FUNCTIONAL AND EVOLUTIONARY GENOMICS STUDIES OF MAIZE SEED DEVELOPMENT

by

NELSON S. GARCIA

A dissertation submitted to the

Graduate School-New Brunswick

Rutgers, the State University of New Jersey

In partial fulfillment of the requirements

For the degree of

Doctor of Philosophy

Graduate Program in Plant Biology

Written under the direction of

Joachim Messing

And approved by

_____

_____

_____

_____

New Brunswick, New Jersey

May 2017

ABSTRACT OF THE DISSERTATION

Functional and Evolutionary Genomics Studies of Maize Seed Development

This dissertation is divided into two chapters with a common theme of investigating the role of several genes in seed development. Studies of the genetic basis of corn seed development provide not only answers to basic biological questions, but also have significant implications for nutritional and industrial uses. For example, the relative concentrations of different types of storage proteins in maize, called zeins, greatly affect the amount of essential amino acids lysine, methionine, and tryptophan, which are important for human and animal nutrition (Mertz, et al. 1964; Messing and Fisher 1991). The interaction of protein bodies with starch granules also affects kernel hardness and consequently the transportation and storage of corn grains (Wu, et al. 2010). The quantity and quality of another important household commodity – corn oil – also depends on the biosynthesis of triacylglycerols and their storage into lipid bodies which are mainly found in the embryo (White and Weber 2003). Clearly, understanding the many pathways involved in seed development is an important step towards improving its uses.

In the first chapter, the conservation of regulatory factors controlling gene expression of *zeins* throughout the Poaceae were investigated by taking advantage of oat-maize addition lines or OMAs. Oats and maize belong to two different subfamilies of the Poaceae, but it was possible to cross pollinate the two and obtain seeds, where one maize

chromosome at a time can be added to the whole set of oat chromosomes (Kynast, et al. 2001; Rines, et al. 2009). Therefore, one can examine whether oat has regulatory factors that can cause the expression of genes added by the single maize chromosomes, but regulated by a different maize chromosome. The results showed that recently diverged genes of the prolamin gene family, the *α-*, *β-*, and *δ-zeins*, were not expressed, whereas the older *γ-zein* genes were expressed. Further studies also showed that the oat homolog of a known regulator of *zein* gene expression called *Prolamin box binding factor 1* (*Pbf1*) was able to trans-activate *γ-zein* expression in transient expression assays, indicating that it can substitute for the function of its maize homolog. The wheat *Dx5* gene, presumably the founding member of the prolamin gene family (Xu and Messing 2009), is also expressed in a maize transgenic line, even when *zein* transcription factors, *Pbf1* and *O2*, are knocked down. Overall, our data indicates that the regulation of gene expression of old copies of seed protein genes is conserved, whereas the regulation of younger copies seems to have diverged.

Mutant collections are also great resources for identifying genes involved in seed development. Many researchers have used two mutant resources to study maize seed development – a collection of *defective kernel* (*dek*) mutants developed from EMS mutagenesis and mutants from *Mutator* insertion collections. Some of the genes that were identified using these two resources encode a heat shock binding protein (Fu, et al. 2002), a chloroplast DNA polymerase (Udy, et al. 2012), an RNA splicing factor (Fouquet, et al. 2011), and several enzymes (Lid, et al. 2002; Wang, et al. 2014). The second chapter uses a new insertion mutant resource that has been developed based on a *Ds* transposable element tagged with GFP (*Dsg*) (Li, et al. 2013). Selection of a seed with a defective

kernel phenotype has led to the identification of *dek38-Dsg*, a recessive lethal mutant that encodes a co-chaperone protein called Tel2-interacting protein 2 (TTI2). TTI2 interacts with two other co-chaperones called Telomere maintenance 2 (Tel2) and Tel2-interacting protein 1 (TTI1) to form the TTT complex that is required to maintain steady-state levels of phosphatidylinositol 3-kinase-related kinases (PIKKs) which are essential for development (Hurov, et al. 2010; Takai, et al. 2010). As a co-chaperone for HSP90, the TTT complex can interact directly with PIKKs to aid in their proper protein folding and assembly into functional complexes. Reversion analysis and multiple *Dsg* excision footprint alleles established the linkage of the gene to the phenotype. Histological sections of developing seeds show that many aspects of development are affected in *dek38-Dsg*. The arrest of embryo development at the transition stage is similar to the PIKK *Target of rapamycin (TOR)* mutation in *Arabidopsis* and is consistent with reduced level of TOR protein in *dek38-Dsg*. Pollen transmission problems, as shown by the significantly lower number of GFP kernels when *dek38-Dsg* is used as a male, indicate that TTI2 is important for male reproductive cell development. Cloning of maize *Tel2* and *Tti1* homologs and yeast two-hybrid assays show that the interaction of TEL2 to TTI1 and TTI2 is conserved in maize. Overall, the results open up new lines of investigations into the roles of co-chaperones in seed development, and show the advantages of the *Dsg* insertion collection in maize for functional analysis of a gene.

To sum up, my dissertation presented evidence of the conservation of gene expression in older copies of seed storage protein genes in maize, and showed the utility of a new mutant resource in maize by characterizing the effects of *Tti2* mutation for the first time in plants. Our findings provide new avenues to the roles of gene expression and

genome evolution, as well as the role of co-chaperones in seed development. For example, how do promoters or DNA binding proteins diverge after gene amplification? What are the direct targets of *Tti2* during seed development? The approaches taken here have advanced our understanding of the genetic basis of seed development and allowed us to suggest further directions in our quest to further improve one of our most important crops through molecular breeding.

References

Fouquet R, et al. 2011. Maize Rough Endosperm3 Encodes an RNA Splicing Factor Required for Endosperm Cell Differentiation and Has a Nonautonomous Effect on Embryo Development. Plant Cell 23: 4280-4297.

Fu S, Meeley R, Scanlon MJ 2002. empty pericarp2 Encodes a Negative Regulator of the Heat Shock Response and Is Required for Maize Embryogenesis. Plant Cell 14: 3119–3132.

Hurov KE, Cotta-Ramusino C, Elledge SJ 2010. A genetic screen identifies the Triple T complex required for DNA damage signaling and ATM and ATR stability. Genes Dev 24: 1939-1950.

Kynast RG, et al. 2001. A Complete Set of Maize Individual Chromosome Additions to the Oat Genome. Plant Physiology 125: 1216-1227.

Li Y, Segal G, Wang Q, Dooner HK. 2013. Gene tagging with engineered Ds elements in maize. In: Peterson T, editor. Plant Transposable Elements: Methods and Protocols: Humana Press. p. 83-99.

Lid SE, et al. 2002. The defective kernel 1 (dek1) gene required for aleurone cell development in the endosperm of maize grains encodes a membrane protein of the calpain gene superfamily. Proc Natl Acad Sci U S A 99: 5460-5465.

Mertz ET, Bates LS, Nelson OE 1964. Mutant gene that changes protein composition and increases lysine content of maize endosperm. Science 145: 279–280.

Messing J, Fisher H 1991. Maternal effect on high methionine levels in hybrid corn. Journal of Biotechnology 21: 229-237.

Rines HW, et al. 2009. Addition of individual chromosomes of maize inbreds B73 and Mo17 to oat cultivars Starter and Sun II: maize chromosome retention, transmission, and plant phenotype. Theor Appl Genet 119: 1255-1264.

Takai H, Xie Y, de Lange T, Pavletich NP 2010. Tel2 structure and function in the Hsp90-dependent maturation of mTOR and ATR complexes. Genes Dev 24: 2019-2030. doi: 10.1101/gad.1956410

Udy DB, Belcher S, Williams-Carrier S, Gualberto JM, Barkan A 2012. Effects of Reduced Chloroplast Gene Copy Number on Chloroplast Gene Expression in Maize. Plant Physiology 160: 1420-1431.

Wang G, et al. 2014. Proline responding1 Plays a Critical Role in Regulating General Protein Synthesis and the Cell Cycle in Maize. Plant Cell 26: 2582-2600.

White PJ, Weber EJ. 2003. Lipids of the kernel. In: White PJ, Johnson LA, editors. Corn: Chemistry and Technology. St. Paul, Minnesota: American Association of Cereal Chemists, Inc.

Wu Y, Holding DR, Messing J 2010. Gamma-Zeins are essential for endosperm modification in quality protein maize. Proc Natl Acad Sci U S A 107: 12810–12815.

Xu JH, Messing J 2009. Amplification of prolamin storage protein genes in different subfamilies of the Poaceae. Theor Appl Genet 119: 1397-1412.

ACKNOWLEDGMENTS

I would like to thank my thesis adviser, Dr. Joachim Messing, who took a chance on me to work in his excellent lab and for letting me explore interesting scientific questions. The wisdom and insights he imparted to me will serve as guide in my future scientific undertakings. I also thank the whole Messing Lab for the support and friendship throughout my doctoral days.

I am also very thankful to Dr. Hugo Dooner, whose greatness in genetics I can only hope to equal someday. His advices have guided me enormously and helped me focus on the questions that matter.

I also got a lot of support and help from the various labs at Waksman Institute. I'm especially thankful to the Gallavotti lab for sharing many of the protocols, plasmids, and microbial strains that I used.

Graduate school means spending countless hours in isolation. In the midst of this, my parents, sisters, relatives, and friends always cheered me on and offered company and words of encouragement. For this I am forever thankful to all of them.

Table of Contents

LIST OF TABLES

LIST OF ILLUSTRATIONS

Chapter 1 Evolution of gene expression after gene amplication

Abstract

A rather unique approach to investigate the conservation of gene expression of prolamin storage protein genes across two different subfamilies of the Poaceae was undertaken. Oat plants carrying single maize chromosomes called oat-maize addition (OMA) lines permitted the determination of whether regulation of gene expression was conserved between the two species. Results show that *γ-zeins* are expressed in OMA7.06, which carries maize chromosome 7 even in the absence of the trans-acting maize Prolamin-box Binding Factor (PBF), which regulates their expression. This is likely because oat PBF can substitute for the function of maize PBF as shown in our transient expression data, using a *γ-zein* promoter fused to GFP. Despite this conservation, the younger, recently amplified prolamin genes in maize, absent in oat, are not expressed in OMAs 4.42 and 7.06. However, maize can express the oldest prolamin gene, the wheat HMW-glutenin *Dx5* gene, even when maize *Pbf* is knocked down (via PbfRNAi), and/or another maize transcription factor, Opaque-2 (O2) is knocked out (in maize *o2* mutant). Therefore, older genes are conserved in their regulation, whereas younger ones diverged during evolution and eventually acquired a new repertoire of suitable transcriptional activators.

Introduction

Cereal species are among the most studied plants because of their agro-economic importance. They are currently categorized into three subfamilies – Ehrhartoideae (rice), Pooideae (oat, wheat, barley), and Panicoideae (maize, sorghum, millets) (Kellogg 2001). Relative to other cereals, oat (*Avena sativa*, 2n=6x=42) and maize (*Zea mays*, 2n=2x=20) are evolutionarily distant, having shared a common ancestor with all the rest of the cereals about 50 million years ago (Gaut 2002; Salse and Feuillet 2007). Despite being remotely related, maize pollen can be used to pollinate oat flowers and can sometimes produce fertile partial hybrids containing single maize chromosome additions. The complete set of 10 maize chromosomes has been added into oat plants as single chromosome additions and is collectively called oat-maize addition (OMA) lines (Kynast, et al. 2001; Rines, et al. 2009). Wheat and maize are also evolutionarily distant, diverging from each other at about the same time when oat and maize formed separate lineages about 50 million years ago (Gaut 2002; Salse and Feuillet 2007).

Aside from being a major source of carbohydrates, cereals can also be a source of protein for both human and livestock nutrition. During seed development, they accumulate storage proteins mainly in the form of alcohol-soluble prolamins, with the exception of rice and oat, whose grains mainly accumulate saline-soluble globulins (Shewry and Halford 2002). The prolamins are deposited into endoplasmic reticulum-associated protein bodies (PBs) in maize, but oat and wheat have protein storage vacuoles (PSVs) in addition to PBs (Lending, et al. 1989; Herman and Larkins 1999).

In maize, prolamins are called zeins and are encoded by a medium-sized gene family. Based on structural differences, zeins are subdivided into α, β, γ, and δ groups.

The α-zeins are encoded by five multi-copy gene clusters and a single gene locus (*19-* and *22-kDa α-zeins*) located on chromosomes 1, 4, and 7 (Song and Messing 2003; Miclaus, et al. 2011), whereas γ-zeins are encoded by three genes each as single copies (*27-* and *50-kDa γ-zeins* on chromosome 7, *16-kDa γ-zein* on chromosome 2). The δ-zeins are encoded by two single copy genes (*18-kDa* on chromosome 6 and *10-kDa* on chromosome 9) and the β-zein is encoded by one single copy gene (*15-kDa β-zein* on chromosome 6) (Xu and Messing 2008; Feng, et al. 2009).

Based on syntenic alignments and nucleotide substitution rates, it has been proposed that prolamins arose from the duplication of a globulin gene and evolved into a high-molecular weight (HMW) glutenin. Thus HMW glutenin genes likely represent the founding genes of this family within the grasses (Xu and Messing 2009). During evolution of the species of this family, the HMW-glutenin gene got amplified in a tandem fashion, but more importantly in dispersed chromosomal locations. Dispersal led to greater divergence so that today there appeared to be three groups of related prolamins. Group III is comprised by the HMW-glutenin, present only in the subfamily Pooideae. Interestingly, they are critical for baking quality and must have been lost in the other two subfamilies of the Poaceae after copies dispersed and diverged. Therefore, we need wheat today for baking bread. The other wheat prolamins, called gliadins, are also important for baking and belong to group II, which includes the maize γ–zein and β-zeins. Group I is absent in wheat and represents the most recent amplification and dispersal of prolamin genes.

Dispersal of copies also led to tandem amplification resulting in gene clusters, but those can vary in maize haplotypes (Song, et al. 2001). There are two transcription

factors – Prolamin-box binding factor (PBF) and Opaque2 (O2) that have been shown so far to regulate the expression of a significant number of *zein* genes. *Pbf* is located on maize chromosome 2 and encodes a Dof class of zinc finger transcription factor that engages in protein-protein interaction with O2 (Vicente-Carbajosa, et al. 1997). The role of PBF in regulating zein gene expression was elucidated in PbfRNAi transgenic lines, showing that knocking down expression of *Pbf* reduces the accumulation of only the *22 kDa α-* and *27-kDa γ-zeins* (Wu and Messing 2012). The *O2* gene on the other hand is located on chromosome 7 and encodes a leucine zipper-type of transcription factor (Schmidt, et al. 1990) that has been shown to affect the accumulation of most 22-kDa α-zeins and the 15-kDa β-zein (Schmidt, et al. 1992; Cord-Neto, et al. 1995; Song and Messing 2003). In maize, PBF has been shown to bind a TGTAAAG conserved motif called the P-box, whereas O2 binds a CACGTA/C (also called A/C-box) motif downstream to the P-box (Schmidt, et al. 1992; Vicente-Carbajosa, et al. 1997). In wheat, the *Dx5* promoter has also been shown to contain P-box motifs, but has also a G-box like motif (TTACGTGG) located upstream of the P-box. Evidence through gel shift assays has shown that the G-box like motif has similar affinity to the A-box in maize, and therefore might be another binding site for O2 (Norre, et al. 2002).

Although DNA binding studies have been useful in characterizing transcriptional activators, they are not conclusive for in vivo interactions. A different approach is facilitated in the OMA, where cis- and trans-acting sites can be combined from evolutionarily diverged regulatory processes. Using this path, we show evidence that oat PBF can substitute for the function of maize PBF (ZmPBF) in OMA lines. In contrast, ZmPBF and maize O2 (ZmO2) do not seem to be major factors for the expression of the

wheat HMW-glutenin *Dx5* (group III prolamin) in transgenic maize. However, group I prolamins, present in the Panicoideae, have diverged to a degree that distantly related transcription factors can no longer activate their expression. The resulting knowledge not only provides us with an understanding of the evolution of gene amplification and expression, but also will be useful for translational genetics/breeding of agronomic traits.

Materials and Methods

Plant Materials

Drs. Ronald Phillips and Howard Rines of the University of Minnesota kindly provided the set of oat-maize addition (OMA) lines. These lines represent single disomic maize chromosome additions (either from maize B73, Mo17, or Seneca 60) in the background of either Starter or Sun2 oat cultivars. For nomenclature purposes, these materials are named in the format OMAy.z, wherein y is the maize chromosome number added, and z is used to trace its original chromosome recovery event (Kynast, et al. 2001; Rines, et al. 2009).

