STRUCTURE-FUNCTION ANALYSIS OF THE RGH3 SPLICING FACTOR IN MAIZE

By

FEDERICO MARTIN

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

2012

To my parents and siblings

ACKNOWLEDGMENTS

I would first like to thank my parents, Oscar and Ana, and my siblings Santiago, Ma. Eugenia, Ma. Cecilia, Ignacio and Ma. Soledad for their tremendous support and unconditional love. They are in great part the reason I am here writing this document.

I would really like to thank my advisor Dr. A. Mark Settles for patiently guiding me from beginning to end, and for challenging me to become a much broader scientist. I also thank my committee members Dr. Christine Chase, Dr. Kenneth Cline, Dr. Richard Condit, and Dr. Bala Rathinasabapathi for their guidance and suggestions. I extend my gratitude to the members of the Plant Molecular and Cellular Biology (PMCB) program for their support, and to all current and former students who have been great partners and friends. I truly need to thank all the members of Dr. Settles' lab, particularly to Dr. Diego Fajardo, Dr. Romain Fouquet, and Christy Gault for all their help and guidance with this project. I also include Dr. Chi Wah Tseung, Dr. Gertie Spielbauer, Dr. Jeff Gustin, John Baier, Joe Black, Diana Grigalba, Alyssa Baggadion, Tyler Policht and Sarah Dailey for their great companionship and support. Finally, I specially thank my friends Carlos, Claudia, Jose, Cynnamon, Eugenia, Andres, Belen, Gabriela, Patan, Florencia, Pablo, Paola, Guy, Raul, Mike and to all of those who have supported me in all my scientific and life endeavors and with whom I have shared countless moments of enjoyment that will never be forgotten. Gracias totales.

TABLE OF CONTENTS

page

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

BP	Branch Point
BSA	Bulk Segregant Analysis
DAP	Days After Pollination
EJC	Exon Junction Complex
HNRNP	HETEROGENEOUS NUCLEAR Ribonucleoprotein Particle
NMD	Non-sense Mediated Decay
Py-tract	Polypyrimidine tract
RBD	RNA Binding Domain
RRM	RNA Recognition Motif
RUST	Regulated Unproductive Splicing and Translation
RGH	Rough Endosperm
SF1-BBP	Splicing Factor 1 Branchpoint Binding Protein
SNP	Single Nucleotide Polymorphism
SNRNP	SMALL NUCLEAR Ribonucleoprotein Particle
SRSF1	Serine-Arginine rich Splicing Factor 1
SR	Serine/Arginine rich
SS	Splice Site
SSR	Simple Sequence Repeat
TE	Transposable Element
UHM	$U2AF^{35}$ Homology Motif
URP	$U2AF^{35}$ Related Protein
U2AF	U2 Auxiliary Factor
ZRSR2	Zinc finger, RNA-binding motif, Serine-Arginine rich 2

Abstract of Dissertation Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

STRUCTURE-FUNCTION ANALYSIS OF THE RGH3 SPLICING FACTOR IN MAIZE

By

Federico Martin

August 2012

Chair: A. Mark Settles
Major: Plant Molecular and Cellular Biology

Alternative RNA splicing produces multiple mRNA species from individual genes increasing protein diversity and regulating gene expression. Genome sequencing projects have shown that about 42% to 45% of intron-containing genes in plants are alternatively spliced, but little is known about how alternative splicing is controlled. The *rough endosperm3* (*rgh3*) mutant causes developmental defects that are either seed or seedling lethal. *Rgh3* encodes a U2AF^{35} related protein (URP), which is a predicted RNA splicing factor. U2AF^{35} proteins identify splice acceptor sites during RNA processing and function through protein-protein interactions by creating complexes with U2AF^{65} and other Serine/Arginine rich-proteins (SR-proteins). Semi-quantitative RT-PCR analyses of alternatively spliced genes showed that *rgh3* affects splicing in a subset of genes supporting a role for RGH3 in alternative splicing. *Rgh3* is alternatively spliced, producing at least 19 different spliced variants. Interestingly, only one variant is predicted to encode a full-length URP ortholog containing an N-terminal acidic domain followed by two zinc fingers (Zn) flanking a U2AF homology motif (UHM) domain and a C-terminal RS-like domain. Several *Rgh3* splice variants produce truncated proteins missing one to several domains. GFP fused to full-length RGH3 localized to the

nucleolus and nuclear speckles. Functional analysis with the endogenous truncated protein variants and an artificial deletion of the UHM domain fused to GFP showed that while the acidic domain contains a nuclear localization signal, the RS-like domain enhances nuclear localization and is also important for protein recruitment to nuclear speckles. The UHM domain is a modified RRM domain that allows protein-protein interaction and in RGH3 it enables interaction with U2AF^{65}. These results suggest that RGH3 participates in the U2-type spliceosome and its function is regulated by alternative splicing creating truncated variants that render the protein unstable and excluded from the spliceosome.

CHAPTER 1
INTRODUCTION

The Spliceosome and the Splicing Reaction

Eukaryotic genes are made up by coding and non-coding regions. Exons generally represent coding regions which are separated by non-coding, intervening sequences known as introns. Once genes are transcribed into precursor-mRNA (pre-mRNA) intron sequences are removed by a mechanism known as RNA splicing (Reddy, 2007). The accurate removal of all intron sequences from pre-mRNA is often called constitutive splicing.

The splicing reaction takes place co-transcriptionally and is catalyzed by a multi-protein complex known as the spliceosome. Two types of spliceosome have been identified: a major spliceosome (U2-type), and a less abundant minor spliceosome (U12-type) (Chen and Manley, 2009). While the major spliceosome is found across all eukaryotic organisms, the minor spliceosome is only found in a reduced group including plants and most metazoans, but not in simple eukaryotic organisms (Patel and Steitz, 2003). The spliceosome is a very dynamic complex assembled in a multistep process that involves RNA-RNA, RNA-protein and protein-protein interactions many of which require ATP hydrolysis. Each step of spliceosome assembly requires formation of protein complexes that recognize exon and intron signals to define splicing sites. Splicing is completed after two *trans*-esterification reactions join concomitant exon sequences (Moore and Sharp, 1993). Signals within the intron include two consensus sequences at the 5’ end (donor site) and the 3’ end (acceptor site) of the intron, which determine splicing boundaries. A polypyrimidine tract (Py-tract) is found at the 3’ end of the intron to identify the branch point, which is located between 17-40 nucleotides

upstream the acceptor site (Reddy, 2001). The first *trans*-esterification reaction takes place when the 2' hydroxyl group of a conserved adenosine at the branch point site attacks the phosphodiester bond of 5' donor site of the intron to form a loop RNA structure know as the lariat intron. Afterwards, the exposed hydroxyl group of the 5' donor sites attacks the 3'end donor site to join the exons together with the second *trans*-esterification reaction (Patel and Steitz, 2003).

The accurate recognition of splicing signals and catalysis of the splicing reactions requires the participation of 5 small nuclear ribonucleoprotein particles (snRNPs) and over 150 other non-snRNP proteins (Wahl et al., 2009). The major U2-type spliceosome includes U1, U2, U4, and U6 snRNPs while the minor U12 spliceosome includes the functionally analogous but not identical U11, U12, U4atac, and U6atac snRNPs (Reddy 2007). The fifth snRNP is U5 and functions in both major and minor spliceosomes. In metazoans, the basic processes of spliceosome assembly and catalysis of the splicing reaction are well documented (reviewed in Wahl et al., 2009). Unfortunately, the lack of an *in vitro* splicing assay in plants has hindered similar biochemical studies in plant systems. Despite differences in intron composition between animals and plants, plants contain homologs of many of animal proteins involved in spliceosome assembly, including the main snRNPs from the major and the minor spliceosome (Wang and Brendel, 2004; Lorkovic et al., 2005), arguing for a conservation of the splicing mechanism.

The first step of the major spliceosome assembly involves the base pairing of the U1 snRNP to the 5' donor or splice site (5' SS) of the intron and the concomitant binding of the splicing factor 1 (SF1-BBP) to the branch point. This is followed by recognition of

the 3’ acceptor site or 3’SS of the intron by the U2 auxiliary factor (U2AF) heterodimer complex which also interacts with SF1-BPP. These initial interactions are known as the E (early) complex. After formation of the E complex, the U2 snRNP is recruited and base pairs to the branch point in an ATP-dependant manner. This process replaces the SF1-BBP to form the pre-spliceosome complex A (Chen and Manley, 2009). Next, a pre-assembled tri-snRNP complex formed by U4/U6-U5 which leads to the formation of the inactivated complex B. In order to become activated, complex B undergoes major conformational and compositional rearrangements that destabilize or release U1 and U4 snRNPs. The activated complex then undergoes the first catalytic step of splicing forming the lariat intron, a step that generates complex C. Complex C undergoes additional reorganization prior to catalyzing the second reaction to join the exon sequences. The final step involves disassembly of the splicing factors and release of the mRNA in the form of messenger ribonucleoprotein (mRNP, Wahl et al., 2009).

The minor U12-type spliceosome catalyzes the splicing of a rare class of introns that are diverged from U2-type introns. U12 introns have a longer and more tightly constrained consensus sequence at the 5’ SS, and by lacking a distinguishable Py-tract. U12 type of introns generally represent less than 1% of the introns in an organism; in *Arabidopsis* this number is about 0.7% (Reddy, 2007). Genes with U12 introns almost always contain other introns of the U2-type (Patel and Steitz, 2003). Despite the intron sequence differences, the assembly steps and conformational changes of the minor spliceosome resembles that of the major U2 spliceosome with the exception that U11 and U12 (U1 and U2 homologs respectively) form a pre-spliceosomal complex prior to binding the 5’ SS and the branch point, respectively (Frilander and Steitz, 1999).

The spliceosome is also is under tight regulation. Splicing signals within the pre-mRNA are not sufficient to effectively guide the assembly of the spliceosome. In fact, some of these signals need to be recognized multiple times by different factors to ensure a precise splicing reaction. Many of the interactions taking place in the spliceosome are weak and require multiple splicing factors to form stable complexes (Wahl et al., 2009). This is a crucial principle to confer flexibility to recognize the short and variable splicing signals within an intron. In addition to the core splicing signals, multiple *cis*-acting regulatory sequences in both introns and exons can enhance or inhibit the splicing reaction. These sequences are recognized by *trans*-acting splicing factors which regulate the identification of splicing signals and the assembly of spliceosomal complexes (reviewed in Chen and Manley, 2009).

