searching
plotExpr(default.trim.mat,"Zm00001d009292","fl3","../results/tissues/embryo/")
plotExpr(default.trim.mat,"Zm00001d037436","dzs18","../results/tissues/embryo/")
plotExpr(default.trim.mat,"Zm00001d045792","expb14","../results/tissues/embryo/")
plotExpr(default.trim.mat,"Zm00001d041781","zag2","../results/tissues/embryo/")
plotExpr(default.trim.mat,"Zm00001d035505","pmpm6","../results/tissues/embryo/")
plotExpr(default.trim.mat,"Zm00001d035505","pmpm6","../results/tissues/embryo/")

# Early seed specific
earlyCandidates <-
findHighTiss(default.trim.mat,4,c(56:58),limit=curated$KEY,ignore=c(59:80))
earlyCandidates <- as.data.frame(earlyCandidates,stringsAsFactors=FALSE)
earlyCandidates <- merge(earlyCandidates,curated,by="KEY")
write.csv(earlyCandidates[,c(4,1,3,2)],file="../results/tissues/embryo/highinearlyseed.csv")

# Endosperm Specific
endoCandidates <-
findHighTiss(default.trim.mat,4,c(76:79),limit=curated$KEY,ignore=c(56:67))
endoCandidates <- as.data.frame(endoCandidates,stringsAsFactors=FALSE)
endoCandidates <- merge(endoCandidates,curated,by="KEY")
write.csv(endoCandidates[,c(4,1,3,2)],file="../results/tissues/endo/highinendo.csv")

# Green-tissue Specific
# Lets look at PPC genes first:
plotExpr(default.trim.mat,"Zm00001d046170","PEP1","../results/")
plotExpr(default.trim.mat,"Zm00001d016166","PEP2","../results/")
plotExpr(default.trim.mat,"Zm00001d020057","PEP4","../results/")
plotExpr(default.trim.mat,"Zm00001d020057","glp1","../results/tissues/leaf/")
# Now lets use our searching technique:
leafCandidates <- findHighTiss(default.trim.mat,4,c(23:41),limit=curated$KEY)
leafCandidates <- as.data.frame(leafCandidates,stringsAsFactors=FALSE)
leafCandidates <- merge(leafCandidates,curated,by="KEY")
write.csv(leafCandidates[,c(4,1,3,2)],file="../results/tissues/leaf/highinleaf2.csv")
plotExpr(default.trim.mat,"Zm00001d017485","test","../results/")
test <- cbind(as.character(default.trim.mat$gene),(default.trim.mat$Whole.Seed_2DAP +
default.trim.mat$Whole.Seed_4DAP))

# Root Specific
rootCandidates <- findHighTiss(default.trim.mat,4,c(2:19),limit=curated$KEY)
rootCandidates <- as.data.frame(rootCandidates,stringsAsFactors=FALSE)
rootCandidates <- merge(rootCandidates,curated,by="KEY")
write.csv(rootCandidates[,c(4,1,3,2)],file="../results/tissues/root/allroot.csv")

crownCandidates <-
findHighTiss(default.trim.mat,4,c(15:19),limit=curated$KEY,ignore=c(2:14))
crownCandidates <- as.data.frame(crownCandidates,stringsAsFactors=FALSE)
```

```
crownCandidates <- merge(crownCandidates,curated,by="KEY")
write.csv(crownCandidates[,c(4,1,3,2)],file="../results/tissues/root/crown.csv")

# Use plotting funciton to explore candidate genes from the list
plotExpr(default.trim.mat,"Zm00001d015327","ubi1",".")
```