Transgenic maize expressing *Dx5* controlled by its native promoter was obtained from the Scott Lab at Iowa State University (Sangtong, et al. 2002) and was backcrossed several times to B73 (Zhang, et al. 2013). A maize transgenic line with a knock-down of *ZmPbf* by PbfRNAi was generated in the Messing Lab (Wu and Messing 2012). The *o2* mutant used was also a Messing Lab stock. A small section was cut from the top of the seed from the cross *Dx5*/-, *O2/o2* X *pbfRNAi*/-, *O2*/*o2* and saved for protein extraction, and the rest of the seed containing the embryo was saved for germination and genotyping. Primers used to genotype *Dx5*, *O2* and *Pbf* were described previously (Sangtong, et al.

2002; Wu and Messing 2012). Individuals with *Dx5*/-, *o2*/*o2*, *PbfRNAi*/- ; *Dx5*/-, *o2*/*o2*, -/- ; and *Dx5*/-, *O2*/*o2*, *PbfRNAi*/- genotypes were selected for prolamin extraction and Dx5 phenotyping.

Dx5 phenotyping

The seed section saved for protein extraction was ground into powder and protein was extracted as described (Zhang, et al. 2013) with some modifications. Briefly, 600 µl of borate extraction buffer was added in the powder and left overnight in a shaker. The mixture was then centrifuged and 200 µl of the supernatant was mixed with equal amount of 100% ethanol and shaken for over 2 hours to extract the prolamins, which include the wheat Dx5 protein. The prolamin extract was then run in an SDS-PAGE gel and blotted on a PVDF membrane. Western blot using anti-Dx5 antibodies was then performed to examine the accumulation of the Dx5 protein (Zhang, et al. 2014).

PCR assays

Genomic DNA was extracted using the CTAB method, whereas mRNA was extracted using Qiagen RNEAsy kit. Maize simple sequence repeat (SSR) markers, one on each chromosome arm, were used to verify the presence of intact maize chromosomes in the OMA lines (Rines, et al. 2009). PCR primers to amplify various zein genes as well as *ZmO2* and *ZmPbf* are described elsewhere (Song and Messing 2003; Wu and Messing 2009; Miclaus, et al. 2011). We also used a previously described primer pair for gene expression of actin as a control (Brautigam, et al. 2005). Total RNA was extracted from leaf (30 days after sowing), stem (30 days after sowing), seedling (14 days after sowing)

and immature endosperm (milk grain stage, 7-18 days post-anthesis) using Qiagen RNEasy Kit. The same amount of RNA for each sample was then treated with DNAse and reverse transcribed to cDNA using SuperScript III First Strand Synthesis Kit (LifeTechnologies) using Oligo-dT primers. The cDNA was then used for PCR using Platinum Pfx polymerase (LifeTechnologies).

The gene expression of *AsPbf* copies was done based on the protocol used by Song and Messing (2002) and Miclaus et al. (2011). Briefly, PCR primers were designed to amplify a 300bp region that contains sequence polymorphisms between the two copies. The RT-PCR products were then ligated into pGEM-Teasy vector and transformed in E. coli DH10B. Plasmid DNA were then extracted from 96 colonies and the clones were sequenced using M13 primers. The proportion of clones that correspond to each copy was taken as an indication of their relative expression levels.

To monitor the expression of *AsPbf* and *27 kDa γ-zein* during OMA 7.06 endosperm development, PCR was performed on cDNA from developing endosperm at 7, 9, 11, and 13 DAA using *AsPbf* and *27 kDa γ-zein* primers.

Cloning of oat *Pbf* cDNA

To clone the oat *Pbf* (*AsPbf*), we designed PCR primers from *Pbf* sequence of barley (Genbank accession no. AJ000991.1). Primers were selected as close to start or stop codons as possible. The primer sequences used were HvPBF-L (GAGGAAGTGTTTTCGTCCAA) and HvPBF-R (CATCAGGGAGGTGCTGTTGA). Total RNA was extracted from immature seeds (15 DAA) of oat (cultivar Starter) and was used for reverse transcription using RT-PCR. The resulting cDNA was used as

template for PCR amplification using the barley *Pbf* primers. Amplicons were cloned in pGEM Teasy vector (Promega) and transformed into E. coli DH10B, after which we picked single colonies for plasmid extraction and DNA sequencing using M13 primers. The sequences were then BLASTed against the NCBI non-redundant nucleotide database to see if they have significant hits against known Pbf sequences. The 5′ and 3′ rapid amplification of cDNA ends (5′ and 3′ RACE) were then conducted using the GeneRacer Kit (LifeTechnologies) to reveal the full coding sequence (CDS) of the candidate oat Pbf.

Plasmid constructs and trans-activation assay

To use GFP as a reporter of PBF activity, we modified an existing plasmid in our lab (PTF102-P4) that has a GFP gene driven by a 27-kDa γ-zein promoter (P27:GFP) (Wu and Messing 2012) (Figure 5). We then removed the Bar gene in the T-DNA of pTF102-P4 using BglII and Xho1 restriction enzymes, and replaced it with our candidate AsPbf1 gene such that it is now under the control of the 35S promoter (35S:AsPbf1). The resulting plasmid now contains a T-DNA with P27:GFP and 35S:AsPbf1 expression cassettes. To clone AsPbf1 into the plasmid, it was amplified using primers with ends that overlap the target insertion site sequence (AsPbf-F AGCTAGATTGTTGAGCATTACATCAGGGAGCTGCT and AsPbf-R TCATTTGGAGAGGACCATGGAGGAAGTGTTGTCAT). This enabled us to use a Gibson reaction kit (NEB) to piece the linearized PTF102-P4 without the Bar gene and the candidate AsPbf together. The ligation ends were then sequenced with the ABI3730XL platform to verify the insertion, and to identify which copy was inserted. This new plasmid was then renamed pTF102-P6. Another plasmid also available in our

lab, pTF102-P5, was used as a positive control. This plasmid contains P27:GFP and *ZmPbf* driven by 35S promoter (35S: ZmPBF) (Wu and Messing 2012). The original pTF102-P4 with only the P27:GFP on the other hand served as the negative control.

The plasmid constructs described above were used in a transient expression experiment by *Agro*-injection method in tobacco (*Nicotiana benthamiana*) leaves as previously described (Gallavotti, et al. 2011). Briefly, the plasmid constructs were transformed into *Agrobacterium tumefaciens* strain EHA101 and plated on YEP-agar media. Single colonies were then picked and grown in liquid YEP media for about 18-20 hours, aiming for an optical density (OD600) measurement of 1. The cultures were then centrifuged and the bacterial pellets were resuspended and washed with 10 mM MgCl2. This cycle of centrifugation and washing was done twice, until the pellet was finally resuspended in 5 ml of 10 mM MgCl2. This was mixed with an equal amount of Agrobacterium suspension containing a plasmid expressing p19 protein to increase the GFP signal. Tobacco leaves were then injected on the abaxial side using a syringe. At least three leaves were injected for each treatment. Imaging of the leaves were done five days after injection using a fluorescence microscope with an appropriate filter. The same camera setting was used for all samples.

GFP quantification

Western blot was used to measure the GFP signal from the trans-activation assay in a semi-quantitative way. Using a leaf punch, leaf tissues from *Agro*-injected leaves were collected in 2ml tubes and ground in liquid nitrogen. Total protein was then extracted using Tris-beta mercaptoethanol buffer. To ensure equal loading, the protein

extracts were quantified using the DC Protein Quantification kit (BioRad). One microgram of the protein extract (three biological replicates per treatment) were loaded onto a 15% SDS-PAGE for electrophoresis, and then blotted on a PVDF membrane (BioRad). The membrane was then blocked with 5% nonfat dry milk in TBST. Rabbit GFP antibody (LifeTechnologies) was then used to probe the membrane, followed by goat anti-rabbit IGG secondary antibody with HRP conjugate (LifeTechnologies). The membrane was then incubated for 5 minutes in HRP substrate (BioRad Western ECL) and exposed on x-ray film.

To quantitate the GFP bands from the Western blot, densitograms were made using ImageJ software (http://imagej.nih.gov/ij/). First, Western blot images were saved as TIFF files and converted to 8-bit image. Bands were selected using the rectangle tool and the background was cropped. The saved band images were then lined up vertically on PowerPoint, allowing sufficient white space in between, and saved as TIFF file. The new TIFF file was opened in ImageJ and converted back to an 8-bit image. The rectangle tool was used to drag a box surrounding the bands to define the lane. The densitogram for the bands was then created using "Plot lanes" command. The area inside each peak in the densitogram, which corresponds to the band density, was then calculated using the "Wand" tool.

GFP fluorescence was also measured directly on the image files using ImageJ. The pictures were opened in ImageJ and then converted to 16-bit image files. For each image, the whole field was selected using the rectangle tool and measurements were set to take integrated density, area measured, and mean gray value (integrated density over area measured). The mean gray values were then compared between treatments.

Sequence analysis

DNA and protein sequence alignment was done in MegAlign program of DNASTAR Lasergene 10 software package. The maximum likelihood phylogenetic tree of predicted protein sequences of previously described PBF from maize (*ZmPBF*, Genbank accession no. NP_001105400.1), rice (OsPBF, Genbank accession no. BAA78574.2), wheat (*TaPBF*, Genbank accession no. CAA09976.1), and barley (*HvPBF*, Genbank accession no. CAA04440.1), and a putative PBF sequence from *Brachypodium* (BdPBF, Genbank accession no. XP_010239574.1) was created using the JTT model (Jones, et al. 1992) as implemented in software package MEGA (Tamura, et al. 2007).

Results

Expression of ZmPbf, O2, and zeins in OMA lines

OMA lines represent single disomic maize chromosome additions. The presence of intact corresponding maize chromosomes on the OMA lines were verified by PCR assay using simple sequence repeat-markers (SSR) (Figure 1-1A). Both SSR primer pairs on the short and long chromosome arms successfully amplified, indicating that the maize chromosomes were present. The zeins and their regulator genes (*ZmO2* and *ZmPbf*) were also PCR-amplified and mapped correctly to the OMA lines with the corresponding maize chromosome addition (Figure 1-1B). *Zein* regulators *ZmO2* and *ZmPbf* were PCR-amplified on OMA7.06 and OMA2.51 respectively, whereas the *zeins* were PCR-

amplified on OMA2.51, 4.42, 6.33, 7.06, and 9.41. The oat parent used in creating the OMA lines did not give any amplification.

For the gene expression assay, we found that both *zein* regulators *ZmO2* and *ZmPbf* were expressed in immature endosperm of OMA7.06 and OMA2.51, respectively (Figure 1-1B). For the zeins, we were only able to detect the expression of two *γ-zein* genes (*27-* and *50-kDa*), both of which are on OMA7.06. No expression was detected for the rest of the zeins. The expressed zeins also maintained their tissue specificity as they were expressed only in endosperm tissues but not in leaves. Therefore, regulation of gene expression must have diverged between group I and group II prolamin genes between these species.

Cloning and characterization of oat *Pbf* (*AsPbf*)

Since previous studies have shown that ZmPBF is the main transcriptional regulator of the *27-kDa γ-zein* (Marzabal, et al. 2008; Wu and Messing 2012), its expression in the absence of *ZmPbf* in OMA7.06 led us to hypothesize that the ZmPbf homologue in oat can substitute for it. Because barley is more closely related to oat than maize, we cloned *AsPbf* by PCR using primers designed from a previously cloned barley *Pbf* (*HvPBF*) gene (Mena, et al. 1998) to validate our assumption. Indeed, PCR of oat immature endosperm cDNA using the barley primers gave us an amplicon of approximately 1 kb in size. Sequencing and BLAST analysis of the amplicon showed that our sequence is very similar to barley and wheat *Pbf* genes, as well as other Dof domain-containing genes from other grasses.

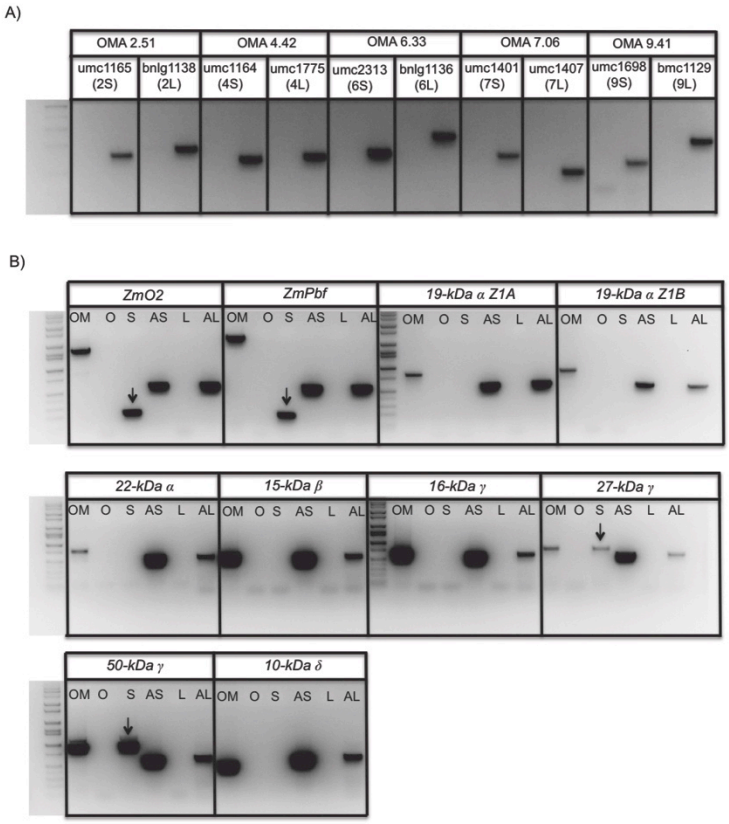


Figure 1-1. A) Verification of presence of maize chromosomes in OMA lines. For each SSR marker, the first lane is the oat parent. The second lane is the OMA line. B) Expression of maize Pbf, maize O2, and zeins in OMA lines. Legend: OM – PCR of OMA line where the gene being assayed is present; O – PCR of oat parent control; S – RT-PCR of seed mRNA; AS – actin control for RT-PCR of seed mRNA; L – RT-PCR of leaf mRNA; AL – actin control for RT-PCR of leaf mRNA. Arrows indicate genes that are expressed in the seed. OMA 2.51 was used to assay for ZmPbf and 16-kDa γ-zein, OMA 4.42 for 19-kDa α-zein Z1A and 22-kDa α-zein, OMA 6.33 for 16-kDa β-zein, OMA 7.06 for 19-kDa α-zein Z1B, 27-kDa γ-zein, 50-kDa γ-zein, and ZmO2, and OMA 9.41 for 10-kDa δ-zein.

The 5′ RACE revealed sequence variations upstream of the start codon, which indicates variation in 5′UTR sequences and the presence of at least two copies of *AsPbf*. The presence of small indels and SNPs within the CDS of these two genes also support the presence of two copies. We named these two copies *AsPbf1* and *AsPbf2*. AsPBF1 (325-aa long) is slightly shorter than AsPBF2 (328-aa long) but they are highly similar, sharing 87.7% predicted protein sequence identity (Figure 1-2B).

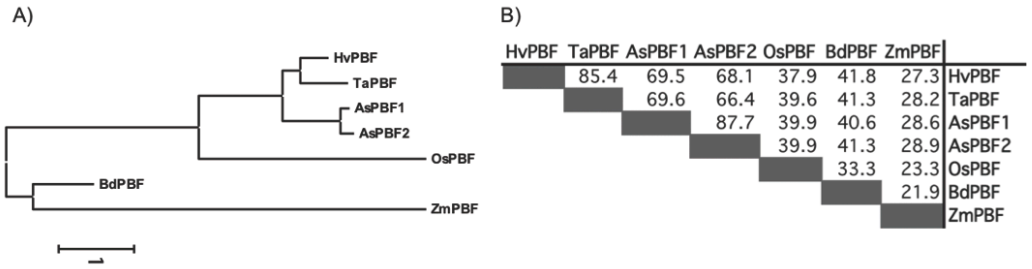


Figure 1-2. A) Phylogenetic tree of PBF proteins from several grass species. The bar below is the distance scale in amino acid substitutions per site. B) Pairwise percent identity of the protein sequences. Legend: HvPBF – barley PBF, TaPBF – wheat PBF, AsPBF – oat PBF, OsPBF – rice PBF, BdPBF – Brachypodium PBF, ZmPBF – maize PBF

The predicted protein sequences of AsPbf1 and AsPbf2 were compared to the Pfam database (http://pfam.sanger.ac.uk/search); indeed both possess a Dof domain, a defining characteristic of PBF proteins. The Dof domains are close to the amino terminal end, consistent with PBF proteins from other grasses (Figure 1-3A). The Dof domain is highly conserved between oat, barley, and wheat PBF proteins, which is located between the 31st and 93rd amino acid residues. The two AsPBF sequences also contain an asparagine-rich C-terminus like the other PBF proteins.