The SR-family of proteins is the best studied group of regulatory splicing factors in both plants and mammals. This group of proteins comprises several phylogenetically conserved and structurally related proteins characterized by a domain rich in arginine and serine residues, known as the RS domain, and one or multiple RNA recognition motifs (RRM). Remarkably, SR-proteins are involved in a variety of functions that not only include regulation of the splicing reaction but also post-splicing processes such as mRNA nuclear export, nonsense-mediated mRNA decay and mRNA translation (reviewed in Long and Caceres, 2009).

Alternative Splicing and Gene Regulation

In addition to constitutive splicing, genes containing multiple exons and introns can display a diverse pattern of splicing known as alternative splicing. Alternative splicing generates multiple mRNA products from a single pre-mRNA sequence. To do this, the spliceosome recognizes alternative 5’ and 3’ SS to change the final length of exons, to

skip entire exons, or to retain entire introns (Nilsen and Graveley, 2010). Alternative splicing can produce unstable mRNA to down regulate protein expression. These transcript isoforms can also produce multiple protein isoforms with defective or divergent functions to enhance the proteome complexity of eukaryotic organisms.

Alternative splicing was initially believed to be a random event that resulted as a by-product of constitutive splicing. However, transcriptomic analyses have revealed that more than 90% of intron containing human genes undergo some kind of alternative splicing (Wang et al., 2008). Similar studies in *Arabidopsis* and rice estimated these numbers to be ~42% and ~48% of intron containing genes, respectively (Filichkin et al., 2010; Lu et al., 2010), arguing that alternative splicing is widespread. The functional and biological significance of alternative splicing is difficult to prove and a matter of intense investigation. Nevertheless, a variety of studies particularly in metazoans, have demonstrated the relevance of alternative splicing in multiple developmental processes (Black, 2003).

Alternative splicing can regulate transcripts by introducing premature stop codons (PTC) (Filichkin et al., 2010). Most PTC-containing transcripts are targeted for degradation by a quality-control survey mechanism known as nonsense-mediated decay (NMD). Even though the regulatory bases of this mechanism are not yet fully understood, recent studies have suggested that coupling of alternative splicing and NMD likely influences the abundance of functional transcripts, and hence protein levels (Lewis et al, 2003; McGlincy and Smith 2008). The coupling of these mRNA regulatory mechanisms is often called regulated unproductive splicing and translation (RUST) (Lareau et al. 2007a). In metazoans, this is exemplified by the alternative splicing of the

Sex-lethal (*Sxl*) and *transformer* (*tra*) genes in *D. melanogaster* where in each gene, sex-specific splicing produces a functional protein product in females. In males, the alternative splicing leads to the inclusion of stop codons, so that no functional protein is produced (Matlin et al., 2005). Similarly, in *Arabidopsis*, autoregulation by alternative splicing of *FCA* limits the amount of functional FCA protein to control the transition from vegetative to floral development (Reddy, 2007).

The regulatory effects of alternative splicing can also take place post-translationaly. Alternatively spliced transcripts may code for protein isoforms with divergent functions via alternative sub-cellular localization or as a protein missing functional domains. For example, it was recently demonstrated that the alternative splicing of the *Arabidopsis* SR-protein SR45 creates two isoforms with spatially different functions. While SR45.1 plays a major role in flower petal development, SR45.2 is relevant for proper root growth (Zhang and Mount, 2009). Similarly, alternative splicing of SR-proteins may lead to isoforms missing the C-terminal RS-domain which is relevant for regulation of protein localization and function by phosphorylation (Long and Caceres, 2009). Moreover, studies with SR-proteins in *Arabidopsis* demonstrated that absence of the RS-domain alters sub-nuclear localization of proteins likely affecting their function (Tillemans et al., 2005, 2006). Taken together, these examples further highlight the relevance of alternative splicing as a post-transcriptional and post-translational regulatory mechanism.

The U2AF Splicing Complex

The heterodimeric U2AF complex identifies and defines the 3' SS of the intron during formation of the pre-spliceosomal complex E (Wahl et al., 2009). The U2AF complex is formed by a large subunit, $U2AF^{65}$, and a small subunit, $U2AF^{35}$. Both

subunits have been identified in most metazoans and plant species based on domain structure and homology (Domon et al., 1998; Wang and Brendel, 2004, 2006a). Each subunit identifies different splicing signals, but cooperatively they define the 3'acceptor site of the intron and facilitate the recruitment of additional splicing factors. On the one hand, the large $U2AF^{65}$ subunit binds the pre-mRNA at the Py-tract with the help of two of its three RRM-domains. (Zamore et al, 1992; Mollet et al., 2006). On the other hand, the $U2AF^{35}$ subunit specifically identifies the dinucleotide signal that defines the 3'SS (Wu et al., 1999).Even though both subunits enhance accurate recognition of the 3' SS of introns, additional factors are necessary for their recruitment to true splicing sites. Recently, Tavanez et al. (2012) demonstrated that the heterogenous nuclear ribonucleoprotein A1 (hnRNP A1) proof-reads the pre-mRNA identifying pyrimidine-rich RNAs followed by 3' SS canonical dinucleotides and facilitates recruitment of the U2AF complex. Moreover, the correct identification of the 3'SS site by the $U2AF^{35}$ subunit is believed to be guided by SR-proteins bound to exonic splicing enhancer signals (ESS, Long and Caceres, 2009).

The binding between both $U2AF^{65}$ and $U2AF^{35}$ subunits confers stability and strength to the U2AF complex facilitating its binding to the 3' SS and its adaptation to the multiple rearrangements of the spliceosome. This binding is enabled by two reciprocal tryptophan (Trp) residues that form a "tongue-in-groove" binding platform (Kielkopf et al., 2001). In the $U2AF^{65}$ subunit, the Trp is found at an N-terminal polyproline region between the RS-domain and the first RRM domain. Conversely, in the $U2AF^{35}$the Trp residue is located in an RRM-like domain known as U2AF homology motif (UHM; Kielkopf et al., 2001, 2004). $U2AF^{65}$ also contains a UHM domain located

at the C-terminal end of the protein. However, this UHM domain does not interact with U2AF^{65}. Instead, it binds the SF1-BPP splicing factor and is believed to enhance subsequent recruitment of the U2 snRNP (Selenko et al., 2003; Mollet et al., 2006).

The many proteins associated with the splicing reaction provide the necessary flexibility that allows the spliceosome to better react to changes in cell state or the environment. Moreover, it enables the spliceosome to be prepared for the splicing of a wide variety of pre-mRNA introns present in an organism and which may require participation of alternative spliceosome components (Wahl et al., 2009). Even though the U2AF complex is involve in the splicing of the great majority of eukaryotic introns, other proteins have been found to be involved in recognition of particular types of 3' SS. A recently characterized protein in vertebrates, known as PUF60, shows high homology to U2AF^{65} and was able to substitute U2AF^{65} in early recognition of 3'SS (Hastings et al., 2007).Unlike U2AF^{65}, PUF60 seems to primarily recognize weak 3'SS. Similarly, CAPERα and CAPERβ are another set of U2AF^{65} related proteins (Dowhan et al., 2005).

The U2AF^{35} Related Protein (URP)

Genome sequence analyses and expression studies have revealed multiple genes encoding protein with similar domain organization and significant homology to the U2AF^{35} protein in vertebrates. These proteins include U2AF^{26} and two highly similar proteins named U2AF-RS1 and U2AF-RS2 (Tronchere et al, 1997; Shen et at., 2010). The human U2AF-RS2, also known as ZRSR2 or U2AF^{35} related protein (URP), shows an identical domain distribution as U2AF^{35} containing a central UHM domain flanked by two Zn fingers and an RS-domain at the C-terminal. In addition, URP is characterized

by the presence of an acidic-domain at its N-terminus. This protein was found to be necessary for proper splicing of U2-type introns as demonstrated by splicing analyzes with URP depleted nuclear cell extracts (Tronchere et al., 1997). Moreover, URP was able to interact with $U2AF^{65}$ *in vitro* through its UHM domain.Despite the structural and binding similarities between $U2AF^{35}$ and URP, their functions do not overlap. Recently, Shen at al. (2010) showed URP participates in splicing of U2-type introns and demonstrated that this protein exclusively contacts the 3' intron acceptor site and is require during the second catalytic step of the spliceosome. Surprisingly, the authors also found URP to participate in splicing of the rare U12-type introns where it also contacts the 3' end SS. URP had previously been associated with the U12 spliceosome by pull-down analyses (Will et al., 2004).

Recently, a homolog of the human URP was identified is maize by Fouquet et al. (2011). The protein, known as ROUGH ENDOSPERM3 (RGH3), was found to be involved in seed and seedling development and is particularly required for proper cell differentiation at the basal endosperm transfer cell layer (BETL) and embryo surrounding region (ESR). In this work, I analyze *Rgh3* transcript processing by alternative splicing and demonstrate that this post-transcriptional regulatory system is likely to control the abundance of functional protein as well as its participation in the spliceosome. Furthermore, I show that RGH3 is able to interact with the $U2AF^{65}$ splicing factor and is involved in regulation of a reduce set of alternative splicing events. Overall, the data presented here strongly argues that RGH3 has an analogous function to the human URP in maize.

CHAPTER 2
ALTERNATIVE SPLICING PRODUCES RGH3 PROTEIN ISOFORMS WITH DIFFERENT FUNCTIONS

Introduction

Once genes are transcribed into precursor-mRNA, intron sequences are removed by the mechanism of RNA splicing. Alternative splicing can generate multiple mRNA products from a single pre-mRNA sequence. The different transcripts from a single gene can either be unstable mRNA or code for protein isoforms with divergent functions. Alternative splicing, then, constitutes a versatile form of genetic regulation to influence protein abundance or function.