The result of the 5′RACE indicated that the two *Pbf* copies are not expressed equally because of the difference in the number of reads of the two copies. To verify this, we cloned the RT-PCR products in pGEM-Teasy vector, transformed them in E. coli DH10B, and selected 96 white bacterial colonies for sequencing with the ABI3730XL platform. The presence of two small indels and SNPs between the two *AsPbf* copies allowed us to make a quantitative comparison of the expression level of the two copies. The result of the sequencing showed that the *AsPbf1* copy is present in 87 out of the 96 clones, while sequence reads that correspond to *AsPbf2* showed up in only 9 clones. This is evidence that *AsPbf1* is more strongly expressed than *AsPbf2*.

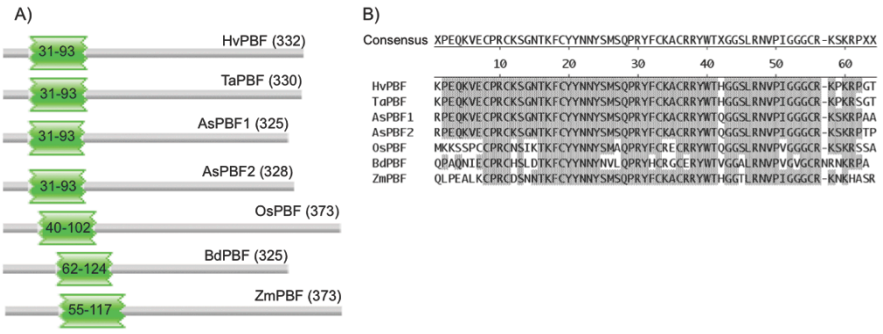


Figure 1-3. A) Comparison of PBF proteins from different grass species. The green boxes with numbers indicate the position of the Dof domain. The number after the protein name is the length of the protein. B) Alignment of Dof domain of PBF proteins from different grasses.

The phylogenetic tree based on predicted protein sequences of PBF sequences from different grasses agrees with their known phylogenetic relationships (Figure 1-2A). HvPBF and TaPBF are grouped together as members of the Triticeae tribe, and the two AsPBF sequences are more closely related to this tribe than to the PBF sequences from other species. It is interesting to see that the putative BdPBF did not show a close relationship to oat, barley, and wheat given that they all belong to the Pooideae subfamily. Because of the absence of the oat genome sequence, it is currently not possible to investigate if the two *AsPbf* copies are being contributed by two of the three subgenomes of oat, or are duplicated genes within one subgenome. However, the presence of only one putative *Pbf* in Brachypodium indicates that the two copies are more likely contributed by two of the three oat genomes.

We then examined the spatial expression pattern of our candidate oat *Pbf* genes by investigating their expression in oat leaf, stem, seedling, and root tissues in addition to

seed endosperm (Figure 1-4A). Like other *Pbf* genes, our candidate *AsPbf* genes are only expressed in the endosperm. In addition, the timing of expression of the *27-kDa γ-zeins* coincides with the expression of the *AsPbf* in OMA 7.06 (Figure 1-4B).

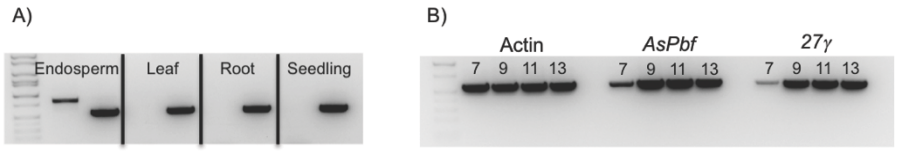


Figure 1-4. A) Expression of *AsPbf* from different oat tissues. For each tissue, the first lane is the RT-PCR for *AsPbf* while the second lane is the actin control. B) Expression of *AsPbf* and *27-kDa γ-zein* in OMA 7.06 at 7, 9, 11, and 13 days after anthesis (DAA).

AsPBF can trans-activate *27-kDa γ-zein*

To investigate whether the AsPBF that we cloned can indeed substitute for ZmPBF′s function in OMA 7.06, a transient expression assay in tobacco leaves was performed using plasmids shown in Figure 1-5. To monitor the activation of *27-kDa γ-zein* promoter by oat PBF, *AsPbf1* was selected since it has higher expression compared to *AsPbf2* based on our previous gene expression analysis. The fluorescences of GFP were compared among treatments in absence of PBF, presence of ZmPBF and presence of AsPBF1. As seen in Figure 1-6A, the fluorescence of GFP in the treatment containing *ZmPbf* and *AsPbf1* was more intense than the control without any *Pbf* for all of the three replicates, indicating that AsPBF1 activates the *27-kDa γ-zein* promoter just like ZmPBF. Quantification of GFP from the image files indicates that the presence of *AsPbf1* resulted in a twofold increase in fluorescence on average over the control without any *Pbf*: this is statistically significant ($p<0.05$) based on a T-test. The densitograms of the Western blot

also showed that the GFP band is twice more intense on average in the treatment with *AsPbf1* than without *AsPbf1* and the difference is also statistically significant (Figure 1-6B).

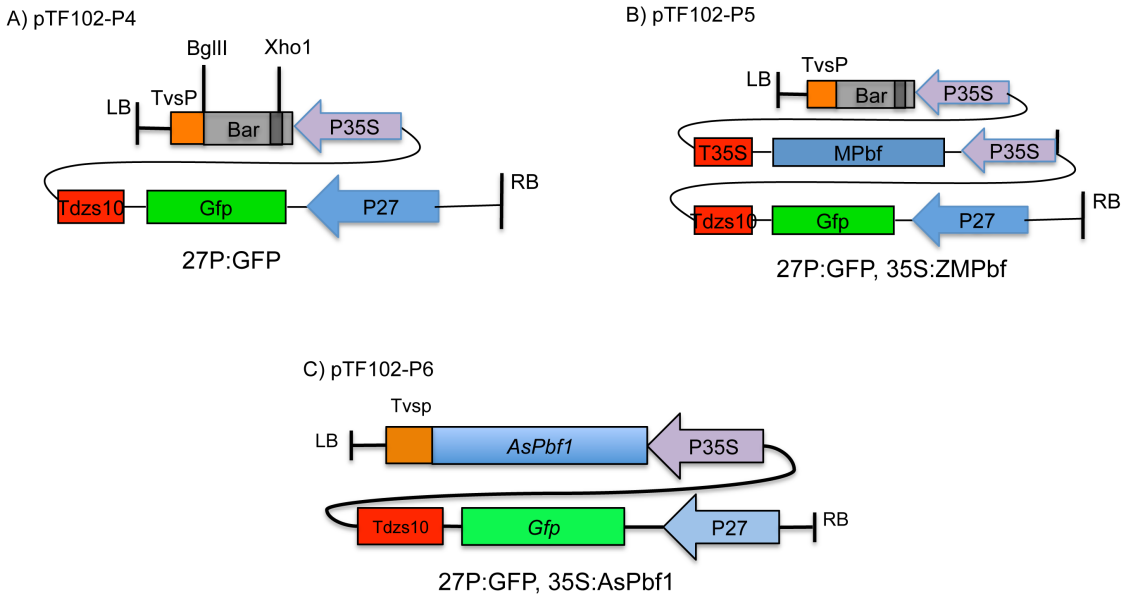
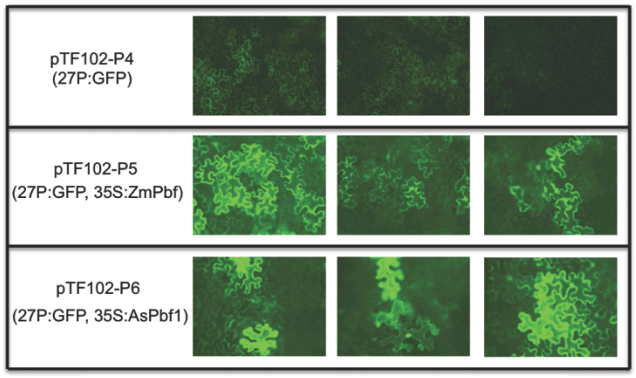


Figure 1-5. Plasmid constructs used for transient expression assay. A) pTF102-P4 contains a *Gfp* gene driven by *27-kDa γ-zein* promoter. B) pTF102-P5 contains maize *Pbf* (*ZmPbf*) driven by 35S promoter in addition to 27P:GFP. C) pTF102-P6 is a modified pTF102-P4, with the *Bar* gene replaced by *AsPbf1*.

Wheat *Dx5* is not controlled by Pbf or O2

If the expression of the maize *γ-zeins* (Panicoideae) is conserved in oat (Pooideae), one could ask whether expression of prolamins from Pooideae will also be conserved in Panicoideae. The closest prolamins to the γ-zeins in the Panicoideae subfamily are the glutenins and the gliadins (Xu and Messing 2009). Indeed, transformation of wheat prolamin *Dx5*, a HMW-glutenin, with its native promoter into maize has led to its successful expression in maize endosperm (Sangtong, et al. 2002), but it was unclear which maize trans-acting factor activated the expression of the wheat glutenin gene. Surprisingly, Dx5 protein accumulation was not reduced in the *PbfRNAi* or *o2* backgrounds. There is also no visible reduction of Dx5 in the *PbfRNAi*;*o2* background (Figure 1-7). These results suggest that the *Dx5* gene was regulated differently from the *γ-zeins*. Given that Dx5 is older (group III) than the *γ-zein* genes, it appears that the group II prolamin genes have acquired cis-acting elements for their regulation by PBF and O2 after amplification from group I.

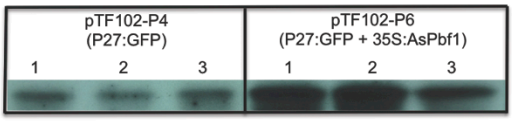
A)



Construct	Replicate	Area	Integrated density	Mean Gray Value
pTF102-P4*	1	153114	3338176	21.80
	2	153340	3128425	20.40
	3	153567	2617827	17.05
			Average	19.75
pTF102-P6*	1	153340	5845432	38.12
	2	153567	5624795	36.63
	3	153340	7514108	49.00
			Average	41.25

*p-value = 0.003

B)



Construct	Replicate	Densitogram value
pTF102-P4*	1	6953.7
	2	5561.4
	3	5608.6
	Average	6041.2
pTF102-P6*	1	12076.8
	2	12938.6
	3	10812.7
	Average	11942.7

*p-value = 0.0008

Figure 1-6. A) Tobacco leaf transient expression experiment to assay the activity of PBF using GFP as a reporter. Software quantification of fluorescence from the image files using ImageJ is shown at the bottom. B) Quantification of GFP signal in Western blot using anti-GFP antibody. The densitogram for the bands are shown below.

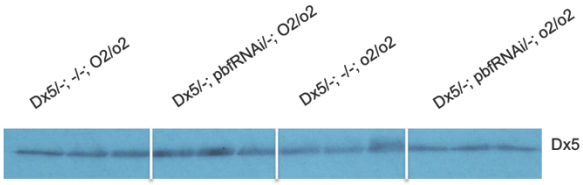


Figure 1-7. Western blot showing the accumulation of Dx5 protein in different *O2*/*Pbf* backgrounds

Discussion

The control of seed protein gene expression in cereals has been studied extensively due to their importance in determining the grain's nutritional and physical properties. Cereal species are highly conserved in terms of gene content and synteny, but their seed protein composition and properties can vary significantly. While most cereals have prolamins as their main seed storage proteins, oat and rice are exceptions in that they contain mostly globulins (Shewry, et al. 1995). Also, whereas the majority of both wheat and maize seed storage proteins are prolamins, maize does not contain the wheat prolamin variants (gluten proteins) that are needed for bread making. The characterization and cloning of many cereal seed protein genes with different properties opens up the prospect of engineering their seed protein compositions and expanding their utilization. Thus, studying conservation or divergence of genetic control is important for heterologous expression of these genes among cereals.

Previous gene regulation studies of *22-kDa α-zeins* showed that their tissue specificity and temporal expression are regulated epigenetically. This occurs at the DNA level through methylation, and at the chromatin level through histone methylation and acetylation marks (Locatelli, et al. 2009). The presence of individual maize chromosomes

in the OMA lines separates it from regulatory elements present in other chromosomes. In our RT-PCR experiments, we did not see any expression of *ZmO2*, *ZmPbf*, or any of the *zeins* in the leaf. However, the coordinated temporal expression of *ZmPbf* and the *27-kDa γ-zein* was well preserved. These results indicate that the epigenetic states of these genes were preserved, and that the oat host genome was able to correctly assign and maintain appropriate epigenetic marks on *ZmO2*, *ZmPbf*, and *27-kDa γ-zein*. On the other hand, failure of some *zeins* to be expressed may be due to either silencing by epigenetic modification, or the absence/incompatibility of the host regulators to trans-activate them.

PBF and O2 are the two major *zein* regulators in maize. Previous reports have provided evidence for the role of O2 in zein gene expression. In an *o2* mutant, both the 22-kDa α- and 15-kDa β-zein proteins were reduced (Vicente-Carbajosa, et al. 1997). We did not detect expression of either gene in the OMA lines where they are located, which can be due to two reasons. First, OMA 4.42 and OMA 6.01 (which contain the 22-kDa α- and 15-kDa β-zein genes respectively) do not contain its native *ZmO2* regulator because it is present on another OMA line (OMA 7.06). Second, in the absence of *ZmO2*, the *AsO2* was not able to trans-activate them. Whether this is due to failure of AsO2 to recognize the 22-kDa α- and 15-kDa β-zein promoters or the silencing of these genes through epigenetic modifications is unknown at this time. The *19-kDa α-zein* cluster on OMA 7.06 is also not expressed even when *ZmO2* is expressed in this OMA line. Interestingly, this cluster arose from one of the progenitor of maize before allotetraploidization, whereas the 19-kDa α-zein cluster on chromosome 1 and 4 arose from the other homoeologue (Xu and Messing 2008).

Previous studies have shown that PBF binds to the P-box motif (TGTAAAG) (Vicente-Carbajosa, et al. 1997). In *ZmPbfRNAi* transgenic lines, only the 22-kDa α- and 27-kDa γ-zeins are reduced but other zeins accumulate normally. This is surprising because analysis of zein promoters revealed that most have P-box motifs (Wu and Messing 2012). Whereas ZmPbf might not be necessary for expression of all zeins, it is clearly required for the expression of the *27-kDa γ-zeins* (Marzabal, et al. 2008). Thus, the expression of *γ-zeins* in OMA 7.06 even without ZmPBF points to a trans-acting action of AsPBF on maize *γ-zeins*. Our transient expression data is evidence that this might indeed be the case.

All PBFs have a Dof domain near the N-terminus, which is highly conserved across the cereals (Figure 1-3B). The high degree of sequence conservation in the position and alignment of the Dof domain might explain the ability of AsPBF to trans-activate the *γ-zeins* in OMA 7.06. This trans-activating property was also observed for ZmPBF, which has been shown to recognize and activate glutelin promoters in transient assay using rice endosperm tissue culture (Hwang, et al. 2004). Based on these evidences from oat and rice, it is very likely that TaPBF, HvPBF, or OsPBF can also bind other prolamins that contain P-box in their promoters.

Oat has an allohexaploid genome consisting of A, C, and D homoeologous chromosome sets. Species that underwent polyploidization often show bias in gene expression between homoeoloci. Studies in maize, an ancient tetraploid, have shown that gene expression can change between homoeologous copies, with one homoeolocus expressed significantly higher than the other (Schnable, et al. 2011; Jiang, et al. 2013). A recent genome-wide study in wheat has also shown that 73-76% of expressed homoeoloci

show significant gene expression difference, and that one of the three homoeoloci dominate the gene expression (Leach, et al. 2014). The divergence in the regulation of homoeoloci after polyploidization may explain the significant difference in gene expression between the two copies of *AsPbf* that we detected.

A previous study based on *Dx5* promoter segment deletions has shown that the G-box like motif is important and fundamental to its expression (Norre, et al. 2002). The deletion of this G-box like motif resulted to a significant reduction in promoter activity. On the other hand, the deletion of a known enhancer element (Thomas and Flavell 1990) that contains a P-box motif resulted in a non-significant reduction of promoter activity. The study also showed that stronger expression can be achieved when the G-box like motif is paired with the P-Box and if these two are repeated in tandem. The same study also found that the G-box like motif can bind at least two types of transcription factors – bZip factors like O2, as well as those belonging to the ASF-1 family – due to the similarity of their binding patterns in gel-shift assays. In the absence of any reduction in Dx5 protein in *PbfRNAi*, *o2*, and *PbfRNAi-o2* backgrounds, our data is in agreement with the finding that other transcription factors can bind to the G-box like motif. Therefore, there seems to be a genetic redundant mechanism that controls the expression of *Dx5* (and perhaps other G-box containing prolamin promoters) that enables them to be expressed in the absence of one of the *Pbfs* and *O2*.

Apparently, the redundancy of transcription factors in these species appears to be a prerequisite for recently amplified gene copies to acquire the capacity to switch to different regulators, as long as they achieve a suitable expression pattern. This is particularly evident from the group I prolamins. Even, when both PBF and O2 are

knocked down, there are still α–zeins being expressed. Moreover, there are even maize haplotypes of the z1C1 locus that can differ in copy number variation and their transcriptional activation by O2 despite the conservation of the *cis*-acting elements defined by DNA-protein binding studies and promoter truncation experiments. Interestingly, this change affects the two youngest, 3' terminal copies in the tandem gene clusters (Song, et al. 2001). Our studies in oats are consistent with this trend, where promoter regions are acquiring mosaics of elements, when copied.

References

Brautigam M, Lindlof A, Zakhrabekova S, Gharti-Chhetri G, Olsson B, Olsson O. 2005. Generation and analysis of 9792 EST sequences from cold acclimated oat, Avena sativa. BMC Plant Biol 5:18.

Cord-Neto G, Yunes JA, Silva MJd, Vettore AL, Arruda P, Leite A. 1995. The involvement of Opaque2 on beta-prolamin gene regulation in maize and coix suggests a more general role for this transcriptional activator. Plant Mol Biol 27:1015-1029.

Feng L, Zhu J, Wang G, Tang Y, Chen H, Jin W, Wang F, Mei B, Xu Z, Song R. 2009. Expressional profiling study revealed unique expressional patterns and dramatic expressional divergence of maize a-zein super gene family. Plant Mol Biol 69:649–659.

Gallavotti A, Malcomber S, Gaines C, Stanfield S, Whipple C, Kellogg E, Schmidt RJ. 2011. BARREN STALK FASTIGIATE1 is an AT-hook protein required for the formation of maize ears. Plant Cell 23:1756-1771.

Gaut B. 2002. Evolutionary dynamics of grass genomes. New Phytologist 154:15-28.

Herman EM, Larkins BA. 1999. Protein Storage Bodies and Vacuoles. Plant Cell 11:601-613.

Hwang Y, Ciceri P, Parsons R, Moose S, Schmidt R, Huang N. 2004. The Maize O2 and PBF Proteins Act Additively to Promote Transcription from Storage Protein Gene Promoters in Rice Endosperm Cells. Plant and Cell Physiology 45:1509–1518.

Jiang WK, Liu YL, Xia EH, Gao LZ. 2013. Prevalent role of gene features in determining evolutionary fates of whole-genome duplication duplicated genes in flowering plants. Plant Physiol 161:1844-1861.

Jones D, Taylor W, Thornton J. 1992. The rapid generation of mutation data matrices from protein sequences. Bioinformatics 8:275-282.

Kellogg EA. 2001. Evolutionary History of the Grasses. Plant Physiol 125:1198-1205.

Kynast RG, Riera-Lizarazu O, Vales MI, Okagaki RJ, Maquieira SB, Chen G, Ananiev EV, Odland WE, Russell CD, Stec AO, et al. 2001. A Complete Set of Maize Individual Chromosome Additions to the Oat Genome. Plant Physiol 125:1216-1227.

Leach L, Belfield E, Jiang C, Brown C, Mithani A, Harberd N. 2014. Patterns of homoeologous gene expression shown by RNA sequencing in hexaploid bread wheat. BMC Genomics 15:276.

Lending CR, Chesnut RS, Shaw KL, Larkins BA. 1989. Immunolocalization of avenin and globulin storage proteins in developing endosperm of Avena sativa L. . Planta 178:315-324.

Locatelli S, Piatti P, Motto M, Rossi V. 2009. Chromatin and DNA modifications in the Opaque2-mediated regulation of gene transcription during maize endosperm development. Plant Cell 21:1410-1427.

Marzabal P, Gas E, Fontanet P, Vicente-Carbajosa J, Torrent M, Ludevid MD. 2008. The maize Dof protein PBF activates transcription of gamma-zein during maize seed development. Plant Mol Biol 67:441-454.

Mena M, Vicente-Carbajosa J, Schmidt RJ, Carbonero P. 1998. An endosperm-specific DOF protein from barley, highly conserved in wheat, binds to and activates transcription from the prolamin-box of a native B-hordein promoter in barley endosperm. The Plant Journal 16:53-62.

Miclaus M, Xu J-H, Messing J. 2011. Differential Gene Expression and Epiregulation of Alpha Zein Gene Copies in Maize Haplotypes. PLoS Genet 7:e1002131.

Norre F, Peyrot C, Garcia C, Rance I, Drevet J, Theisen M, Gruber V. 2002. Powerful effect of an atypical bifactorial endosperm box from wheat HMWG-Dx5 promoter in maize endosperm. Plant Mol Biol 50.

Rines HW, Phillips RL, Kynast RG, Okagaki RJ, Galatowitsch MW, Huettl PA, Stec AO, Jacobs MS, Suresh J, Porter HL, et al. 2009. Addition of individual chromosomes of maize inbreds B73 and Mo17 to oat cultivars Starter and Sun II: maize chromosome retention, transmission, and plant phenotype. Theor Appl Genet 119:1255-1264.

Salse J, Feuillet C. 2007. Comparative Genomics of Cereals. In: Varshney RK, Tuberosa R, editors. Genomics-Assisted Crop Improvement. Dordrecht, The Netherlands: Springer. p. 177–205.

Sangtong V, Moran L, Chikwamba R, Wang K, Woodman-Clikeman W, Long J, Lee M, Scott P. 2002. Expression and inheritance of the wheat Glu-1DX5 gene in transgenic maize. Theor Appl Genet 105:937-945.

Schmidt RJ, Burr FA, Aukerman MJ, Burr B. 1990. Maize Regulatory Gene Opaque-2 Encodes a Protein With a Leucine-zipper Motif That Binds to Zein DNA. Proc Natl Acad Sci U S A 47:46-50.

Schmidt RJ, Ketudat M, Aukerman MJ, Hoschek G. 1992. Opaque2 Is A Transcriptional Activator That Recognizes A Specific Target Site in 22-kD Zein Genes. Plant Cell 4:689-700.

Schnable JC, Springer NM, Freeling M. 2011. Differentiation of the maize subgenomes by genome dominance and both ancient and ongoing gene loss. Proc Natl Acad Sci U S A 108:4069–4074.

Shewry PR, Halford NG. 2002. Cereal seed storage proteins: structures, properties and role in grain utilization. J Exp Bot 53:947–958.

Shewry PR, Napier JA, Tatham AS. 1995. Seed Storage Proteins: Structures and Biosynthesis. Plant Cell 7:945-956.

Song R, Llaca V, Linton E, Messing J. 2001. Sequence, regulation, and evolution of the maize 22-kD alpha zein gene family. Genome Research 11:1817-1825.

Song R, Messing J. 2003. Gene expression of a gene family based on non-collinear haplotypes. Proc Natl Acad Sci U S A 100:9055–9060.

Tamura K, Dudley J, Nei M, Kumar S. 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol Biol Evol 24:1596-1599.

Thomas MS, Flavell RB. 1990. ldentification of an Enhancer Element for the Endosperm-Specific Expression of High Molecular Weight Glutenin. Plant Cell 2:1171-1180.

Vicente-Carbajosa J, Moose SP, Parsons RL, Schmidt RJ. 1997. A maize zinc-finger protein binds the prolamin box in zein gene promoters and interacts with basic leucine zipper transcriptional activator Opaque2. Proc Natl Acad Sci U S A 94:7865-7690.

Wu Y, Messing J. 2012. Rapid divergence of prolamin gene promoters of maize after gene amplification and dispersal. Genetics 192:507-519.

Wu Y, Messing J. 2009. Tissue-specificity of storage protein genes has evolved with younger gene copies. Maydica 54:409-415.

Xu J, Messing J. 2008. Organization of the prolamin gene family provides insight into the evolution of the maize genome and gene duplications in grass species. Proc Natl Acad Sci U S A 105:14330–14335.

Xu JH, Messing J. 2009. Amplification of prolamin storage protein genes in different subfamilies of the Poaceae. Theor Appl Genet 119:1397-1412.

Zhang W, Ciclitira P, Messing J. 2014. PacBio sequencing of gene families - a case study with wheat gluten genes. Gene 533:541-546.

Zhang W, Sangtong V, Peterson J, Scott MP, Messing J. 2013. Divergent properties of prolamins in wheat and maize. Planta 237:1465-1473.

Chapter 2 A maize defective kernel mutant generated by insertion of a novel *Ds* element in a gene coding for a highly conserved TTI2 co-chaperone

Abstract

We have employed the newly engineered transposable element *Dsg* to tag a gene that gives rise to a defective kernel (*dek*) phenotype. *Dsg* requires the autonomous element *Ac* for transposition. Upon excision, it leaves a short DNA footprint that can create in-frame and frameshift insertions in coding sequences. Therefore, we could create alleles of the tagged gene that confirmed causation of the *dek* phenotype by the *Dsg* insertion. The mutation, designated *dek38-Dsg*, is embryonic lethal, has a defective basal endosperm transfer (BETL) layer, and results in a smaller seed with highly underdeveloped endosperm. The maize *dek38* gene encodes a TTI2 (Tel2-interacting protein 2) molecular co-chaperone. In yeast and mammals, TTI2 associates with two other co-chaperones TEL2 (Telomere maintenance 2) and TTI1 (Tel2-interacting protein 1) to form the triple T complex. Therefore, we cloned the maize *Tel2* and *Tti1* homologs and showed that TEL2 can interact with both TTI1 and TTI2 in yeast two-hybrid assays. The levels of Ataxia Telangiectasia Mutated (ATM) and Target of Rapamycin (TOR) proteins were severely reduced in the *dek38* mutant, consistent with its known function in regulating cellular levels of phosphatidylinositol 3-kinase-related kinases (PIKKs). *dek38-Dsg* also displays reduced pollen transmission, indicating TTI2's importance in male reproductive cell development.

Introduction

Maize has a long history of transposon genetics dating back to the discovery of transposable elements *Activator* (*Ac*) and *Dissociation* (*Ds*) by Barbara McClintock in the 1940s (McClintock 1984). Transposable elements comprise about 85% of the maize genome (Schnable, et al. 2009) and are mostly inert due to mutations and epigenetic silencing by DNA methylation and histone modifications (Almeida and Allshire 2005; Slotkin and Martienssen 2007). Several of them, however, like the *Ac*/*Ds* and *Mutator* families can be activated and used for forward or reverse genetics purposes (Brutnell 2002; Cowperthwaite, et al. 2002; McCarty, et al. 2005). The *Ac*/*Ds* system is the method of choice when performing targeted transposon tagging, but limitations such as lack of "*Ds*-launching pads" near the gene of interest, as well as identifying germinal excisions and tracking *Ds* can be a challenge. To address these major limitations, a new resource has been developed in maize utilizing an engineered *Ds* that can be tracked phenotypically for germinal excision and reinsertion through a change in seed color and green fluorescent protein (GFP) signal (Li, et al. 2013). The *Ds* element contains GFP (*Dsg*) and has the 5' and 3' end sequences that are recognized by *Ac* for transposition. As such, transpositions from these *Dsg* launching pads are also being made and new insertion sites being mapped to develop a collection of tagged sites for the maize genetics community (http://www.acdsinsertions.org/). The GFP tag makes the *Ds* element dominant, and therefore allows for the phenotypic selection of seeds either heterozygous or homozygous for the *Dsg* insertion.

We have used this system to identify recessive mutations that prevent normal seed development. Such a screen resulted in the identification of a defective kernel mutant

(*dek*), called *dek38-Dsg*. Using the unique *Dsg*-adjacent sequences, the corresponding gene was cloned. Based on information from studies in animal and yeast systems, its gene product could be characterized as a co-chaperone hypothesized to enable early seed development by controlling the action of phosphatidylinositol 3-kinase-related kinases (PIKKs) that regulate cellular signaling pathways related to gene expression, cellular growth, response to DNA damage, and nonsense-mediated mRNA decay (Abraham 2004). The protein, called Tel2-interacting protein 2 (TTI2), associates with Telomere maintenance 2 (TEL2) and Tel2-interacting protein 1 (TTI1) forming what is called the TTT complex (Hurov, et al. 2010). The TTT complex regulates the steady state levels of PIKKs, as well as their assembly into functional complexes in a heat shock protein 90 (HSP90)-dependent manner (Takai, et al. 2010). Here we use the *dek38-Dsg* mutation to report the first description of the TTT complex in plants. The mutation is lethal and results in an arrested embryo and a highly-underdeveloped endosperm. The mutation also reduces protein levels of two members of maize PIKKs called Target of Rapamycin (ZmTOR) and Ataxia Telangiectasia Mutated (ZmATM), suggesting a conserved function throughout eukaryotic species.

Materials and Methods

Plant Materials

dek38-Dsg was identified in 2013 from the available *Dsg* insertion collection after a screen of ears segregating for defective kernel phenotypes. It was previously referred to as *dek34-Dsg*. Seeds were deposited in the Maize Genetics Cooperative Stock Center, and can be obtained under the accession number tdsgR07A07. The *dek18* mutant used for

allelism test was obtained from the Maize Genetics Stock Center. The *wx-m7*(*Ac*) stock that was used as the transposase source for the reversion and *Dsg* excision footprint experiments is in a predominant W22 background.

Dsg excision footprint and reversion analysis

Revertants were isolated by first crossing *dek38-Dsg* to *wx-m7*(*Ac*). 500 GFP-fluorescent seeds from this cross were then planted in the field and self-pollinated. The ears from these plants were then evaluated for GFP and defective kernel segregation and ears that segregated for GFP but not defective kernels were identified. GFP kernels from these putative revertants were then planted and genotyped for two polymorphisms that distinguish the WT allele from *wx-m7*(*Ac*) and the *Dek38* revertant allele from *dek38-Dsg*. These individuals were also genotyped for *Dsg* presence/absence using PCR primers from Table 2-1.