Splicing is catalyzed by a multi-protein complex know as the spliceosome. The spliceosome requires the dynamic interaction of 5 small ribonucleoprotein particles (snRNPs) and over 170 other proteins (Wahl et al., 2009). Two types of spliceosome exist in both plants and animals: a major spliceosome (U2-type), and a minor spliceosome (U12-type). Despite the large number of proteins involved in this process, many of them share common structural domains and are regulated in a similar fashion.One common form of regulation is alternative splicing of these factors. In plants, for example, *in silico* analysis using *Arabidopsis* EST and cDNA sequences showed a high frequency of alternative splicing events among splicing-related genes (Wang and Brendel, 2004). Further support for this mechanism was drawn from analyses of a well studied group of splicing factors, known as SR-proteins, demonstrating that in both mammals and plants these proteins undergo high levels of alternative splicing (Lareau et al., 2007b; Palusa et al., 2007). SR-proteins are a conserved family of proteins that play crucial roles as regulators of constitutive and alternative splicing by identifying enhancers or repressor sequences, as well as recruiting splicing factors and guiding

spliceosome assembly (Isshiki et al., 2006; Shen and Green, 2006). Interestingly, in A*rabidopsis*, splicing of SR-proteins is regulated in a developmental- and tissue-specific manner, and is influenced by hormones and multiple abiotic stresses indicating a true functional control mechanism rather than noise created by the splicing process (Palusa et al., 2007; Ali and Reddy, 2008b). Even though the regulatory bases of this mechanism are not yet fully understood, recent studies have suggested that coupling of alternative splicing and nonsense mediated decay (NMD) likely influences the abundance of functional transcripts through a mechanism known as regulated unproductive splicing and translation (RUST) (Lewis et al, 2003; Lareau et al. 2007a; McGlincy and Smith 2008). In plants, strong indications for this type of coupling have been demonstrated for SR-proteins as well as for several polypyrimidine track binding proteins (Palusa and Reddy, 2010; Stauffer, 2010).

In mammals, as in plants, members of the pre-mRNA splicing machinery typically show a punctuated distribution throughout the nucleoplasm while recruited into sub-nuclear structures such as the nucleolus, cajal bodies, and speckles (Lamond and Spector, 2003; Tillemans et al., 2005; Shav-Tal et al., 2005; Lorkovic et al., 2008; Koroleva et al., 2009; Stauffer et al., 2010). The nucleolus is the most prominent sub-nuclear compartment, though very few splicing factors were found to exclusively localize to this compartment (Pendle et al., 2005). However, multiple factors, especially snRNPs, have been shown to transiently pass through it as part of their maturation process (Lorkovic et al., 2004; Ali and Reddy, 2008a). Interestingly, several factors that do not localize to the nucleolus under normal conditions were seen to re-localize into it as a product of different stresses (Tillemans et al., 2005, 2006). Nuclear speckles are

located in the interchromatin space in the nucleoplasm, and are one of the major splicing factors storage sites (Spector and Lamnod, 2011). Mostly known for recruiting SR-proteins, nuclear speckles also house snRNPs and non-snRNPs as well as additional RNA processing proteins. In addition, they have been observed near active transcription sites and recently pre-mRNA have been detected in areas immediately adjacent to nuclear speckles indicating that these structures serve as storage and assembly areas for RNA processing factors (Reddy et al., 2012). Importantly, nuclear speckles are dynamic structures which formation and size are influenced by multiple processes including cell cycle, cell type and transcriptional activity (Spector and Lamnod, 2011; Ali and Reddy, 2008a and ref. therein).

Here, I describe the maize RGH3 protein, a SR-like protein that shows high degree of homology to the human U2AF^{35} Related Protein (HsURP), also known as ZRSR2. HsURP was first characterized by Tronchere et al. (1997) as an alternative splicing factor that participates in the U2 spliceosome. It was later demonstrated that HsURP also interacts with U12 spliceosome complex members and participates in splicing of rare U12 introns (Will et al., 2004; Shen et al., 2010). Through structural and functional analysis of *Rgh3* transcripts and RGH3 protein isoforms, I demonstrate that alternative splicing of *Rgh3* affects protein localization, likely regulating protein function.

Results

The *Rgh3* Transcript is Alternative Spliced

Rgh3 was originally identified by Fajardo (2008) while screening for seed mutants affecting endosperm-embryo interactions in maize. The *Rgh3* locus is not correctly assembled in the current B73 AGPv2 reference genome (www.maizesequence.org). To sequence and assemble the locus, I cloned the genomic locus as two overlapping

fragments from the maize BAC ZMMBBc497J22. Once cloned, the fragments were sequenced by primer walking and assembled into a single contig. To identify the *Rgh3* exons, I designed primers spanning the entire predicted coding region and tested seedling cDNA by RT-PCR. Numerous products were amplified andcloned, and from these 45 independent RT-PCR products were sequenced. These represented 19 *Rgh3* splice variants with the potential to encode seven protein variants (Genbank Accessions: JN791417 to JN791436). *Rgh3* transcript isoforms result from variable intron retention, exon skipping, and alternative 5' and 3' splice sites. Figure 2-1A shows schematics forrepresentative *Rgh3* mRNA isoforms that code for the seven protein isoforms. The *Rgh3*α isoform contains the full length coding sequence and is composed of 10 exons and retention of intron 1. The sequence encodes a predicted 755 amino acid peptides and shares significant protein sequence identity with HsURP at the central region composed of a UHM domain (Kielkopf et al., 2004) flanked by two CCCH type Zn-Finger domains. Similar to HsURP, RGH3 contains an N-terminal acidic domain, and an arginine/serine rich (RS) region at its C-terminus. By contrast, HsU2AF35 is less identical to RGH3 in the central domains and does not contain an acidic N-terminal region. For these reasons, I conclude RGH3 is homologous to human U2AF35 Related Protein (URP) (Tronchere et al., 1997). RGH3 shows considerable sequence identity and domain structure with a likely ortholog in Arabidopsis (Figure 2-1B).

The 18 other *Rgh3* isoforms identified in this study have a range of coding potential. Splicing of intron 1 produces a likely noncoding message; nevertheless it retains an uninterrupted URP reading frame that could initiate translation at a downstream start codon found within the UHM domain. Similar to *Rgh3*α, *Rgh3*ε

isoforms skip exon 6 and 7 but shows alternative splicing at exon 10 that creates a frame shift and premature termination codon (PTC) at the second zinc finger. *Rgh3ζ* also excludes exons six and seven but retains intron three. Retention of intron three causes a frame shift that incorporates a PTC and produces a truncated protein containing the N-terminal acidic domain exclusively. Isoforms *Rgh3β*, *Rgh3γ*, *Rgh3δ*, and *Rgh3η* have variations involving exons five, six, seven, and/or introns five, six and seven. These splicing events create frame shifts in the coding sequence introducing several new amino acids but ultimately producing a PTC either at the first zinc finger or the UHM domain (Figure 2-2A). Among the 19 identified splice variants, many include multiple alternative splicing events 3' of the key events that change coding potential and are technically different splice variants even though they code for one of the seven protein isoforms. For example, 4 different classes of transcripts were found to retain exon 6 but also included additional splicing events downstream. Exon 6 retention results in a transcript that codes for RGH3β, and a total of 11 of the 45 sequenced cDNA clones belonged to one of these 4 classes (Figure 2-2A, Table 1-1).

Protein Coding Potential of *Rgh3* Transcript Isoforms

The alternatively spliced *Rgh3* isoforms have the potential to regulate RGH3 protein expression either by destabilizing *Rgh3* transcripts or by altering RGH3 protein domains (Zang and Mount, 2009; Stauffer et al., 2010). To determine whether *Rgh3* variants code for proteins *in vivo*, I designed peptide antibodies (Ab) targeting the N- and C-terminus of the RGH3 protein. I then tested the Ab on protein extracts from multiple normal tissues of several maize inbred backgrounds (Figure 2-2). The RGH3α protein is predicted to be ~86 kDa (Figure 2-2A). As observed in Figure 2-2B, western

blot analyses show mixed results. The N-terminal Ab seems not to detect any proteins in either seedling or root tissues. On the contrary, it detects multiple bands in 24 days after pollination (DAP) seed protein extracts. None of these bands match the predicted sizes for RGH3 protein isoforms. The C-terminal Ab, detected a band of ~125 kDa only in seedlings tissues, and a single band of ~56 kDa in seed tissue.

In order to reconcile these results and to experimentally determine the SDS-PAGE electromobility of the RGH3 protein isoforms, I cloned *Rgh3*α, *Rgh3*ε, *Rgh3*β, and *Rgh3*γ cDNA and tested them in *in vitro* transcription/translation reactions using wheat germ extracts (Figure 2-3). The *Rgh3*ζ variant was not included because the N-terminal peptide antibody epitope lies just downstream of the PTC created in the variant (see Materials and Methods for information regarding Ab). Both, the N- and C-terminal Ab confirmed that RGH3α protein is ~125 kDa. These data suggest the proteins detected in western blots using *in vivo* protein extracts represent RGH3α. The three alternatively spliced variants produced translation products that cross-reacted with the N-terminal anti-RGH3 antibody, but not with the C-terminal antibody (Figure 3). As observed for RGH3α, the detected proteins are slightly larger than the expected 42, 33.5 and 30 kDa molecular weight for RGH3ε, RGH3β, and RGH3γ isoforms respectively. The results suggest that alternatively spliced *Rgh3* isoforms are capable of producing truncated protein versions of RGH3 and are likely to be found *in vivo*. Due to technical difficulties involving protein extraction and blotting, I have not been able to further confirm these results.

Subcellular Localization of RGH3 Protein Isoforms

To study the sub-cellular localization and distribution of RGH3 protein isoforms, I created green fluorescent protein (GFP) N-terminal fusions with RGH3α, RGH3β, RGH3ε, RGH3γ, and RGH3ζ. Each construct was transiently expressed in *N. benthamiana* leaves using Agrobacterium-mediated infiltration. In these experiments, RGH3α localized to the nucleolus at all times and was frequently found in nuclear speckles (Figure2-4). Since RGH3α is the predicted full length protein, it is likely that its localization reveals structures where the protein is functional. It is also possible that transient over-expression may reveal locations where the protein is sequestered if it accumulates to excessively high levels.

The RGH3β, RGH3ε, RGH3γ, and RGH3ζ isoforms localized only to the nucleolus and faintly disperse in the nucleoplasm (Figure 2-4). Interestingly, all tested truncated protein variants also localized in the cytoplasm. Under similar conditions of laser intensity and gain at the fluorescent microscope, the fluorescent signal emitted by these protein fusions was much lower compared to that emitted by GFP-RGH3α. The difference in signal could be attributed to a lower stability of the truncated mRNA which translate into lower protein abundance, or to the stability of the protein variants themselves. Together, these data suggest that RGH3 localization results from interacting signals within the acidic, and RS domains. On one hand, the acidic domain contains a nuclear localization signal, while on the other the RS-domain appears to be important for recruitment of the RGH3 protein into nuclear speckles, a phenomenon already observed with other RS-domain containing proteins (Tillemans et al., 2005). In

addition, localization of the truncated protein variants to both the nucleus and cytoplasm indicates that the RS-domain also influences nuclear localization.