For isolating *Dsg* footprint alleles, the *dek38-Dsg* mutant was also crossed to *wx-m7*(*Ac*). The F1 were then self-pollinated to produce F2 seeds. From these, 2,500 GFP and 2,500 non-GFP seeds were planted in the field. At the seedling stage, DNA was extracted from leaf samples bulked from 10 individuals and assayed for *Dsg* excision footprints by PCR with primers that amplify 100 bp regions flanking the *Dsg* insertion (Table 2-1). PCR products were electrophoresed in 3% Metaphor agarose (Lonza) to enable resolution of the excision footprint PCR band which appears as a slightly slower band above the wild type band. To identify the actual plants within the bulk that possessed the footprint, DNA from individual plants were extracted and genotyped for the presence of the footprint. Individual plants from the GFP population identified to

Table 2-1. PCR primers used in this study

Primer name	Sequence	Purpose
Telo2_5UTR_F	AAAGACTCGTCTATTCGAACTGTCC	full length CDS cloning
Telo2_3UTR_R	AGGACACATCCCTGCACAAC	
Tti1_5UTR_F	GTATTGGAAAGCAGGCAAGC	full length CDS cloning
Tti1_3UTR_R	GGGTGCTAGCATGACCAAAG	
Tti2_5UTR_F	ACCCCTTCAGTGCCTTGACT	full length CDS cloning
Tti2_3UTR_R	GTGACAGGATAGGCGATGGT	
Telo2_qPCR_F	TATTGTTTGCAGCGTCTTGC	qRT-PCR
Telo2_qPCR_R	TGATTTGCAGATTCCAGTGC	
Tti1_qPCR_F	TGATGTTCTTGCTTCCATGC	qRT-PCR
Tti1_qPCR_R	TCATCTGGCAAACAAGTGGA	
Tti2_qPCR_F	GATGCGTGCTGTCAGAACAT	qRT-PCR
Tti2_qPCR_R	GAGGCCCATAGCATCAAAAA	
Ubiquitin_qPCR_F	CGCACCCTAGCAGACTACAA	qRT-PCR
Ubiquitin_qPCR_R	TACGCACACACAACACAACC	
Dsg_left_F	TCTTCGCATTGTTGTTCCTG	Left Dsg insertion PCR
Dsg_left_R	TTCGCTCATGTGTTGAGCAT	
Dsg_right_F	CGGGATCATTCCGATTAAAA	Right Dsg insertion PCR
Dsg_right_R	ATTCGGTTAGGGGGATTGAG	
ZmMRP1_RT_F	GGCTTGTTCAGGGACAACAA	ZmMRP1 RT-PCR
ZmMRP1_RT_R	TCGTTGATGCTAAAGCGTTG	
Betl10_RT_F	CATGCATGAGCAACAAGGAC	Betl10 RT-PCR
Betl1_RT _R	AAATGGAGGACATGCCAAAC	
Pbf_RT_F1	TGCTGAGAGGAGGGATTTTT	Pbf RT-PCR
Pbf_RT_F1	GCTCTTCATGATGGCCTTG	
Telo2Sfi1A_F	AAGGCCGTCAAGGCCATGGCGAGCACGCGAAGC	YFP-fusion
Telo2Sfi1B_R	AAGGCCCATGAGGCCCATCATGTTGCCGAATGG	
Tti1Sfi1A_F	AAGGCCGTCAAGGCCATGGCGATGGAGGCGTCGGCG	YFP-fusion
Tti1Sfi1B_R	AAGGCCCATGAGGCCCACACCCTGCATCCTTCTGAA	
Tti2Sfi1A_F	AAGGCCGTCAAGGCCATGGCGAGCACGCGAAGC	YFP-fusion
Tti2Sfi1B_R	AAGGCCCATGAGGCCCATCATGTTGCCGAATGG	
Telo2Sfi1A_F	AAGGCCGTCAAGGCCATGGCGAGCACGCGAAGC	Yeast two-hybrid
Telo2Sfi1B_R_Stp	AAGGCCCATGAGGCCTCACATCATGTTGCCGAA	
Tti1Sfi1A_F	AAGGCCGTCAAGGCCATGGCGATGGAGGCGTCGGCG	Yeast two-hybrid
Tti1Sfi1B_R_Stp	AAGGCCCATGAGGCCTCACACACCCTGCATCCTTCT	
Tti2Sfi1A_F	AAGGCCGTCAAGGCCATGGCGAGCACGCGAAGC	Yeast two-hybrid
Tti2Sfi1B_R_Stp	AAGGCCCATGAGGCCTCACGGTGAACTGTTTTT	

contain footprints were then genotyped for the presence or absence of *Dsg* using PCR primers designed to amplify the insertion junctions (Table 2-1). Individual plants containing a footprint were then tagged, self-pollinated, and crossed to *dek38-Dsg* for allelism test.

Histological analysis

Developing kernels were fixed in 4% paraformaldehyde in PBS and processed according to a previously described protocol (Gallavotti, et al. 2010). Sections were then mounted on slides, stained with Toluidine blue, and documented using a Leica DM5500B light microscope system. A previously published method was also used for transmission electron microscopy of developing kernels (Wu, et al. 2010).

Transient expression in tobacco leaves

Maize Tel2, Tti1, and Tti2 C-terminal YFP fusions were created by first PCR amplifying the coding sequences (without stop codon) using primers with Sfi1A and Sfi1B restriction sites (Table 2-1) from pGEM-Teasy full length CDS clones of *Tel2*, *Tti1*, and *Tti2*, respectively. The PCR products were then directionally cloned into pENTRSfi-223.1 plasmid by standard cloning protocol. The pENTR vectors containing *Tel2*, *Tti1*, and *Tti2* were then Gateway-recombined into pEARLEYGATE101 (Earley, et al. 2006) using LR clonase. The resulting gene-YFP fusion constructs were then transformed into *Agrobacterium* and used for transient expression by leaf injection. Leaf punches were then imaged 3-5 days after injection in Leica SP8 confocal microscope using 514 nm excitation and 520-550 nm emission ranges.

Yeast two-hybrid

Maize *Tel2*, *Tti1*, and *Tti2* full length CDS were cloned into pENTRSfi-223.1 as described above. The pENTR-Tel2 construct was then Gateway-recombined into pDestDB (ZmTel2-DB) while pENTR-Tti1 and pENTR-Tti2 were Gateway-recombined into pDestAD (ZmTti1-AD and ZmTti2-AD). The DB and AD constructs were then transformed into yeast strain Y8930 and Y8800 respectively using the lithium acetate method (Consortium 2011). Auto-activation of pDestDB-Tel2 was then tested by growing in a medium lacking histidine with 1mM 3AT. Transformed DB and AD constructs were mated by plating on top of each other in YPDA medium overnight. Diploids (ZmTel2-DB/ZmTti1-AD and ZmTel2-DB/ZmTti2-AD) were then selected and grown on media lacking leucine, tryptophan, and histidine (-LTH) with 1mM 3AT for three days to test for reporter gene activation.

Western Blotting

To extract total buffer-soluble proteins, seed samples were powdered in liquid nitrogen and lysed in NETT buffer (20 mM Tris-HCl, 5 mM EDTA, 100 mM NaCl, 1 mM DTT, 0.5% Triton X) supplemented with 1 mM NaF, 1mM **β**-glycerophosphate, 1 mM Na_2VO_4 and cOmplete™ Ultra Tablet Mini protease inhibitor cocktail (Roche). Samples were then centrifuged at 15,000 g for 15 minutes at 4^0 Celsius, and the cleared supernatant was saved to another clean tube. To ensure equal loading, the protein extracts were quantified using the DC Protein Quantification kit (BioRad). Samples were then loaded in SDS-PAGE for electrophoresis and blotted on PVDF membrane (BioRad). The blots were then further processed according to membrane manufacturer's protocol. The

ZmATM antibody development was outsourced to GenScript (Piscataway, NJ, USA) by using the carboxyl terminal 200 amino acids as the antigen, the same part that was used for previous development of ATM antibody in *Arabidopsis* (Garcia, et al. 2003). *Maize* and *Arabidopsis* ATM proteins have 83% sequence identity on this part. The antibody against TOR is commercially available and was purchased from AgriSera (http://www.agrisera.com/en/artiklar/tor-target-of-rapamycin.html).

Gene expression analysis

Total RNA was extracted using Qiagen RNEasy Mini Kit, treated with DNAse (Invitrogen) and reverse-transcribed to cDNA with oligo-DT primers using qScript Supermix (Quanta Biosciences). The first strand cDNA was then used as template for qRT-PCR. The PerfeCTa SYBR Green Fastmix (Quanta Biosciences) was used for qPCR in an Illumina Eco Real Time PCR System. Three biological and technical replicates were used. Relative gene expression was calculated using the ΔΔCT method as implemented in the Illumina EcoStudy software using ubiquitin gene expression as a reference. The primers used are listed in Table 2-1.

Protein folding prediction

The translated protein sequences of ZmTTI2 (WT and with *Dsg* insertion) were submitted to Phyre (Kelley and Sternberg 2009) for *de novo* protein folding prediction. Although such methods do not substitute for crystal structure analysis of proteins, they still can provide information on protein truncations, as it is the case here. The top hit models were then visualized in the free visualization software RasMol.

Results

Verification and genetic characterization of the *dek38-Dsg* mutant

The *dek38-Dsg* mutant was isolated from a screen of a *Dsg* insertion collection for defective kernel (*dek*) phenotypes in a predominantly W22 genetic background (Figure 2-1). *Dsg*-adjacent sequences derived from sequence indexing of this collection indicated that the *Dsg* was inserted in exon 6 of gene model GRMZM2G048851 in *dek38-Dsg*, based on mapping sequences to the maize inbred B73 reference genome assembly version 3 (http://www.acdsinsertions.org). The *Dsg* insertion was verified by designing PCR primer pairs on the left and right sides of the insertion, with one primer designed from the gene model, and the other from the *Dsg* insertion (Table 2-1). PCR using these two primer pairs gave amplification on the GFP containing *Ds* elements, but not on non-GFP containing elements. The PCR products were also sequenced and showed that left and right insertion junctions contain an 8-bp direct repeat (Figure 2-2), which is characteristic of *Ac*/*Ds* transposon insertion sites (Pohlman, et al. 1984). RT-PCR experiments showed that the *Dsg* insertion does not affect transcription of GRMZM2G048851. If translated, the transcript is predicted to result in a shorter protein from a premature stop codon, containing amino acid residues from parts of the *Dsg* sequence. Protein-folding modeling also predicts a drastic change in the 3D conformation of the protein (Figure 2-2C).

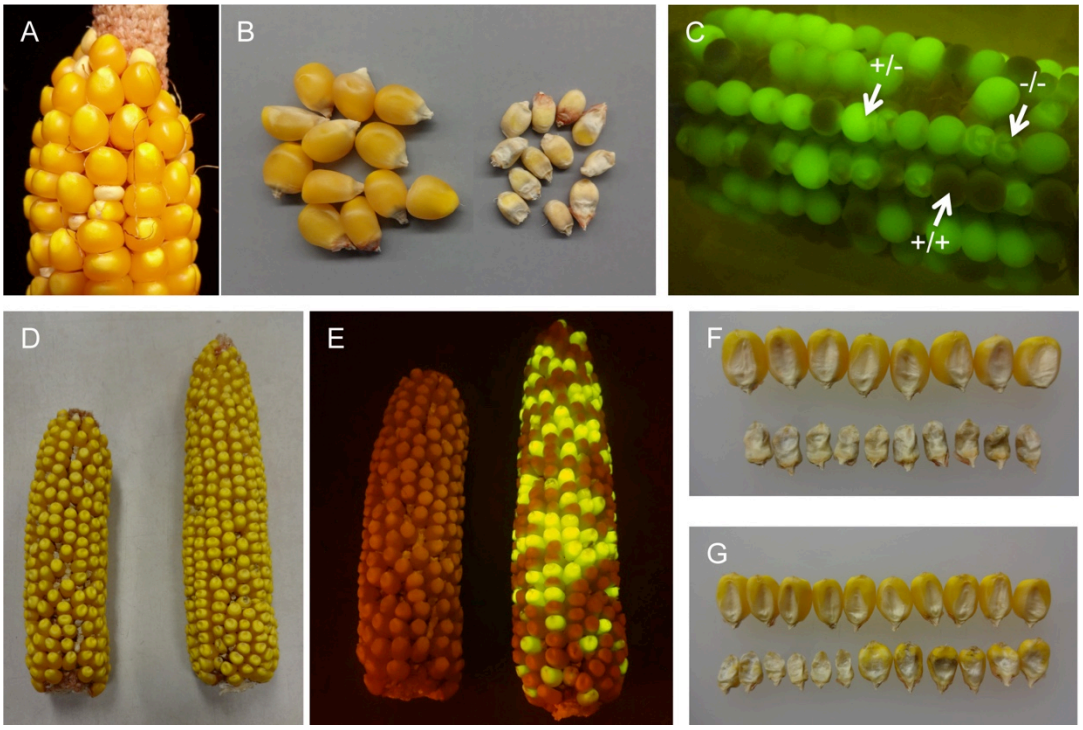


Figure 2-1. The *dek38-Dsg* mutant segregating in an ear (A, small kernels) and after shelling (B, right). When viewed under a blue light (C), segregation between normal non-GFP (+/+), normal GFP (+/-), and *dek* GFP (-/-) is visible. The change in color of normal kernels not fluorescing GFP is because of exposure to blue light in a dark background. (D) Left, ear from a *dek38* footprint allele with with a frameshift mutation. Right, ear from the allelism test between the same footprint allele and the *dek38-Dsg* parent. (E) Same ears as in D, but under blue light. The frameshift allele on left lost *Dsg* (and fluorescence), but the kernels on the right fluoresce because of the *Dsg* element in the *dek38-Dsg* mutant. Also shown are pictures of defective kernels (embryo side up) from the footprint allele (F) and from the allelism test (G). Normal kernels are shown on top for comparison.

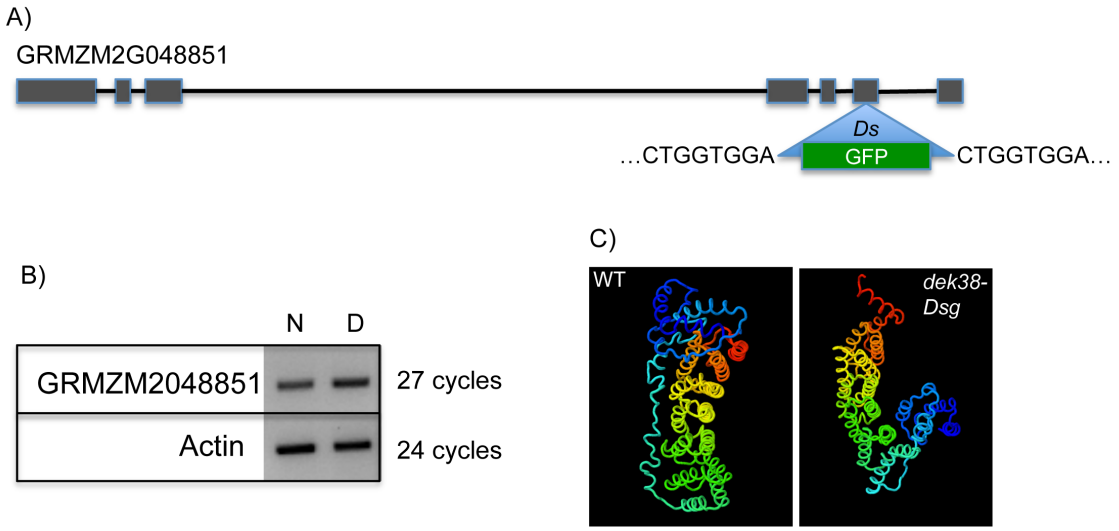


Figure 2-2. (A) The *Dsg* is inserted in exon 6 of gene model GRMZM2G048851 (*ZmTti2*). The 8-bp target site duplication (TSD) sequence is indicated on the left and right sides of the insertion. B) Gene expression of GRMZM2G048851 between WT and *dek* using mRNA extracted from immature seeds. C) Predicted protein folding of ZmTTI2 in WT and *dek38-Dsg* alleles

An advantage of having a *Dsg*-tagged gene is that it is possible to select heterozygotes among the normal seeds in the *dek38-Dsg* mutant by simple GFP selection. This made it easier to conduct a co-segregation analysis between the *dek* phenotype and the *Dsg*. Selfed progenies from 25 normal GFP seeds segregated for the *dek* phenotype. Conversely, none of the selfed progenies from 25 normal non-GFP seeds segregated for the *dek* phenotype. This indicates a tight linkage between GFP (and thus *Dsg*) and the *dek* phenotype (0/50 recombinants). Another advantage is that it is possible to monitor the transmission of the gene (either through male or female) by counting the number of GFP and non-GFP seeds after performing a cross. To determine if *dek38* is important for male or female reproductive cell development in maize, reciprocal crosses were conducted

using *dek38-Dsg* either as a male or female parent. Table 2-2 indicates that when *dek38-Dsg* is used as female, the segregation of GFP and non-GFP seeds falls along the 1:1 ratio that is expected. However, when used as a male, the expected 1:1 segregation is often distorted (based on chi-squared tests), indicating a defect in pollen transmission in the *dek38-Dsg* mutant. Based on the B73 genome reference V3, GRMZM2G048851 is located on the short arm of chromosome 5 close to another *dek* mutant (*dek18*). *dek18* (accession 527A from the Maize Genetics Stock Center) also has collapsed and floury kernels like *dek38-Dsg*. We tested whether these two *dek* mutants were allelic, but they were not.

Table 2-2. Pollen transmission defect in *dek38-Dsg*

	Observed		Chi-squared (p-value)
Cross	GFP(+)	GFP(-)	
dek38-Dsg x *wx-m7(Ac)*	157	177	0.27
dek38-Dsg x *wx-m7(Ac)*	92	84	0.55
dek38-Dsg x *wx-m7(Ac)*	106	131	0.10
dek38-Dsg x *wx-m7(Ac)*	211	218	0.74
dek38-Dsg x *wx-m7(Ac)*	213	233	0.34
wx-m7(*Ac*) x *dek38-Dsg*	91	140	0.00*
wx-m7(*Ac*) x *dek38-Dsg*	60	48	0.25
wx-m7(*Ac*) x *dek38-Dsg*	125	172	0.01*
wx-m7(*Ac*) x *dek38-Dsg*	152	205	0.01*
wx-m7(*Ac*) x *dek38-Dsg*	137	251	0.00*

Chi-squared – Chi-squared Test. Note the number GFP seeds are lower than expected (bold numbers, significant chi-squared values) when *dek38-Dsg* is used as male.

A third advantage of a marked *Ds* element is that it enabled us to select revertants from a lethal mutation. Revertants serve to confirm that the mutation in question is, in fact, tagged by *Ds*, but they are difficult to isolate from a lethal mutation (Ma and Dooner 2004). Putative revertants are isolated by planting large number of normal, green fluorescent individuals, selfing them, and screening for ears that do not segregate the *dek* mutation, but still segregate green fluorescent kernels. These are revertant candidates that presumably arose from a *Dsg* transposition that generated a normal *Dek* allele. Among 500 self-pollinated ears, three putative revertants segregated GFP, but not the *dek* phenotype (Figure 2-3A-B). Genotyping these three plants showed that the *Dsg* was no longer present at the *dek38* locus and that its excision had left no footprints, restoring the original wild type sequence. To distinguish this *dek38-Dsg* clean excision allele from the normal allele coming from the *wx-m7*(*Ac*) parent, GFP seeds from revertants were genotyped for two polymorphisms that mark the two parental alleles: a 6-bp insertion-deletion (indel) polymorphism and a SNP present in exon 1 of GRMZM2G048851 (Figure 2-3C). Sequencing several GFP seeds from each of the three putative revertant ears showed that, in fact, they were segregating both polymorphisms, indicating that the ears do not segregate the *dek* phenotype because the gene's function had been restored by excision of *Dsg*.

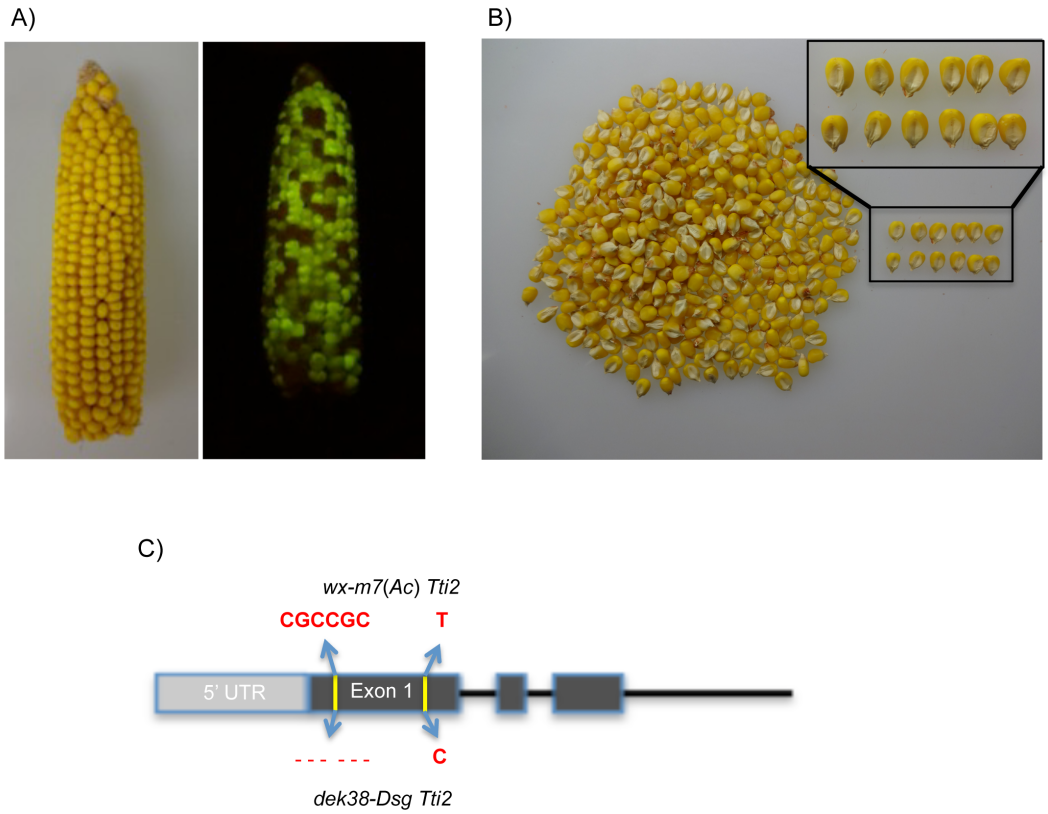


Figure 2-3. A) One of the identified revertants showing a completely normal ear (left). The same ear is shown on the right under blue light segregating for GFP. B) Kernels from (A) with a subset shown on the right. Inset shows a more close-up view of the normal kernels. C) Sequence polymorphisms found in exon 1 of GRMZM2G048851 which enabled differentiation between the WT allele contributed by the *wx-m7*(*Ac*) parent and the WT progenitor allele of *dek38-Dsg*.

Besides clean *Ds* excision, the addition of extra nucleotides (usually variations of the TSD sequence) can cause either frameshift or in-frame mutations. The existence of these footprint alleles was first identified by PCR screening of 5,000 F2 plants grown from 2,500 GFP and 2,500 non-GFP seeds from self-pollinated F1 plants from the cross

of *dek38-Dsg* to *wx-m7*(*Ac*). To manage the large population size, DNA were first extracted in bulks of 10 plants and genotyped for the presence of footprint using PCR primers that flank the insertion site (Table 2-1). Bulks that indicated the existence of a footprint by the presence of a larger PCR band above the wild type band on a 3% Metaphor agarose gel were then selected for individual plant genotyping. The identified individuals from the GFP population that possessed the footprint were then genotyped with another pair of PCR primers to detect the presence or absence of the *Dsg* insertion (Table 2-1). A total of 112 individuals (102 from the GFP population and 10 from the non-GFP population) with the presence of a footprint and absence of the *Dsg* insertion (germinal *Dsg* excisions) were then tagged for self-pollination, as well as cross-pollination to *dek38-Dsg* for allelism tests.

Thirty individuals that were identified by PCR screening described above as containing germinal excision footprints were sequenced. Out of these, we identified four different types of frameshift insertion footprints – two types with eight extra nucleotides (fp+8), and another two with seven extra nucleotides (fp+7) (Table 2-3). All the selfed ears from these individuals segregated *dek* kernels. Crosses with *dek38-Dsg* also resulted in ears segregating for the *dek* phenotype indicating that the *dek* mutations are allelic. A notable observation is the isolation of a *dek38* frameshift footprint allele from the non-GFP population, which is a result of excision and loss of *Dsg* that either segregated away during meiosis or did not reinsert into the genome (Figure 2-1D-E). Two in-frame mutation footprint alleles were also identified, one predicted to add an extra amino acid (fp+3) and another one adding two extra amino acids (fp+6) (Table 2-3). Examination of selfed ears from these individuals showed that a great majority of the kernels are normal,

but occasionally *dek* seeds can be found randomly on the ear. A possible explanation for this is that the addition of extra amino acids has only a small effect on the proper function of the protein (partial reversion) (Figure 2-4).

Table 2-3. *dek38* alleles from Dsg excision footprints

Footprint	Number of plants	Consequence
CTGGTGGA GTGGTGGA CTGGTGGA <Dsg> CTGGTGGA	12	Frameshift
CTGGTGGT CTGGTGGA CTGGTGGA <Dsg> CTGGTGGA	4	Frameshift
CTGGTGG GTGGTGGA CTGGTGGA <Dsg> CTGGTGGA	1	Frameshift
CTGGTGGT CGGTGGA CTGGTGGA <Dsg> CTGGTGGA	1	Frameshift
CTGGTGG TGGA CTGGTGGA <Dsg> CTGGTGGA	5	Inframe (+1 amino acid)
CTGGTGGT GGTGGA CTGGTGGA <Dsg> CTGGTGGA	7	Inframe (+2 amino acids)

Note: The footprint sequences are in red. The target site duplication (TSD) sequence is indicated below the footprint sequence for comparison.

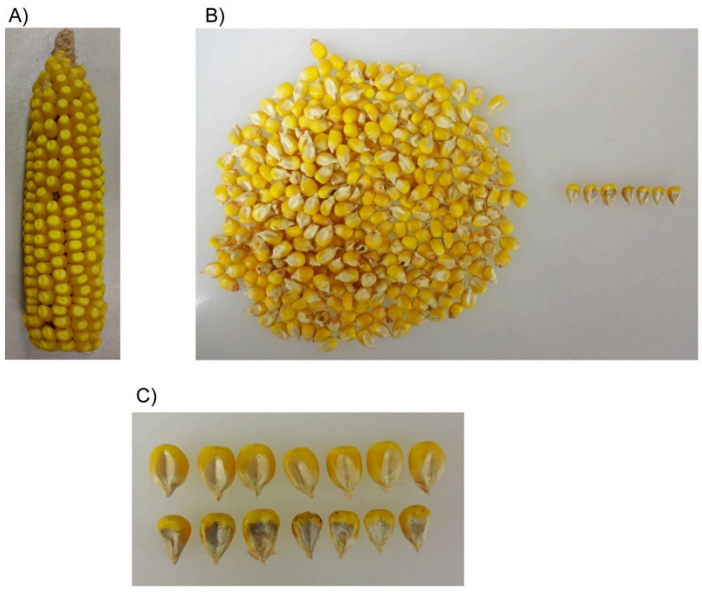


Figure 2-4. A) An ear from one of the identified partial revertants. B) Kernels from (A) showing mostly normal kernels (left) and seven *dek* kernels (right). C) Close up of the seven *dek* kernels compared to the wild type.

dek38-Dsg has impaired seed development

Developing ears were sampled at different days after pollination to determine the progression of the *dek* phenotype. As early as 8 DAP, *dek* kernels can already be distinguished from WT kernels because of their smaller size (Figure 2-5A). At around 12 DAP, the *dek* seeds begin to cave in at the crown due to incomplete seed filling (Figure 2-5B). Proximate analysis of mature seeds reveals a lower crude protein and oil concentration, and a higher carbohydrate concentration in the *dek* seeds (Table 2-4). The lower protein concentration is reflected in the smaller protein bodies in the *dek* endosperm compared to normal (Figure 2-6A). The increase in carbohydrate concentration in *dek* kernels is consistent with the unchanged size of the starch granules

and the smaller overall seed size. Zein storage proteins, which make up approximately 50% of the total seed protein (Lending and Larkins 1989), are also greatly reduced in the *dek* seeds. The 27-kDa gamma-zein has the largest reduction among different zeins proteins, with no detectable protein in the SDS PAGE gel (Figure 2-6B). Indeed, its regulator, the transcription factor *Prolamin box binding factor 1* (*Pbf1*), whose function is conserved across cereals (Garcia, et al. 2015), is also reduced in the *dek* seeds (Figure 2-6C).

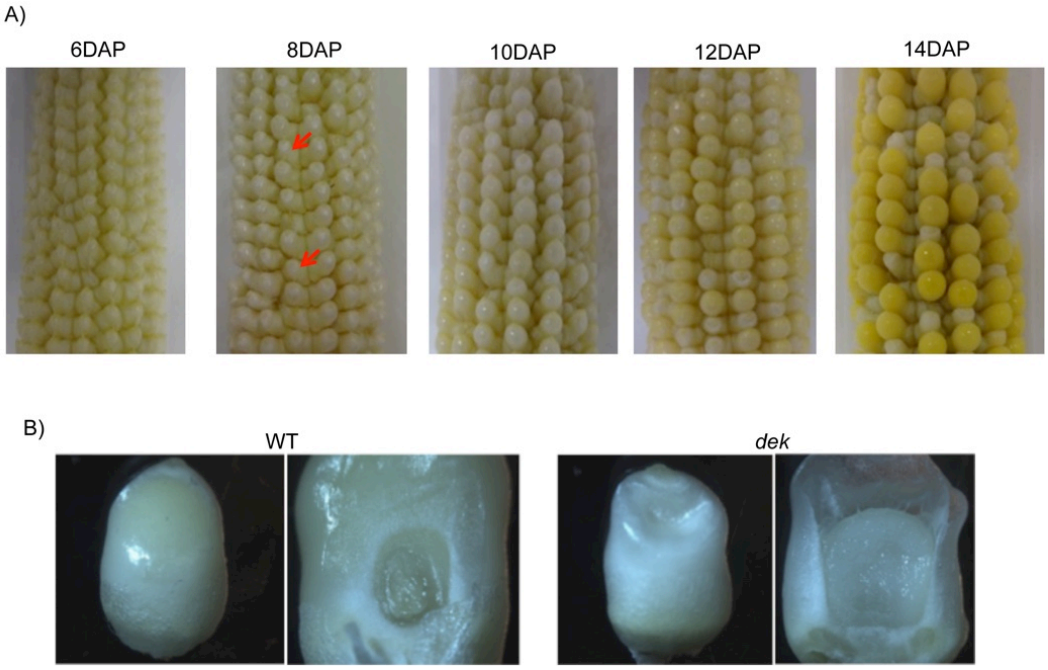


Figure 2-5. A) Heterozygote *dek38-Dsg* ears at different stages of development. Arrows indicate the smaller *dek* kernels. B) 12DAP WT and *dek* kernels

Table 2-4. Proximate analysis of *dek* seeds compared to WT

	WT (%)	*dek* (%)
Protein (crude)	11.12	7.84
Fat (crude)	3.95	1.05
Carbohydrates	72.56	79.99
Moisture	10.60	9.35
Ash	1.56	1.77

*Proximate analysis was done by New Jersey Food Labs Inc (http://www.njfl.com)

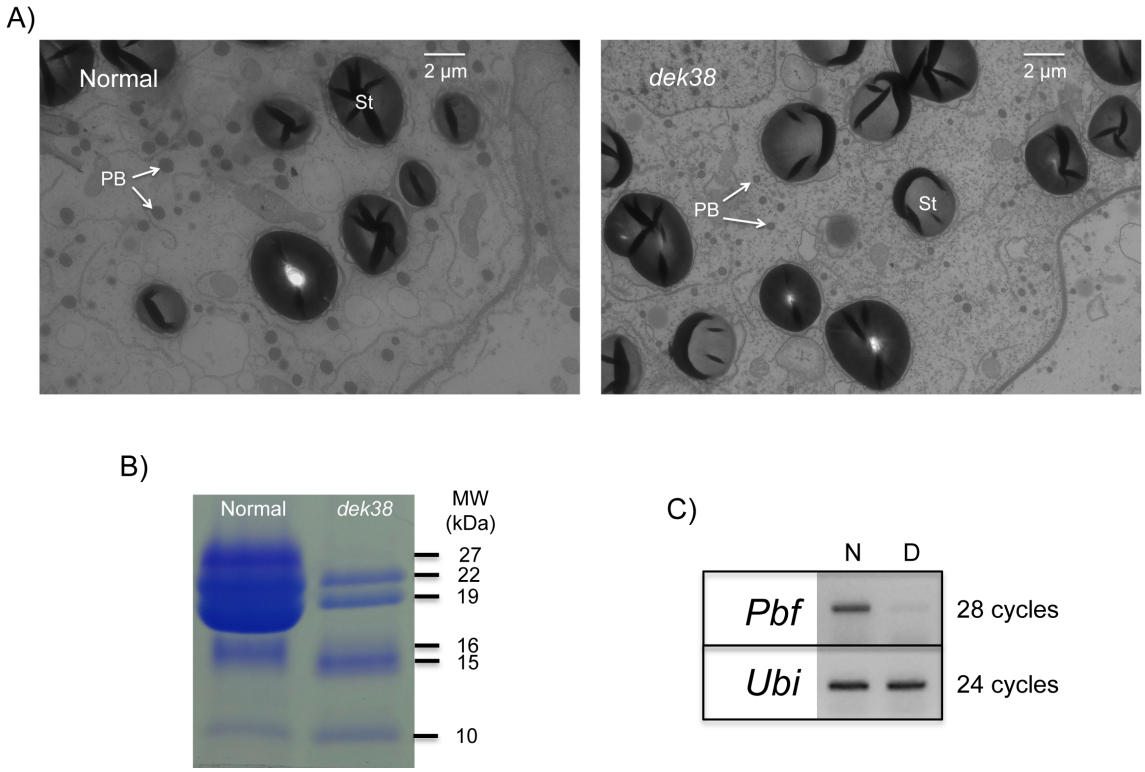


Figure 2-6. A) Transmission electron micrograph of a normal (left) and *dek* (right) seed from the *dek38-Dsg* mutant (PB – protein body, St – starch granule). B) Zein protein profile of a *dek* seed from the *dek38-Dsg* mutant compared to a normal kernel (MW – protein molecular weight in kilodaltons). C) Gene expression (RT-PCR) of *Pbf* gene (N – normal, D – *dek*)

The development of the embryo and the endosperm was also studied through histological sections of seeds collected at different days after pollination. Dramatic differences between normal and mutant embryos are seen in Figure 2-7. First is the arrest in the development of the embryo, which stopped growing at the dermatogen stage indicating failure of further cell division and differentiation. Second are the smaller cells in the endosperm of *dek* kernels, which did not seem to expand as much as the normal kernels (Figure 2-8D). Third is the conspicuous absence of a fully-developed basal endosperm transfer layer (BETL), a darkly-staining cell layer at the bottom of the developing endosperm which is specialized for nutrient transport (Figure 2-8A-B). The aberrant BETL formation is also reflected by the low gene expression of BETL-specific genes in maize (Figure 2-8C). Also noticeable is the formation of a gap between the developing endosperm and the placento-chalazal cells (PC). Gap formation is visible at around 8 DAP and continues until the endosperm is fully dissociated from the PC layer at around 10 DAP (Figure 2-8B).