The dynamic localization shown by the truncated variants seems not to be influenced by the presence or absence of the UHM domain. To test this hypothesis, I created a UHM domain deletion transcript and fusedit to GFP. As observed in Figure 2-4B, the fusion protein localizes to the nucleus. Like the full-length RGH3α, the protein is also found in the nucleolus but it localized to nuclear speckles much more often and readily than the full-length protein. The localization of this construct demonstrated that the acidic domain is necessary for nuclear localization, though it is not sufficient for this task as the RS-domain seems to aid proper nuclear and sub-nuclear localization. Also, the experiment confirms the importance of the RS-domain for nuclear speckle recruitment. UHM domains have the ability to act as protein-protein and/or protein-RNA binding domains (Kielkopf et al., 2004). It is therefore feasible that the RGH3 UHM domain facilitates localization to the nucleolus by enhancing interaction with other factors within this sub-nuclear compartment. Absence of the UHM domain may have weakened RGH3 ability to target other proteins or mRNAs in the nucleolus and thus enhance RS-domain capacity to recruit the protein into speckles.

Discussion

All higher eukaryote genes containing intronic regions must undergo splicing of their pre-mRNA in order to produce a mature mRNA ready for translation. Through alternative splicing, a single pre-mRNA with multiple introns can give rise to various mRNAs which can generate either unstable mRNA or code for different protein isoforms with defective or divergent functions (Reddy, 2007). Originally considered to be a

random phenomenon, alternative splicing is increasingly gaining attention as a versatile form of post-transcriptional regulation with vast implications for the transcriptome and proteome of eukaryotic organisms. In plants, recent studies in *Arabidopsis* and rice estimated that ~42% and ~48% of intron containing genes, respectively, undergo some type of alternative splicing (Filichkin et al., 2010; Lu et al., 2010). In humans, this number increases to over 90% of the transcripts (Wang et al., 2008) indicating that this is a widespread mechanism rather than a by-product of constitutive splicing.

The complete biological consequences of alternative splicing are difficult to realize. Nevertheless, the involvement of this mechanism in relevant developmental processes is well established and characterized. One of the best understood developmental pathways concerning alternative splicing is the sex determination pathway in *Drosophila* (reviewed in Matlin et al., 2005). In plants, the role of alternative splicing as a regulatory system is as not as defined as in metazoans. However, several investigations have demonstrated its effects in multiple developmental and defense response pathways. For example, the alternative splicing of FCA pre-mRNA acts as a developmental switch between vegetative and reproductive phases in *Arabidopsis* (Quesada et al., 2003; reviewed in Lorkovic, 2009).

Regulation of splicing factors by alternative splicing provides conclusive evidence of this mechanism's ability to control gene expression. Among all the proteins involved in splicing, the SR-family of splicing factors is the best studied group in metazoans as well as in plants. In humans, 11 SR-proteins have been identified compared to 18 found in *Arabidopsis* and 24 in rice (Barta et al., 2008). Plant and mammalian SR-proteins undergo high levels of alternative splicing, and many of the splicing events are

conserved across species (Iida and Go, 2006; Lareau et al., 2007a; Palusa et al., 2007). In fact, SR-proteins in both mouse and humans were shown to undergo the same alternative splicing patterns which take place at ultra-conserved elements within their sequences (Lareau et al., 2007a). In plants, how alternative splicing is regulated and the actual function of alternative spliced variants is not well understood, however, many of alternative spliced variants were found to include pre-termination codon (PTC) (Palusa et al., 2007; Tanabe et al., 2007; Filichkin et al., 2010). Most PTC-containing transcripts are targeted for degradation by nonsense-mediated decay (NMD), a common RNA surveillance mechanism found in most eukaryotes (Chang et al., 2007; Muhlemann et al, 2008). If left unprocessed, translation of PTC-containing transcripts could produce truncated proteins, as is the case for RGH3 isoforms (Figure 2-2), with potential negative effects for the cell. Based on multiple studies showing extensive coupling between alternative splicing and NMD, it was suggested that alternatively spliced variants containing PTCs may regulate the abundance of functional transcripts through the RUST mechanism (reviewed in Lareau et al., 2007b). Consistent with this hypothesis, extensive coupling between NMD and the abundance of alternative spliced variants of SR-proteins and other important genes was recently observed in *Arabidopsis* (Palusa and Reddy, 2010; Kalyna et al., 2012). Here, I demonstrate that the *Rgh3* transcript undergoes high levels of alternative splicing, producing spliced variants containing PTC (Figure 2-1B). The observed frequency of *Rgh3* spliced variants could be explained by the RUST mechanism. However, the high frequency of *Rgh3* cDNA clones containing a PTC suggests that the PTC transcripts are either stable and translated or they are produced at very high frequency when compared to the full-length

Rgh3α mRNA. Additional experiments are required to determine whether *Rgh3* transcripts are subject to RUST regulation. Overall, this example further demonstrates the importance of alternative splicing as a mean of post-transcriptional control.

Alternative splicing can also have post-translational repercussions by creating protein isoforms with alternative functions or sub-cellular localization. Once again, regulation of SR-proteins provides a clear example for this mechanism. SR-proteins have a modular structure, typically containing one or two copies of an RRM (RNA recognition motif) domain at the N-terminus and a C-terminal RS domain. RRM-domains provide RNA-binding specificity and RS-domains promote protein–protein interactions that facilitate recruitment of the spliceosome. In addition, RS-domains tend to act as a nuclear localization signals (NLS) affecting the subcellular localization of SR proteins by mediating the interaction with the SR protein nuclear import receptor, transportin-SR (Long and Caceres, 2009). RS-domains are highly phosphorylated, offering an additional level of control for the activity of SR-proteins. For example, in mammals, phosphorylation state of the RS-domain of the serine-arginine-rich splicing factor 1 (SRSF1) is known to control sub-cellular localization of the protein and to modulate interaction with the U1-70k unit of the U1 snRNP to initiate spliceosome assembly (Cho et al., 2011). Based on these examples, the RS-domain of RGH3 may also be required for localization of the protein into the nucleus indicating that even though the NLS is found in the acidic domains, this domain may not be sufficient for nuclear localization. Moreover, since the RS-domain of SR-proteins is located at the C-terminal end, alternative splicing of these proteins is likely to produce truncated protein isoforms missing the RS-domain. This, in turn, not only would prevent protein regulation

by phosphorylation but also affect sub-cellular localization and recruitment to the spliceosome. For example, deletion of the RS domain of the *Arabidopsis* SR-protein RSZp22 influenced re-distribution of the protein from nuclear speckles to the nucleolus (Tillemans et al., 2006). Consistent with these observations, RGH3 isoforms lacking the RS-domain fail to localize into speckles suggesting the RS-domain is needed for their recruitment into speckles. In addition, the presented data argues that RGH3 truncated protein variants are likely to be non-functional due to the alternative sub-cellular localization and failure to be recruited to spliceosomal speckles.

Finally, the full-length RGH3α protein was found to be localized in the nucleolus (Figure 2-3). The nucleolus is not a common storage site of splicing factors under normal conditions. RNA splicing factors typically accumulate in nuclear speckles and/or cajal bodies (Reddy et al., 2012). Different stress conditions including hypoxia, heat-shock, phosphorylation state or inhibition of transcription can cause some splicing factors to move into the nucleolus in plants (Tillemans et al., 2005, 2006; Koroleva et al., 2009). Based on these prior observations, the transient expression of RGH3 isoforms may cause cellular stress leading to localization to the nucleolus. Historically, the nucleolus has been known as the site of ribosomal RNA transcription and processing. Moreover, it has been linked to processing and assembly of a variety of RNPs, control of cell cycle and senescence, and as a sensor of stress (Brown and Shaw, 2008). Data from recent studies, however, has implicated the nucleolus in a wider range of functions including transcriptional gene silencing, and mRNA export and surveillance (Pendle et al. 2005; Brown and Shaw, 2008). Additionally, a higher abundance of aberrantly spliced mRNA were found in plant nucleolus compared to the

nucleoplasm suggesting that the processing of aberrant mRNAs may take place in this compartment (Kim et al., 2009). Interestingly, many splicing factors, particularly SR-family members, have been implicated in post-splicing mechanisms including mRNA nuclear export, NMD, and mRNA translation (Long and Caceres, 2009). Based on these observations, RGH3α localization in the nucleolus argues for its potential involvement in pre-mRNA splicing as well as post-splicing activities that may take place in the nucleolus.

In conclusion, I have shown that the *Rgh3* splicing factor gene produces numerous splice variants that introduce PTCs in the resulting mRNAs. Based on *in vivo* and *in vitro* reactions, and transient GFP fusion expression assays these variants have coding capacity. However, expression of the truncated forms suggests *Rgh3* may be under RUST regulation in order to modulate the level of functional transcripts. The data indicate that alternative splicing of *Rgh3* produces non-functional proteins acting as an additional level of regulation of RGH3. RGH3α localizes to nuclear speckles, consistent with the localization of multiple other splicing factors, and also into the nucleolus. Data from the human homologue and observed nucleolus localization of RGH3 argues for a potential involvement of RGH3 in splicing and post-splicing related functions. Combined these data suggest alternative splicing of the RGH3 gene regulates functional transcripts and protein function.

Materials and Methods

Cloning and Sequencing of *Rgh3*

The *Rgh3* locus was subcloned from the B73 maize BAC ZMMBBc497J22. The BAC was digested with *HindIII* and the fragments were cloned into pBluescript vector

(Thermo Scientific, USA). *Rgh3* contains a *HindIII* restriction site within intro 4 splitting the locus in half. Clones carrying the 5' end section were detected using standard southern blot technique using a 535bp probe spanning intron 1, exon 2 and intron 2. The 3' end section of the *Rgh3* locus was amplified in its entirety by PCR using Takara *LA* high fidelity taq DNA polymerase (New England Biosciences) and cloned into pTOPO vector (Invitrogen) according to manufacturer's instructions. Both clones were sequenced by primer walking. Sequenced fragments were then assembled into separate contigs and the 5' and 3' contigs were then assembled into a single locus using Vector NTI software (Invitrogen).

Rgh3 spliced variants were amplified from seed cDNA by RT-PCR using Phusion high fidelity enzyme (Finnzymes) following manufacturer's recommendations. Amplified RT- PCR products were cloned into pTOPO vector (Invitrogen). 45 total clones containing insertions over 2 kb were selected and sequenced by Sanger sequencing.

Western Blot Analysis of RGH3 Proteins

Two regions at the N-terminus (SAQEVLDKVAQETPNFGTE aa 202 to 220), and C-terminus (STKDDKRRKHHSGNRWH, aa 692 to 709) of the RGH3 proteins were targeted to produce peptide antibodies. Peptides were synthesized, purified and used to raise polyclonal antibodies in rabbit (Bio Synthesis, Lewisville, TX). Extraction of total protein was performed as described in Abdalla et al. (2009) with the following modifications: fresh instead of frozen tissue was used for extraction, and a single filtration step was performed instead of two. Total protein extraction was used for western blot analyzes instead of nuclei fraction.