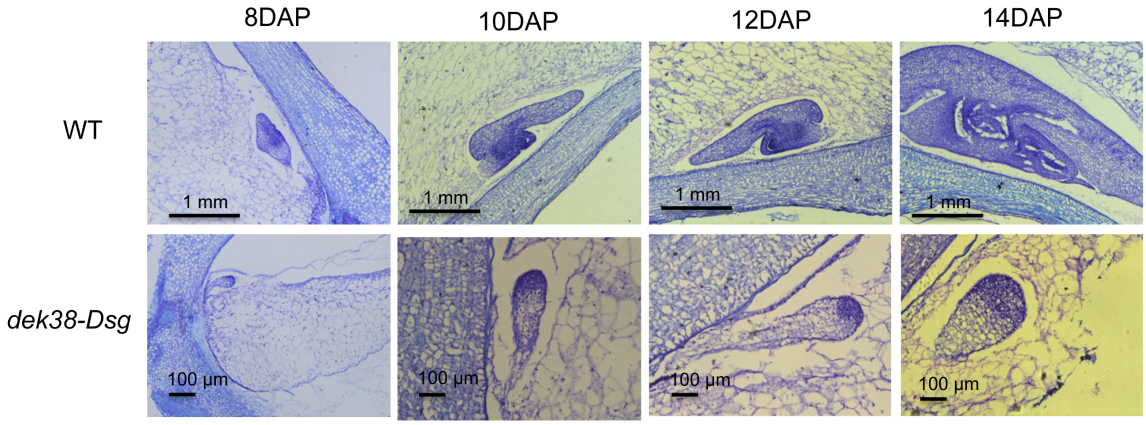


Figure 2-7. Development of the embryo at different days after pollination. Top Panel – Wild type. Bottom Panel – *dek38-Dsg*

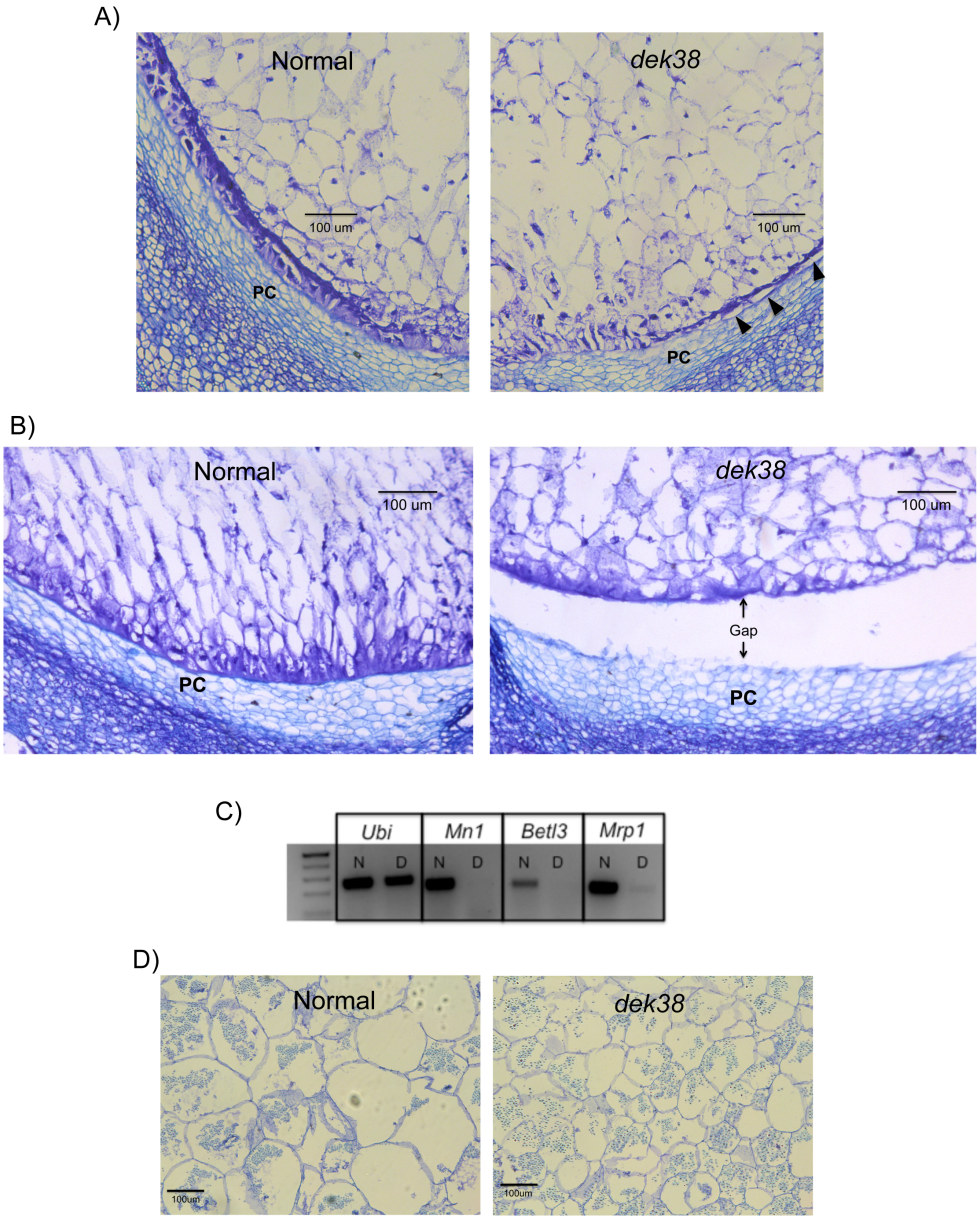


Figure 2-8. Defects in endosperm development of *dek38-Dsg*. A) 8DAP endosperm sections showing underdeveloped BETL in *dek* mutant. Arrows indicate sections that appear to be separating from the base of the seed. PC – placento-chalazal layer B) 10DAP *dek* endosperm showing detachment from the base of the seed. C) RT-PCR of BETL specific genes (N – normal sample, D – *dek* sample). D) Cell size comparison in endosperm cells between a normal and *dek* seed

dek38-Dsg is homologous to yeast and mammalian *Tti2*

The entire *dek38-Dsg* genic region is predicted to be 9,127 bp long and corresponds to the gene model GRMZM2G048851 based on current annotation (Version 3). Despite its large size, the predicted transcript is only 1,263 bp long which is divided into 7 exons. The gene contains a large LTR retrotransposon insertion in the third intron (more than 6 kb), which is also present in the W22 allelic sequence obtained from PacBio sequencing (Dong, et al. 2016). Despite this intron expansion, the gene is expressed in many tissues and at different developmental stages based on available gene expression data at MaizeGDB (http://www.maizegdb.org). The gene annotation indicates that the predicted GRMZM2G048851 protein product contains armadillo repeats, but provides no functional data. Using primers designed from the predicted UTRs (Table 2-1), the full length CDS of GRMZM2G048851 was cloned from mRNA extracted from two-week old maize seedlings. From this, an ORF of 1,269 bp in length was predicted to encode a protein of 422 amino acids. A BLASTP in UniProt (http://www.uniprot.org/) showed that GRMZM2G048851 has similarity to a protein in yeast and animals termed Tel2-interacting protein 2 (TTI2). Therefore, GRMZM2G048851 was designated as a putative *Tti2* (*ZmTti2*) homolog. The ZmTTI2 protein shares 23.4% identity with its human counterpart, with the highest similarity concentrated at the TTI2 domain, which overlaps with the armadillo domain. BLAST analysis of ZmTTI2 in maize B73 reference and the W22 PacBio genome sequences (Dong, et al. 2016) indicates that it is a single copy gene.

TEL2 and TTI1 are also present in maize

In animals and yeast, TTI2 partners with TEL2 and TTI1 to form the TTT complex (Hayashi, et al. 2007; Hurov, et al. 2010). Therefore, a search for TEL2 and TTI1 homologs in maize was conducted. Human TEL2 and TTI1 protein sequences were used to identify their homologous counterparts in maize via BLASTP. Both BLAST searches resulted in single hits, indicating that *Tel2* and *Tti1* are also single copy genes in maize. The putative *Tel2* homolog corresponds to gene model GRMZM2G144166 (*ZmTel2*) whereas the putative *Tti1* homolog corresponds to GRMZM2G056403 (*ZmTti1*) in maize. PCR primers were designed from the annotated 5' and 3' UTR regions of these genes to clone the full length CDS (Table 2-1). For *ZmTel2*, a PCR band about 3 kb long was amplified, cloned, and sequenced. An ORF of 3,051 bp was identified and predicted to encode a protein 1,016 amino acids long. Sequence analysis of this predicted protein product at InterPro (http://www.ebi.ac.uk/interpro) shows the conserved TEL2 domain at amino acid residues 634-745. The TEL2 domain overlaps with an armadillo repeat, which is then flanked by two more armadillo repeats. It also shares 24.4% protein identity with human TEL2 protein. For *Tti1*, a 4 kb PCR product was amplified, cloned and sequenced. This clone contains a 4,035 bp-long ORF that encodes a protein that is 1,344 amino acids long. Similar to ZmTEL2 and ZmTTI2, the predicted ZmTTI1 protein contains armadillo repeats based on analysis with InterPro. Both ZmTEL2 and ZmTTI1 have very similar structures like their respective human homologs.

Functional characterization of the maize TTT complex

The interaction of maize TEL2, TTI1, and TTI2 was tested via yeast two-hybrid. As seen in Figure 2-9, both ZmTel2-DB/ZmTti1-AD and ZmTel2-DB/ZmTti2-AD yeast diploids survived in –LTH triple dropout media, indicating ZmTEL2's interaction with both ZmTTI1 and ZmTTI2. Gene expression of *ZmTel2*, *ZmTti1*, and *ZmTti2* in many different maize tissues was detected using qPCR. Expression is particularly strong in young/developing tissues like two-week old shoots and roots, as well as developing tassels and ears (Figure 2-10). In contrast, gene expression was weaker in mature leaf, and almost no expression was detected in mature pollen. ZmTel2, ZmTti1, and ZmTti2 were fused to YFP and transiently expressed in tobacco leaves to determine their subcellular localization. Microscopic examinations of the leaves show YFP signals in both cytoplasm and nucleus for all the three TTT complex members (Figure 2-11). Our results are consistent with cytoplasmic and nuclear localizations of TEL2 in humans (Horejsi, et al. 2010) and TTI2 in yeast (Genereaux, et al. 2012). In yeast and mammals, the TTT complex is required for the stability of PIKK proteins as well as their assembly into complexes. Western blots with proteins extracted from developing mutant and wild type seeds show that the mutant has no detectable levels of ZmATM and ZmTOR proteins (Figure 2-12).

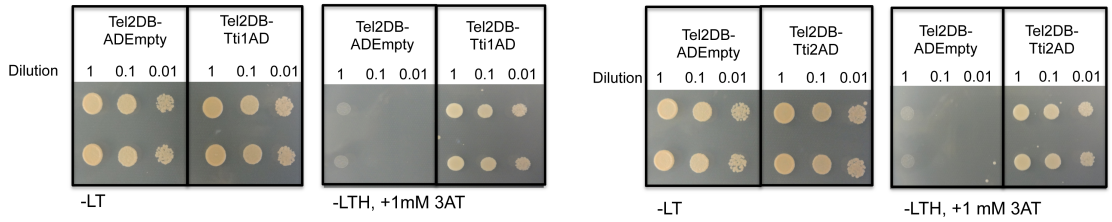


Figure 2-9. Yeast two-hybrid between ZmTel2 and ZmTti1 (A) and between ZmTel2 and ZmTti2 (B). Growth in –LTH media indicates interaction.

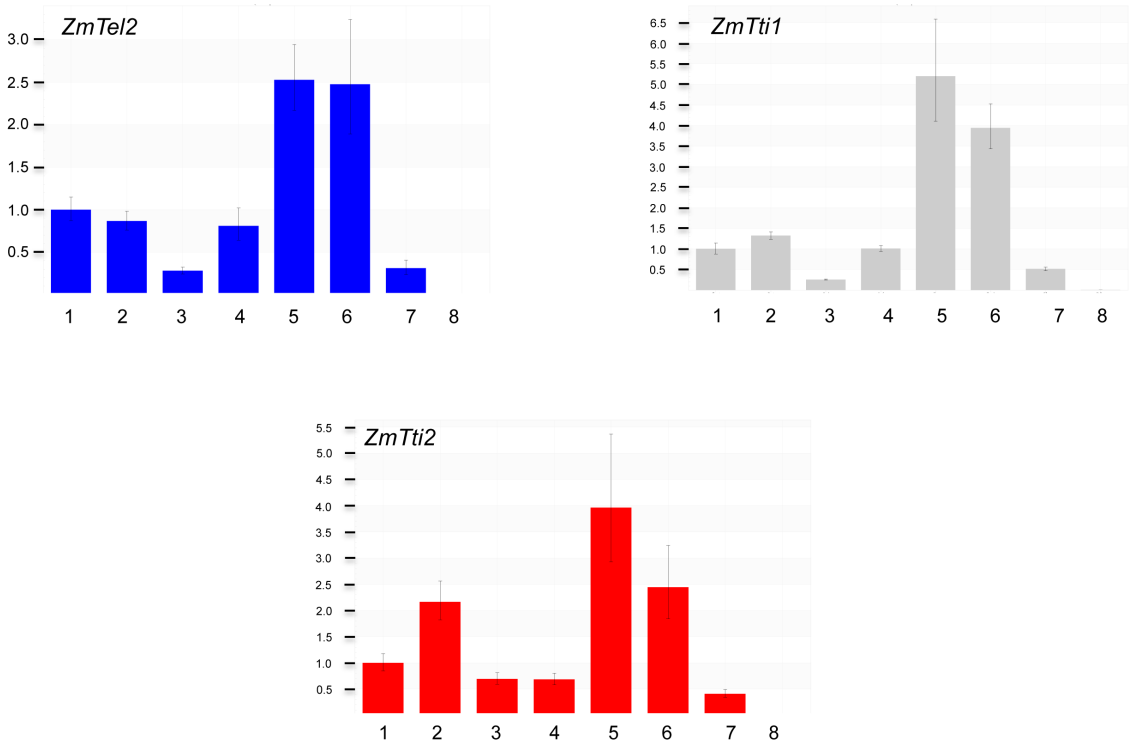


Figure 2-10. Relative gene expression of *ZmTel2*, *ZmTti1*, and *ZmTti2* in different maize tissues using ubiquitin as a reference. 1) 2-week old shoot, 2) 2-week old root, 3) mature leaf, 4) 14DAP seed, 5) immature (1cm) ear, 6) immature (1 cm) tassel, 7) silk, and 8) pollen

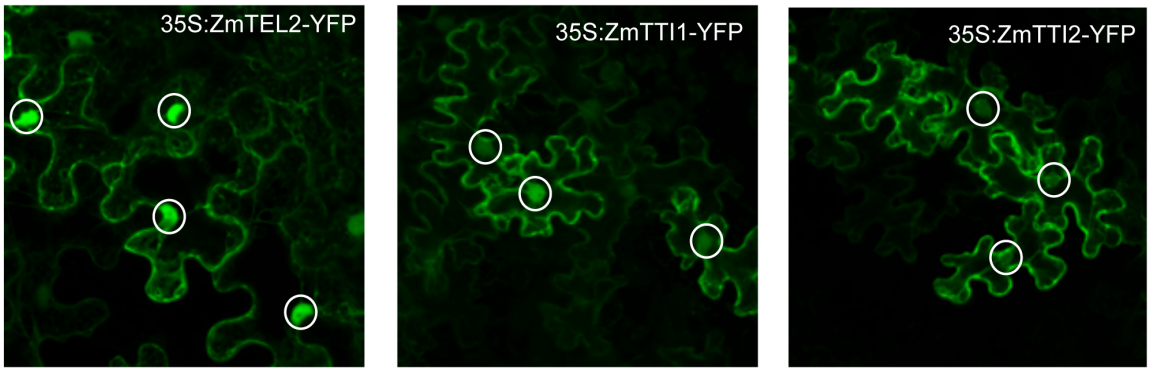


Figure 2-11. Cellular localization of ZmTel2, ZmTti1, and ZmTti2 proteins by transient expression of YFP fusion constructs in tobacco leaves. Circles indicate YFP signal in the nucleus.