For *in vitro* reactions, *Rgh3* variants were sub-cloned from pENTR vectors (see below) using the following primers. Primers contain restriction digestion sites for *Sgf1*

(Promega) and *Pme1* (New England Bio Labs) enzymes. PCR products were cloned into pF3AWG vectors (Promega, #L5671) through restriction digestion using above mentioned restriction enzymes and ligated using T4 DNA ligase enzyme as per manufacturer's instructions (Invitrogen). A total of 3µg of vector DNA was used for transcription/translation reaction using TnT SP6 high yield wheat germ protein expression system following manufacturer's instructions (Promega).

A total of 10 µl of protein extract and 12µl of TnT reaction were separated in an 8% gel by SDS-PAGE and blotted with anti-RGH3 peptide antibodies raised against the N-terminal acidic domain (1/3000 dilution) and the C-terminal RS-domain (1/5000 dilution) of RGH3α.

Subcellular Localization Studies

N- and C-terminal GFP fusion proteins were constructed by cloning cDNAs for RGH3 isoforms into pDONR221 or pENTR vectors (Invitrogen) according to the manufacturer's instructions. The ORFs were subcloned into pB7- FWG2 for C-terminal GFP fusions or pB7-WGF2 for N-terminal GFP fusions (Karimi et al., 2002, 2007) by LR recombination (Invitrogen). To build the RGH3ΔUHM deletion construct, cDNA from *Rgh3*α was used as template to amplify two fragments: one from the 5' ATG to the end of the first zinc finger, and the second from the 5' end of the second zing finger to the C-terminal stop codon. Both fragments were later ligated by overlap extension (sewing) PCR (Horton et al., 1989), and the obtained product was cloned into pENTR (Invitrogen) and sequenced. The entire fragment was then cloned into pB7-WGF2 as explained above. Transient expression experiments in *Nicotiana benthamiana* were completed essentially as described by Kapila et al. (1996) with the following modifications. Binary

vectors were transformed into Agrobacterium tumefaciens strain ABi by a freeze-thaw method (Wise et al., 2006). MES was not included in the Agrobacterium growing media, and *N. benthamiana* infiltration was completed with a 10-mL needleless syringe on 4- to 5-week-old plants grown in growth chamber at 22 to 24C with 16/8-h day/ night. For colocalization experiments, Agrobacterium strains carrying individual plasmids were mixed in a 1:1 ratio prior to infiltration. Fusion protein expression was visualized 24 to 48 h after transient transformation, and representative pictures of subcellular localization were obtained using a Zeiss Pascal LSM5 confocal laser scanning microscope as previously described (Pribat et al., 2010).

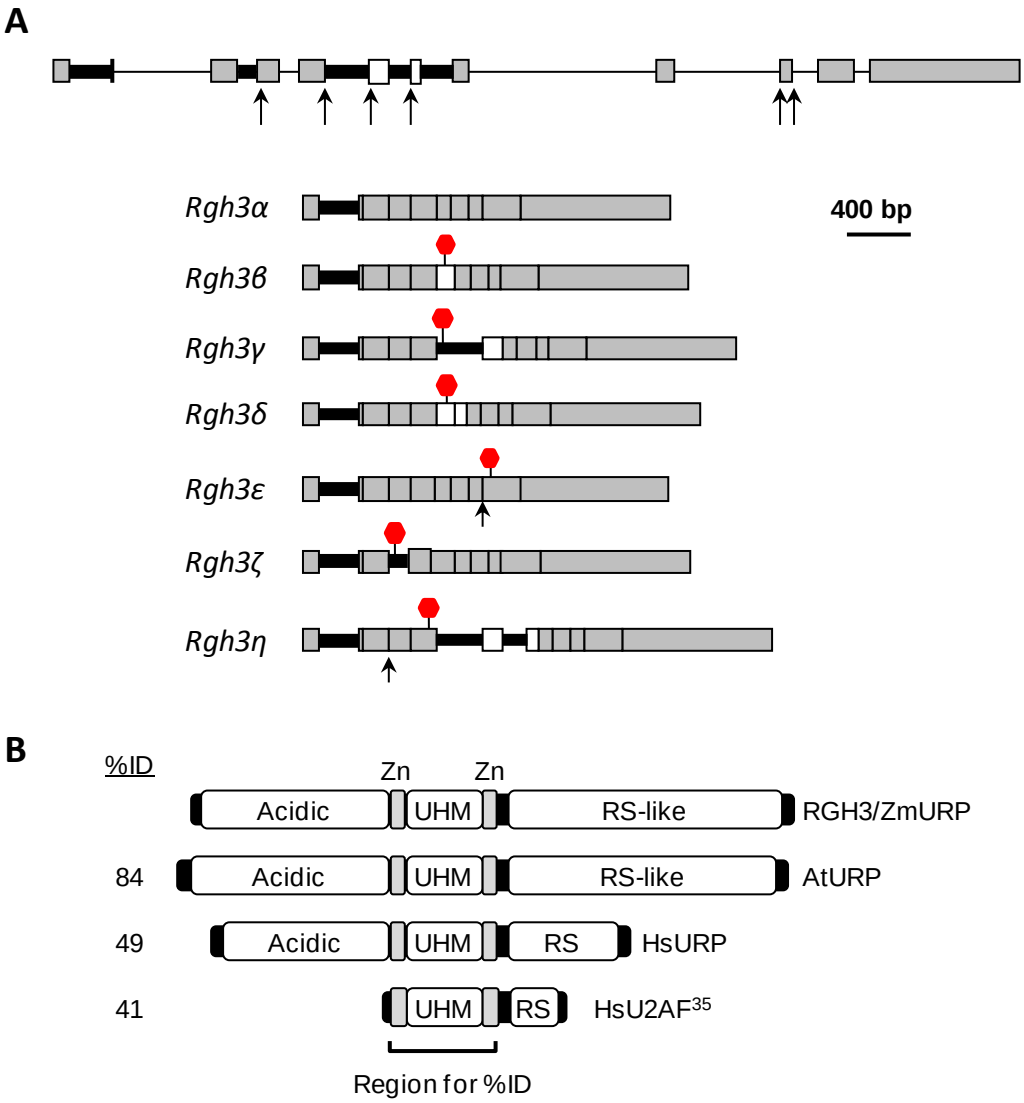

Figure 2-1. RGH3 is homologous to the human URP and its message is alternatively spliced. A) Schematic of the *Rgh3* locus (top panel). Alternative splicing gives rise to at least 19 splice variants, signified by Greek letters. Gray boxes indicate exons required to code the RGH3a protein, and open boxes are skipped exons for this isoform. Thick black lines indicate introns that are retained in one or more splice variants. Arrows indicate alternative 5' and 3' splice sites. Examples of splicing patterns that code for seven RGH3 protein variants are shown at the bottom. Alternatively splicing events introduce premature termination codons (PTC) indicated by red signs. B) Schematic of protein domains of URP and U2AF35 homologs. Sequence identity to the RGH3 Zn-UHM-Zn region is indicated. Zn, zinc finger. At, *Arabidopsis thaliana*; Hs, *Homo sapiens.*

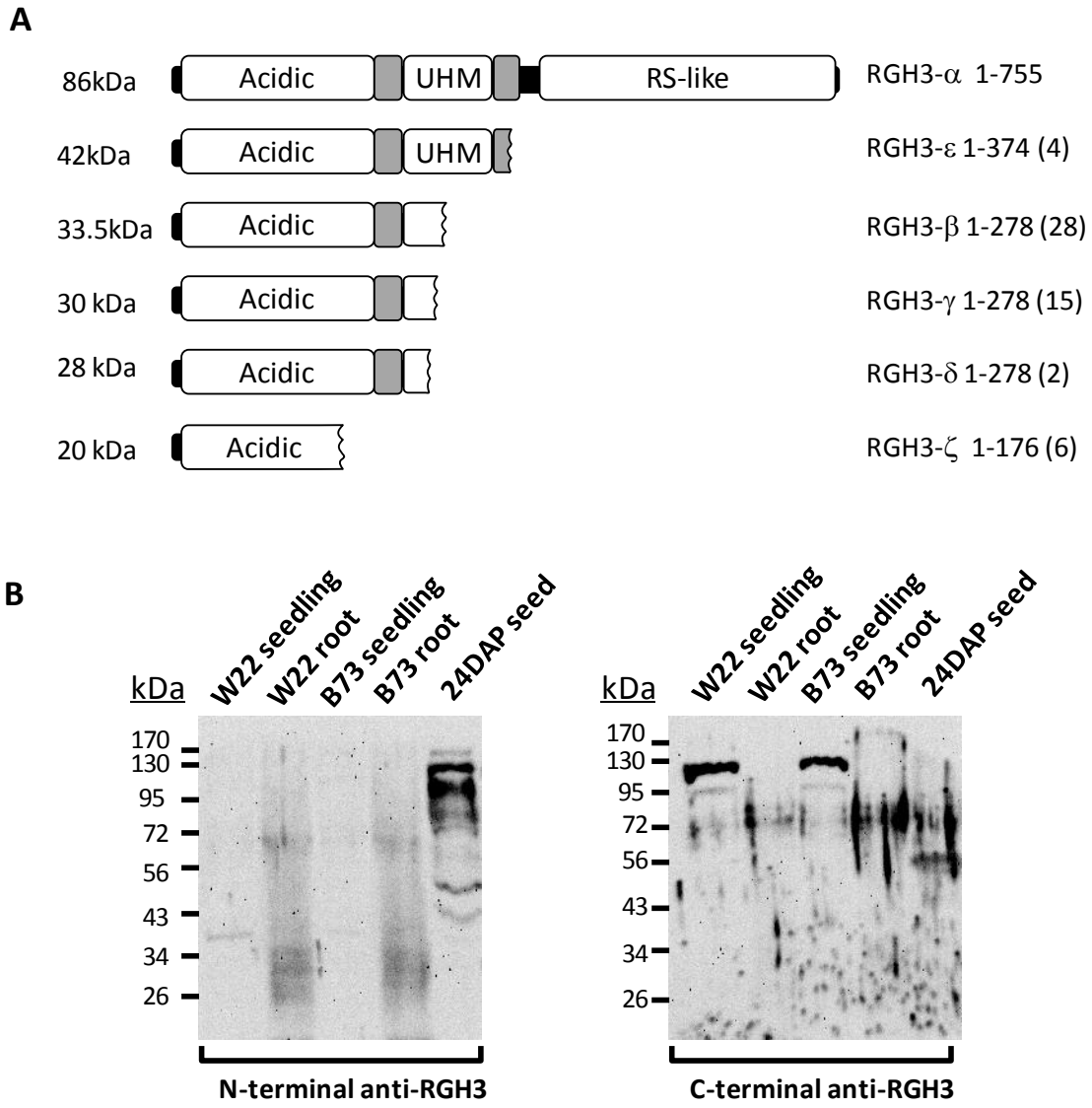

Figure 2-2. Alternative spliced Rgh3 variants produce truncated proteins. A) Schematic of the full-length and several truncated RGH3 protein isoforms detailing present and missing domains of each isoform. Peptides conforming each isoform are detailed on the right with additional amino acids introduced by changes in the frame of translation are indicated in parentheses; predicted protein sizes are specified on the left (not drawn to scale). B) Western blot analysis of protein extracts from multiple Rgh3/rgh3 tissues in W22 and B73 maize inbred backgrounds. Proteins were detected with N-terminal (left panel) and C- terminal (right panel) peptide antibodies (see Material and Methods for antibody information).

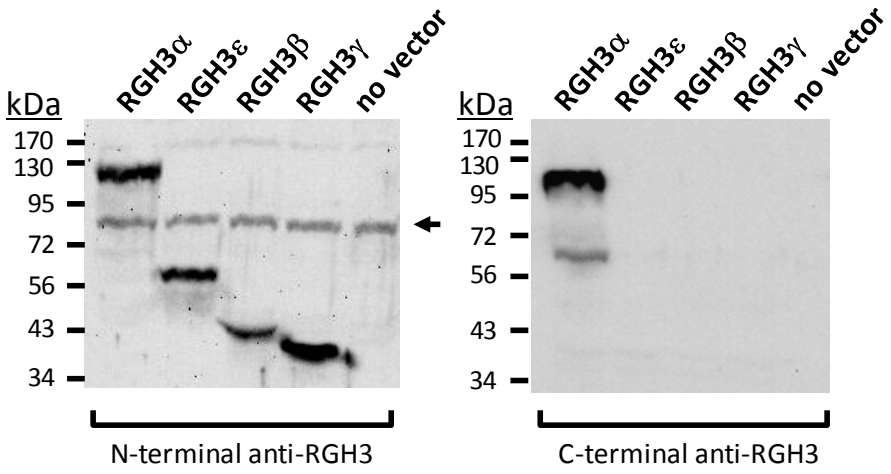

Figure 2-3. Western blot analysis of *in vitro* produced RGH3 normal and truncated protein isoforms. Proteins were detected with N-terminal (left panel) and C-terminal (right panel) peptide antibodies. Detected bands are of similar sizes to those detected in protein extracts from normal tissues indicating that sizes for protein isoforms were underestimated and are present *in vivo*. A non-specific band was detected by the N-terminal antibody that was also present in the no vector control (arrow).

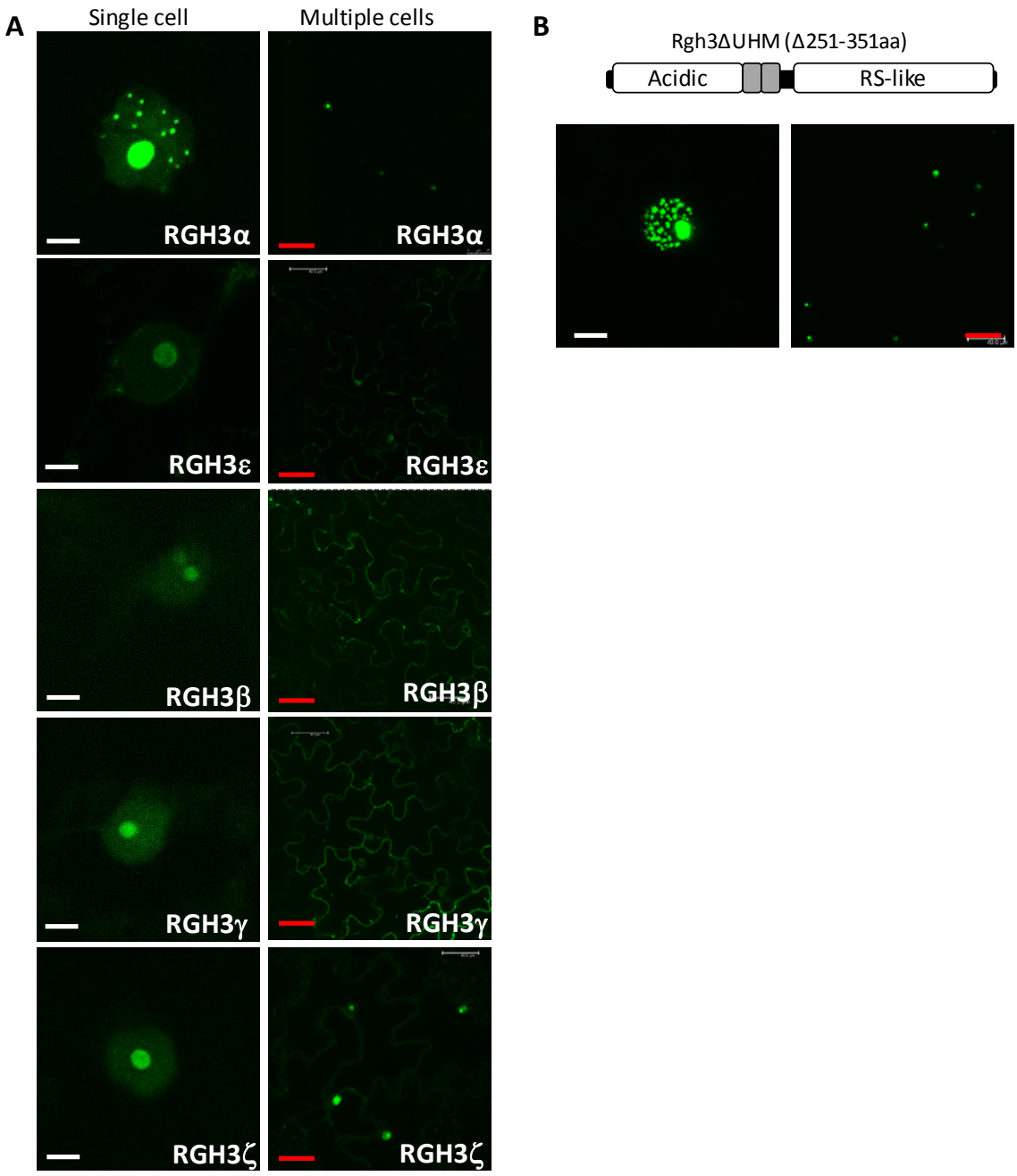

Figure 2-4. Transient expression of GFP tagged RGH3 variants in *N. benthamiana* leaves. A) Some of the RGH3 isoforms identified were tagged with GFP to analyze their localization. RGH3α localized to the nucleolus, to nuclear speckles and diffuse in the nucleoplasm. The remaining protein variants localized to nucleolus and remained diffuse in the nucleoplasm but not nuclear speckles. Expanded field of view demonstrates localization to cytoplasm. B) RGH3ΔUHM deletion variant localized to the nucleolus and was enriched in nuclear speckles but not in the cytoplasm. White scale bar = 5μm; red scale bar = 40μm.

Table 2-1. Number and classes of *Rgh3* alternative spliced variants

Isoform group	*No. of transcripts coding for the same isoform*	*Percentage*	*No. of different events producing same isoform**
RGH3α	11	24.4%	1
RGH3β	11	24.4%	4
RGH3γ	19	42.2%	8
RGH3ε	1	2.2%	1
RGH3δ	1	2.2%	1
RGH3ζ	1	2.2%	1
RGH3η	1	2.2%	1

*Isoform groups may include multiple transcripts containing several different splicing events that ultimately give rise to one of the indicated protein isoforms

CHAPTER 3
RGH3 FUNCTION IN THE SPLICEOSOME

Introduction

Splicing of introns by the spliceosome is a dynamic event that takes place co-transcriptionally. Proper formation of the spliceosome requires the stepwise assembly of smaller protein complexes that identify intronic as well as exonic signals, and catalyze two *trans*-esterification reactions (Chen and Manley, 2009). In metazoans, it has been shown that the earliest step of spliceosome assembly is identification of the 5' intronic splicing signal by the U1-70K snRNP in cooperation with the SR-protein SRSF1 (Cho et al., 2011). In addition to recognizing the 5' splice site (SS), other early events in spliceosome assembly include identifying the branch point (BP) by SF1-BBP. Subsequently, SF1-BBP interacts with the U2 auxiliary factor (U2AF) complex, which binds the polypyrimidine (Py) track and defines the BP and 3' SS (Wahl et al., 2009). Together, these interactions are known as the spliceosome E (early) complex. The next spliceosome assembly step is the recruitment and binding of the U2 snRNP by the U2AF complex. This allows U2-snRNP to base-pair with the BP in an ATP-dependant manner while displacing SF1-BBP to form what is known as complex A (Wahl et al., 2009).

The U2AF is a heterodimer complex formed by two sub-units, the large $U2AF^{65}$ and the smaller $U2AF^{35}$. Both subunits are SR-like proteins containing an RS-domain and RRM domains. $U2AF^{35}$ contains a single RRM, while $U2AF^{65}$ has three RRM domains (Mollet, 2006). Interestingly, the RRM domain in $U2AF^{35}$ and the third RRM domain in $U2AF^{65}$ belong to a new class of RRM domain termed U2AF homology motif (UHM) (Kielkopf et al., 2001; Selenko et al., 2003). UHM domains are able to adopt a

typical RRM-topology and have the potential to interact with RNA. However, they contain unique features that allow them to bind proteins as well (Kielkopf et al., 2004). The UHM domains found in the U2AF complex's members play crucial roles in spliceosome functioning. Through its UHM domain $U2AF^{65}$ is able to interact with SF1-BPP to enable recognition of the BP (Selenko et al., 2003, Banerjee et al., 2004). Similarly, $U2AF^{35}$ binds $U2AF^{65}$ through its UHM domain to enhance recognition of the 3' splicing site (Wu et al., 1999; Kielkopf et al., 2001). The UHM domains from the U2AF large and small sub-units do not interact directly with each other. Instead, the $U2AF^{35}$ UHM binds $U2AF^{65}$ at an N-terminal polyproline region between the RS-domain and the first RRM domain via a reciprocal "tongue-in-groove" interaction of two tryptophan residues (Kielkopf et al., 2001). The Trp of the UHM domain in $U2AF^{35}$ is flanked by an arginine (Arg) and a phenylalanine (Phe) residue, a motif known as Arg-X-Phe motif (where X is any amino acid), located in a hydrophobic pocket where the interaction takes place. The Arg-X-Phe motif is one of the key features that distinguish UHM domains from canonical RRM domains (Kielkopf et al., 2004).