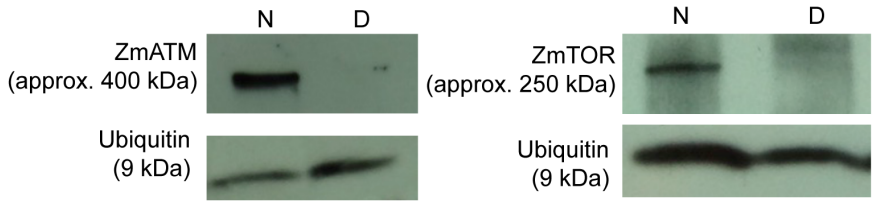


Figure 2-12. Western Blots comparing ZmATM and ZmTOR protein levels between normal (N) and *dek* (D) kernels

Discussion

Mutant resources are very important for functional genomics. However, crop species are still notoriously hard to transform, creating an impediment towards deployment of gene editing technologies (Altpeter, et al. 2016). While progress in transformation has been made in some monocots (Lowe, et al. 2016), transformation still necessitates specialized stocks and constructs and entails tissue culture capabilities. The use of transposons in creating mutants in maize only needs a greenhouse, field space, and

manual pollinations, and therefore can be an alternative for research groups without transformation capabilities. Also, as shown in our work, *Dsg* tagging has several benefits. First, it became easier to track the kernels that contain the *Dsg* element and identify their zygosity just by selecting for GFP fluorescence. Second, since GFP is dominant, linking the phenotype to the *Dsg* insertion in a lethal mutation was made easier by conducting a trait and GFP co-segregation test made possible by the phenotypic identification of seeds heterozygous for the *Dsg* insertion. Third, *Dsg* footprint sequences that can vary from different individuals provide multiple alleles that can have different phenotypic consequences, while clean excision events provide a clear "cause and effect" verification of the mutation. In the case of lethal mutations like the one described here, the GFP fluorescence from the *Dsg* enables the isolation of revertants. Some alleles that can be derived from footprint sequences have beneficial consequences such as the increased seed weight observed in a maize *shrunken2* allele with two amino acid residues added as a result of a *Ds* footprint (Giroux, et al. 1996).

The *dek38-Dsg* which encodes a co-chaperone is the first mutant to be characterized from the *Dsg* insertion collection, and the first member of the TTT complex to be characterized in plants. Due to their important roles in abiotic stress tolerance and growth regulation in response to changing environmental cues, molecular chaperones have been widely studied in plants. Most of the studies, however, have focused almost exclusively on heat shock proteins (HSPs), and little is known about their associated co-chaperones. Current evidence from yeast and mammals suggests that the TTT complex associates with chaperones and various other co-chaperones to regulate cellular levels of PIKK proteins. Studies in yeast and animals showing that TEL2 binds preferentially with

newly synthesized PIKKs, that TEL2, TTI1, and TTI2 are required to stabilize cellular PIKK levels, and that PIKK complexes are disrupted when TEL2 or HSP90 are depleted, all point to a co-chaperone activity of the TTT complex (Horejsi, et al. 2010; Hurov, et al. 2010; Kaizuka, et al. 2010; Takai, et al. 2010).

The defects in many aspects of seed development reflect the multitude of developmental pathways that are affected by the *dek38* mutation, indicating that TTI2 is required in early stages of development. *ZmTti2's* roles in storage protein accumulation and BETL development should be investigated further. Nevertheless, the limited nutrient transport predicted to result from a defective BETL could certainly play a role in the abnormal seed development of *dek38*. However, this alone cannot account for the embryo lethal phenotype as other mutants with defective BETL in maize like *miniature seed 1* and *zmsweet4c* have well-differentiated and viable embryos (Cheng, et al. 1996; Sosso, et al. 2015). Interestingly, the early embryonic lethality observed in *dek38-Dsg* resembles the *TOR* mutation in *Arabidopsis*. Like *dek38-Dsg*, the embryonic development of the *AtTOR* mutant is arrested at the dermatogen stage (known as transition stage in maize) which is characterized by the formation of the protoderm (Menand, et al. 2002). This implies that the embryonic lethality observed in *dek38-Dsg* might be due to aberrant ZmTOR function, which is supported by reduction of ZmTOR protein. The lethal phenotype in *dek38-Dsg* also implies that the C-terminal end of TTI2, where the *dek38* mutations are located, is important for protein function. This is supported by the protein sequence alignment of ZmTTI2 with its putative plant homologs showing high degree of conservation at several amino acid residues at the C-terminal end (Figure 2-13). In addition, the *dek38-Dsg* revertants that were isolated in our screen all show complete

restoration of the WT sequence. This is also consistent with findings in yeast wherein C-terminal end truncations of TTI2 are also lethal (Hoffman, et al. 2016).

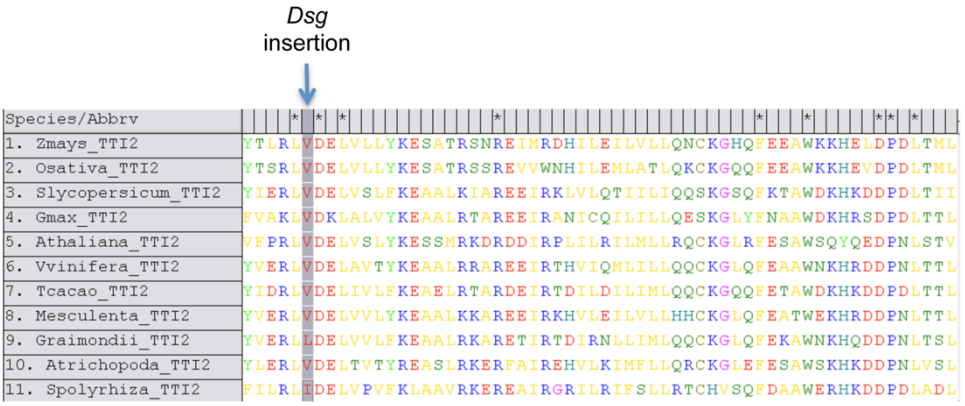


Figure 2-13. Multiple sequence alignment of maize TTI2 protein with other putative plant TTI2 proteins showing high degree of conservation at the C-terminal end. The *Dsg* insertion point is indicated. The accession numbers of the sequences are 1) *Z. mays* – GRMZM2G048851; 2) *O. sativa* – LOC_Os06g47900; 3) *S. lycopersicum* – Solyc06g009690; 4) *G. max* – Glyma.18G031000; 5) *A. thaliana* – AT2G39910; 6) *V. vinifera* – GSVIVG01016380001; 7) *T. cacao* – Thecc1EG042614; 8) *M. esculenta* – Manes.10G018500; 9) *G. raimondii* – Gorai.010G199300; 10) *A. trichopoda* – evm_27.TU.AmTr; 11) *S. polyrhiza* – Spipo12G0055600

The abnormal segregation of GFP and non-GFP seeds from the expected 1:1 ratio when *dek38-Dsg* is used as a pollen donor indicates *ZmTti2* is important for male reproductive cell development. This is likely due to ZmTti2's role in stabilizing the ATM protein. In humans, the *ATM* gene is expressed four times higher in testes than in somatic cells (Galetzka, et al. 2007) and mutations in mice result either in embryonic lethality

(Yamamoto, et al. 2012) or complete infertility (Barlow, et al. 1998). In *Arabidopsis*, ATM mutants are only partially sterile, although defects in meiosis have also been observed in both male and female reproductive cell development (Garcia, et al. 2003). If the pollen transmission defect in *dek38-Dsg* is actually due to reduced ATM function caused by the *ZmTti2* mutation, then ATM may not be as critical for female reproductive cell development in maize compared to *Arabidopsis* and mammals. It is also possible that ZmTTI2 functions independently of ATM in reproductive cell development.

The conserved interaction of TEL2, TTI1, and TTI2 in yeast, mice, humans, and now maize, provides strong evidence of the importance of this complex in organismal development. In a large-scale mutagenesis experiment, the homolog of *Tel2* in *Arabidopsis* has been identified as one of the many genes that affect embryo development and has been named *embryo defective 2423*, which corresponds to gene model AT3G48470 (Tzafrir, et al. 2004). The mutant, however, remains uncharacterized. No *Tti1* mutants have been reported in the literature outside of yeast, but it is very likely that null mutants will also be lethal. A recent finding that a "chaperome" network which includes TTT complex members facilitates tumor survival in humans (Rodina, et al. 2016) raises the possibility of their potential use to genetically engineer plants with enhanced stress tolerance. Our analysis now provides a foundation for new lines of investigation in maize in terms of embryo and endosperm development, BETL formation, and DNA damage repair among others. Because additional mutants in TTT function are likely to be lethal, weaker mutant alleles or conditional mutant alleles would be preferred. The partial *dek38* revertant alleles isolated in this study can be utilized in the future to study other aspects of TTT complex function that were not possible with lethal alleles,

such as DNA damage response. In the case of PIKKs (or other proteins with multiple domains), *Dsg* can be very useful for targeting specific domains of these proteins for mutagenesis via intragenic transposition.

References

Abraham RT 2004. PI 3-kinase related kinases: 'big' players in stress-induced signaling pathways. DNA Repair (Amst) 3: 883-887.

Almeida R, Allshire RC 2005. RNA silencing and genome regulation. Trends Cell Biol 15: 251-258.

Altpeter F, et al. 2016. Advancing crop transformation in the era of genome editing. Plant Cell 28: 1510-1520.

Barlow C, et al. 1998. Atm deficiency results in severe meiotic disruption as early as leptonema of prophase I. Development 125: 4007-4017.

Brutnell TP 2002. Transposon tagging in maize. Funct Integr Genomics 2: 4-12.

Cheng WH, Taliercio EW, Chourey PS 1996. The miniaturel seed locus of maize encodes a cell wall invertase required for normal development of endosperm and maternal cells in the pedicel. Plant Cell 8: 971-983.

Consortium AIM 2011. Evidence for network evolution in an Arabidopsis interactome map. Science 333: 601-607.

Cowperthwaite M, et al. 2002. Use of the transposon Ac as a gene-searching engine in the maize genome. Plant Cell 14: 713-726.

Dong J, et al. 2016. Analysis of tandem gene copies in maize chromosomal regions reconstructed from long sequence reads. Proc Natl Acad Sci U S A 113: 7949–7956.

Earley KW, et al. 2006. Gateway-compatible vectors for plant functional genomics and proteomics. The Plant Journal 45: 616-629.

Galetzka D, et al. 2007. Expression of somatic DNA repair genes in human testes. J Cell Biochem 100: 1232-1239.

Gallavotti A, et al. 2010. The control of axillary meristem fate in the maize ramosa pathway. Development 137: 2849-2856.

Garcia N, Zhang W, Wu Y, Messing J 2015. Evolution of gene expression after gene amplification. Genome biology and evolution 7: 1303-1312.

Garcia V, et al. 2003. AtATM Is essential for meiosis and the somatic response to DNA damage in plants. Plant Cell 15: 119-132.

Genereaux J, et al. 2012. Genetic evidence links the ASTRA protein chaperone component Tti2 to the SAGA transcription factor Tra1. Genetics 191: 765–780.

Giroux MJ, et al. 1996. A single gene mutation that increases maize seed weight. Proc Natl Acad Sci U S A 93: 5824-5829.

Hayashi T, et al. 2007. Rapamycin sensitivity of the Schizosaccharomyces pombe tor2 mutant and organization of two highly phosphorylated TOR complexes by specific and common subunits. Genes Cells 12: 1357-1370.

Hoffman KS, et al. 2016. Saccharomyces cerevisiae Tti2 regulates PIKK proteins and stress response. G3 (Bethesda) 6: 1649-1659.

Horejsi Z, et al. 2010. CK2 phospho-dependent binding of R2TP complex to TEL2 is essential for mTOR and SMG1 stability. Molecular Cell 39: 839-850.

Hurov KE, Cotta-Ramusino C, Elledge SJ 2010. A genetic screen identifies the Triple T complex required for DNA damage signaling and ATM and ATR stability. Genes Dev 24: 1939-1950.

Kaizuka T, et al. 2010. Tti1 and Tel2 are critical factors in mammalian target of rapamycin complex assembly. J Biol Chem 285: 20109-20116.

Kelley LA, Sternberg MJ 2009. Protein structure prediction on the web: a case study using the Phyre server. Nat Protoc 4: 363-371.

Lending CR, Larkins BA 1989. Changes in the zein composition of protein bodies during maize endosperm development. Plant Cell 1: 1011-1023.

Li Y, Segal G, Wang Q, Dooner HK. 2013. Gene tagging with engineered Ds elements in maize. In: Peterson T, editor. Plant Transposable Elements: Methods and Protocols: Humana Press. p. 83-99.

Lowe K, et al. 2016. Morphogenic regulators Baby boom and Wuschel improve monocot transformation. Plant Cell. doi: 10.1105/tpc.16.00124

Ma Z, Dooner HK 2004. A mutation in the nuclear-encoded plastid ribosomal protein S9 leads to early embryo lethality in maize. The Plant Journal 37: 92-103.

McCarty DR, et al. 2005. Steady-state transposon mutagenesis in inbred maize. The Plant Journal 44: 52-61.

McClintock B 1984. The significance of responses of the genome to challenge. Science 226: 792-801.

Menand B, et al. 2002. Expression and disruption of the Arabidopsis TOR (target of rapamycin) gene. Proc Natl Acad Sci U S A 99: 6422-6427.

Pohlman RF, Fedoroff NV, Messing J 1984. The nucleotide sequence of the maize controlling element Activator. Cell 37: 635-643.

Rodina A, et al. 2016. The epichaperome is an integrated chaperome network that facilitates tumour survival. Nature 538: 397-401.

Schnable PS, et al. 2009. The B73 maize genome: complexity, diversity, and dynamics. Science 326: 1112-1115.

Slotkin RK, Martienssen R 2007. Transposable elements and the epigenetic regulation of the genome. Nat Rev Genet 8: 272-285.

Sosso D, et al. 2015. Seed filling in domesticated maize and rice depends on SWEET-mediated hexose transport. Nat Genet 47: 1489-1493.

Takai H, Xie Y, de Lange T, Pavletich NP 2010. Tel2 structure and function in the Hsp90-dependent maturation of mTOR and ATR complexes. Genes Dev 24: 2019-2030.

Tzafrir I, et al. 2004. Identification of genes required for embryo development in Arabidopsis. Plant Physiol 135: 1206-1220.

Wu Y, Holding DR, Messing J 2010. Gamma zeins are essential for endosperm modification in quality protein maize. Proc Natl Acad Sci U S A 107: 12810–12815.

Yamamoto K, et al. 2012. Kinase-dead ATM protein causes genomic instability and early embryonic lethality in mice. J Cell Biol 198: 305-313.