The human URP, a protein showing similar domain structure and homology as $U2AF^{35}$, was found to support splicing of U2 type introns and interacted with $U2AF^{65}$ *in vitro* through its UHM domain (Tronchere et al, 1997). Moreover, URP function is non-redundant with that of $U2AF^{35}$ as $U2AF^{35}$ fails to complement URP-depleted nuclear extracts for in vitro splicing activity. Interestingly, Shen et al. (2010) revealed that URP is also able to participate in U12-type intron splicing demonstrating that URP can interact with both spliceosomes even though it influences distinct steps of splicing for U2 and U12 introns. The RGH3 protein in maize shows significant conserved homology

with human URP at its central region and shares the URP domain structure suggesting it is likely an ortholog of URP. In this section, I provide evidence suggesting that RGH3 participates in the U2 spliceosome and shows similar protein-protein interactions as the human URP.

Results

RGH3 UHM-Domain is Structurally Fit for Protein-Protein Interaction

The central Zn-UHM-Zn domain in RGH3 shows significant homology with URP proteins in human and is a bit more distantly related to the same domains in $U2AF^{35}$ (Figure 2-1). In plants, neither $U2AF^{65}$/$U2AF^{35}$ nor $U2AF^{65}$/URP interactions have been studied in detail. To investigate whether the same interactions could occur in plants, I compared the peptide sequences from the Zn-UHM-Zn domains of $U2AF^{35}$ and URP from human, maize, rice, and *Arabidopsis* (Figure 3-1). The protein sequence alignment demonstrates that the hydrophobic pocket of the $U2AF^{35}$ protein is poorly conserved between human and plant species, but is highly conserved among plant homologs (Figure 3-1A, red underline). Unlike $U2AF^{35}$, the hydrophobic pocket of URP proteins is better conserved across species, even though they are not identical (Figure 3-1B, red underline). Comparison of the Arg-X-Phe motif presents a different scenario. In the UHM domain of both $U2AF^{35}$ and URP, the Arg-X-Phe motif shows higher conservation between human and plant samples (Figure 3-1, orange box). Even though the Arg-X-Phe motif is a signature motif that distinguishes UHM-domains from canonical RRM-domains, variations at the Phe position for similar bulky aromatic amino acids have been observed in other UHM containing proteins (Kielkopf et al., 2004). While the human $U2AF^{35}$ protein presents a signature Arg-Trp-Phe motif (Trp denotes Tryptophan), maize and rice proteins show a conserve Arg-Tyr-Tyr (Tyr denotes

Tyrosine) and *Arabidopsis* shows an Arg-Trp-Tyr. Similar residue changes are observed in URP proteins (Figure 3-1B, orange box). The human URP protein contains an Arg-Trp-Tyr motif while all the plant proteins have a conserved Arg-Tyr-Phe. It is important to highlight that the above mentioned amino acid changes involve substitutions for similar bulky aromatic amino acids which may not significantly alter the phobicity of the binding pocket. Interestingly, plant homologs of both U2AF^{35} and URP contain a Tyr residue instead of a Trp at the X position which in humans is required for the "tongue-in-groove" binding between U2AF^{35} and U2AF^{65}. This change, however, is consistent with a Trp to Tyr residue exchange observed in plant U2AF^{65} proteins at the likely binding site with U2AF^{35}. These residue changes argue that the "tongue-in-groove" biding evolved to take place among Tyr residues instead of Trp (Kielkopf et al., 2004). Taken together, the data argues that the changes in the UHM-domain hydrophobic pocket of plant URP proteins may not significantly alter the phobicity or conformation of the binding pocket. Furthermore, the residue changes in the Arg-X-Phe motif of UHM domains is compensated by a similar change in U2AF^{65} arguing for a conserved re-arrangement enabling the "tongue-in-groove" biding to take place.

RGH3 Co-localizes with U2AF^{65}a at Sites of Transcription

If RGH3 interacts with maize U2AF^{65} in a similar manner as found in vertebrates, I expect both proteins to localize to overlapping sites within the nucleus. To test this hypothesis, I cloned the maize versions of the U2AF^{65}a and U2AF^{35}a proteins and fused them to either GFP or RFP. The constructs were expressed either individually or in combination with RGH3α through agrobacterium-mediated transient expression in *N. benthamiana* leaves (Figure 3-2). U2AF^{65}a localized to the nucleoplasm and was

concentrated into nuclear speckles. Similarly, U2AF^{35}a localized to the nucleoplasm and showed concentration to nuclear speckles as previously observed for the *Arabidopsis* protein (Wang and Brendel, 2006a). In contrast to RGH3, none of the U2AF factors localized to the nucleolus (Figure 3-2A). When RGH3 was co-expressed with U2AF^{65}a, both proteins showed a similar distribution in the nucleoplasm (Figure 3-2B). Even though nuclear speckles were not clearly defined, both proteins concentrated in similar structures and also showed a similar disperse localization within the nucleoplasm. The co-localization was not always observed since at times both proteins localized to different locations in the nucleoplasm (data not shown), possibly due to either the viability of the tissue or cell cycle state of the cell which has been shown to influence re-arrangement of nuclear speckles or transcriptional activity of the cell (Fang et al, 2004; Spector and Lamond, 2011). Interestingly, U2AF^{65}a signal was also observed into the nucleolus suggesting a re-arrangement of its localization due to the presence of RGH3α (Figure 3-2B, red arrowhead). As a control, I also tested the co-expression of U2AF^{65}a and U2AF^{35}a tagged with GFP and RFP, respectively (Figure 3-2C). Expression of both proteins showed a similar distribution in the nucleoplasm but not in the nucleolus, indicating a potential conserved interaction between the proteins to form the U2AF heterodimer complex in plants. Once again, nuclear speckles were not clearly defined and the observed co-localization was dispersed in the nucleoplasm at likely sites of transcription. Taken together, the data indicate that RGH3α and U2AF^{65} have overlapping localization in the nucleus and high levels of RGH3α expression seem to influence U2AF^{65} sub-nuclear localization. These observations support a conserved interaction of the proteins in vertebrates and plants (Tronchere et al, 1997).

RGH3 Natural Protein Isoforms Fail to Co-localize with U2AF^{65}

In chapter 2, I showed that truncated RGH3 variants express at low levels and do not localize like full length RGH3α (Figure 2-3). Potentially, the missing domains of RGH3 variants could affect colocalization with U2AF^{65}a. To analyze this possibility, I co-expressed several natural RGH3 truncated variants with U2AF^{65}a-RFP fusion protein. The tested isoforms include RGH3ε, which is only missing the RS-domain, and RGH3β and RGH3γ which contain the acidic-domain, the first Zn-finger and small a portion of the UHM domain (Figure 2-2B). As shown by the images in figure 3-3, none of the tested variants co-localizes with U2AF^{65}a. While thesevariants are observed in the nucleolus and partially disperse in the nucleoplasm, U2AF^{65}a remains in the nucleoplasm. Moreover, U2AF^{65}a localization is concentrated in nuclear speckles suggesting that truncated RGH3 proteins do not influence U2AF^{65}a localization. Taken together, these data argues that the RS and UHM domains are necessary for a potential interaction with U2AF^{65}a.

RGH3α Interacts with U2AF^{65}a *in planta*

Although co-localization data are suggestive of a conserved URP/U2AF^{65} interaction in plant cells, this is not direct evidence for protein-protein interactions. I used the bimolecular fluorescence complementation (BiFC) assay to address whether the interaction is real and is found *in vivo*. BiFC involves the fusion of two proteins, one as the bait and the other as a prey, to either the N- or C-terminal half of a fluorescence protein. If the tested proteins interact, the two halves of the fluorescence protein would come into close proximity, fuse and emit a fluorescent signal upon excitation (Citovsky et al., 2006). Thus, I fused RGH3α to the N-terminal half of the yellow fluorescent

protein (YFP) and U2AF^{65}a to the C-terminal half of YFP. These constructs were transiently co-expressed in *N. benthamiana* leaves. Figure 3-4A shows images from two independent biological tests demonstrating the emission of fluorescent signal from the nucleus. Interestingly, the signal was found dispersed in the nucleoplasm particularly concentrated around the nucleolus and in structures that resemble nuclear speckles. The distribution within the nucleus observed in this test is reminiscent and consistent with the co-localization of both proteins (Figure 3-2A) indicating that RGH3 interacts with U2AF^{65}a. To further test the nature of this interaction I co-expressed the UHM domain deletion construct (GFP-RGH3ΔUHM) with U2AF^{65}a-RFP fusion protein. If the RGH3/U2AF^{65}a interaction requires the UHM domain, the co-expressed constructs should fail to co-localize. Consistent with this hypothesis RGH3ΔUHM localizes to the nucleolus and speckles, while U2AF^{65}a is dispersed throughout the nucleoplasm (Figure 3-4B). In addition to being diffused in the nucleoplasm U2AF^{65}a shows localization to speckles, but these speckles are not the same as those where RGH3ΔUHM localizes. From these data it can be concluded that RGH3α is able to interact with the splicing factor U2AF^{65}a and the RGH3 UHM-domain is necessary for this interaction to take place. Moreover, the BiFC data suggest that co-localization between RGH3 and U2AF^{65}a is likely to indicate an interaction.

Discussion

In metazoans, the process of spliceosome assembly and catalysis of the splicing reaction are well documented (reviewed in Wahl et al., 2009). The lack of a plant *in vitro* splicing assay has hindered investigations of the assembly of the spliceosome and regulation of the splicing reaction. Nevertheless, plants contain homologs of many of

animal proteins involved in spliceosome assembly and catalysis (Wang and Brendel, 2004; Lorkovic et al., 2005) arguing for a conservation of the constitutive and alternative splicing mechanism. The first task of the spliceosome is recognize the 5'SS, branch point, and 3'SS. Splicing sites, by themselves, are usually not sufficient to guide assembly of the spliceosome and require additional enhancing signals recognized by factors such as SR-proteins (Long and Caceres, 2009). Due to the variety of pre-mRNA existing in an organism and to be able to adapt to changes in cell state or environment, the spliceosome requires participation of many loosely associated proteins (Wahl et al., 2009). Interactions between splicing factors and pre-mRNA tend to be weak but are enhanced by the collaboration of multiple factors each with weak binding sites (Hastings et al., 2007). These weak interactions allow the spliceosome to be a dynamic complex. Interestingly, as demonstrated by its stepwise assembly, not all the proteins associated with the spliceosome participate in every complex. Many of them only aid the proper formation of a specific complex, are involved in splicing of specific events, or act as a proof-read mechanism to help identify vague splicing signals (Wahl et al., 2009).

In vertebrates, the U2AF heterodimer complex identifies the Py-tract and the 3' SS of the intron, a process that leads to the formation of the spliceosome complex E (Wu et al., 1999; Black , 2003). Both, small and large subunits the U2AF complex have been identified in plants based on their high structural homology indicating a conserved functional identification of the 3'SS (Domon et al., 1998; Wang and Brendel, 2004, 2006a). However, 3'SS recognition is dynamic and does not always require the U2AF heterodimer. A recently characterized protein in vertebrates, known as PUF60, shows high homology to $U2AF^{65}$ and was able to substitute $U2AF^{65}$ in early recognition of 3'SS

(Hastings et al., 2007). PUF60 primarily recognizes weak 3'SS further highlighting the dynamic recognition of different and alternative splicing signals by the spliceosome. Similarly, URP interacts with U2AF^{65} *in vivo* to exclusively contact the 3' SS of the intron and to participate in spliceosome complexes in vertebrates (Tronchere et al., 1997; Shen et al., 2010). I confirmed a conserved interaction of RGH3α and U2AF^{65} *in planta* through co-localization and BiFC assays (Figure 3-4). Interestingly, the BiFC interaction takes place in areas within or immediately adjacent to the nucleolus indicating that the re-localization of U2AF^{65} into the nucleolus observed in co-localization assays is not an artifact produced by the high abundance of both proteins. This localization, further suggests the involvement of RGH3 in activities that take place within the nucleolus including post-splicing RNA surveillance functions. Based on the observed speckled distribution of the interaction, it can be speculated that U2AF^{65}/RGH3 complex forms prior to being recruited to functional sites. Interestingly, though, the localization of U2AF^{65} and RGH3ΔUHM deletion protein to different nuclear speckles indicates that the interaction does not occur in nuclear speckles. This suggests that the proteins are stored independently but may be recruited to an additional sub-nuclear structure where they form pre-spliceosomal complexes reminiscent of the U11/U12 pre-spliceosomal complex that initiates assembly of the minor spliceosome (Frilander and Steitz, 1999; Will et al., 2004) (Figure 3-3B). Furthermore, the conserved UHM domains from U2AF^{35} and URP (Figure 3-1) suggest that U2AF^{35} and URP interact with U2AF^{65} through a similar binding site in the UHM domain. The RGH3/U2AF^{65} interaction argues for a conserved URP function in plants and thus may also participate in assembly of the

major spliceosome. In humans, U2AF^{35} and URP do not overlap functionally indicating that their interaction with U2AF^{65} serves different purposes in the identification of 3' SS.

Interestingly, none of the tested RGH3 protein variants co-localize with U2AF^{65} (Figure 3-3). The RGH3ε and RGH3ΔUHM proteins delete the RS-domain and UHM domain, respectively. The absence of co-localization with U2AF^{65} for both of these proteins suggests that both domains are necessary for interaction yet neither is sufficient. The lack of co-localization with the natural protein variants supports the hypothesis that the alternative splicing of *Rgh3* results in proteins that are not recruited to the spliceosome. In Fouquet et al. (2011), a quantitative RT-PCR assay was designed to amplify the *Rgh3* region between exon 4 and exon 8, where most of the alternative splicing events takes place (see Figure 2-1A), in order to detect and estimate the relative concentration of transcripts coding for multiple *Rgh3* isoforms. Due to the extensive alternative splicing throughout the length of the Rgh3 transcript, the assay did not unambiguously measure *Rgh3α* versus all other transcripts. However, the relative level of the key splicing event required to code RGH3α protein was found to represent no more than 20% of *Rgh3* transcripts suggesting the abundance of the functional *Rgh3*α transcript is regulated. Overall, the data presented here strengthens the argument that alternative splicing of *Rgh3* regulates RGH3 protein levels which may take place in a developmental- or tissue-specific manner.

Materials and Methods

Protein Sequence Alignment

Predicted protein sequences of URP and U2AF^{35} were identified through BLASTP and TBLASTN searches in the National Center for Biotechnology Information databases (http://www.ncbi.nlm.nih.gov/), the Joint Genome Institute (www.jgi.doe.gov) servers,

and the plaza plant comparative genomics resource (Proost et al., 2009) using default settings at each server. Conserved domains were identified using Prosite scans (http://www.expasy.ch/tools/scanprosite/) at the European Bioinformatics Institute server. Protein sequence alignments were completed with ClustalW2 at the European Bioinformatics Institute server using default parameters (http://www.ebi.ac.uk/Tools/msa/clustalw2/): Gonnet protein weight matrix, gap open = 10, gap extension = 0.2, gap distances = 5, no end gaps = no, iteration = none, numiter = 1, clustering = NJ.

Subcellular Localization Studies

U2AF^{65} and U2AF^{35} coding sequences were extracted from www.maizesequence.org and amplified by RT-PCR using normal 14 days old seedling leaf cDNA. N- and C-terminal protein fusions, as well as agrobacterium transformation and localization analyses were conducted as previously described in the materials and methods section of chapter 2.

Vectors for bimolecular fluorescent complementation assays were created by transferring RGH3α and U2AF^{65}a cDNA from pENTR vector clones to pSAT4-DEST-nEYFP and pSAT5-DEST-cEYFP (Tzfira et al., 2005; Citovsky et al., 2006) respectively by LR clonase reaction (Invitrogen). Individual transcription cassettes were digested from cloned vectors with the rare cutters *I-CeuI* and *I-Sce-I* and ligated into the same pPZP-RCS2 binary vector using T4 DNA ligase (Invitrogen) in a stepwise manner. Agrobacterium transformation and transient infiltration procedures were conducted as previously described in the materials and methods section of chapter 2. Protein expression was visualized 24 to 48 h after transient transformation, and representative

images were obtained using a Leica TCS SP5 confocal laser scanning microscope (Leica Microsystems). YFP was excited at 514nm and detected with an emission band of 525 to 565nm.

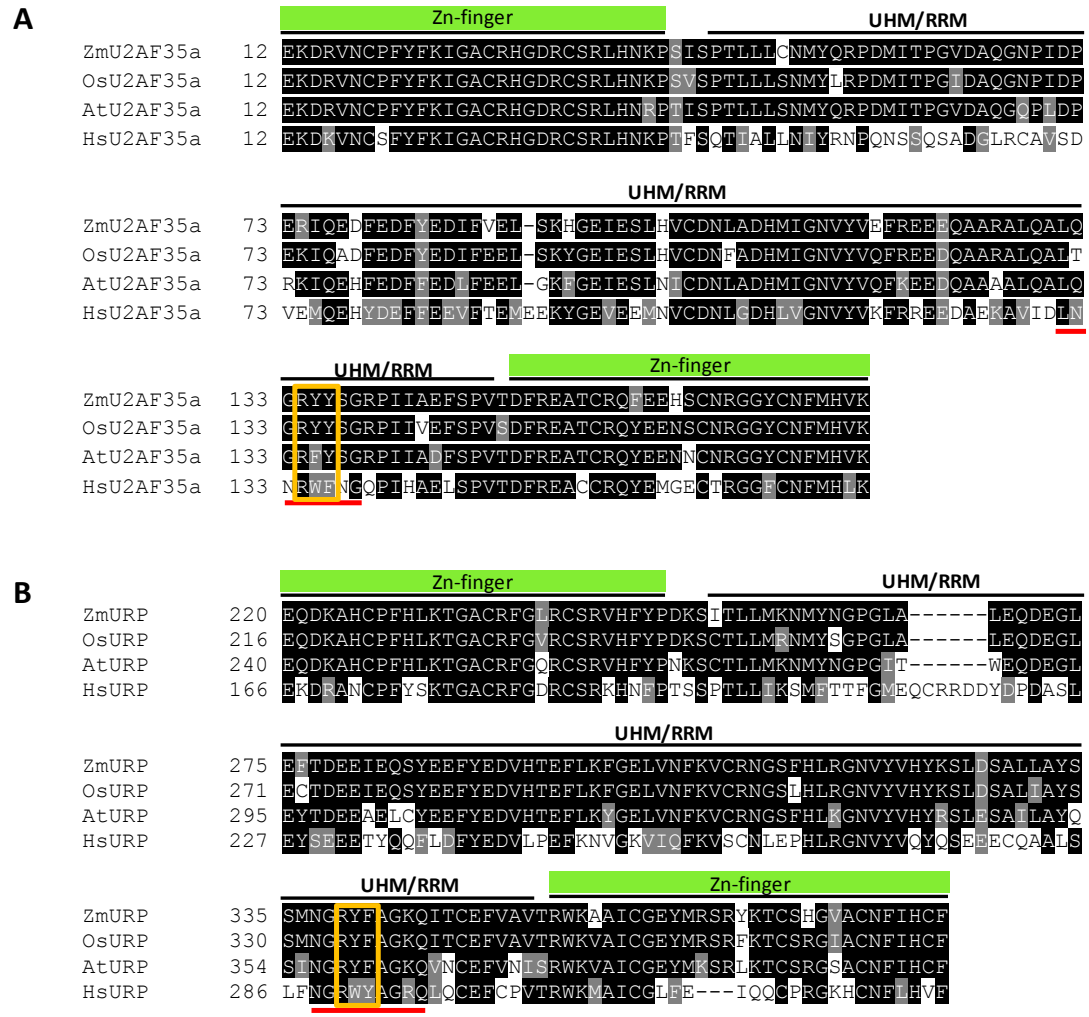

Figure 3-1. Multiple sequence alignment of the conserved zinc finger-UHM-zinc finger region of U2AF35 (A) and URP (B) proteins from multiple species. Identical residues are shown in white letters with black shading, similar residues are highlighted in gray. Zn fingers and UHM domain are labeled on top of the sequence. Residues involved in intermolecular contacts as previously shown by Kielkopf et al. (2001) are underlined in red. Arg-X-Phe motif is highlighted by orange box. The conserved domains were identified within each protein using Prosite scans (http://www.expasy.ch/tools/scanprosite/) and the sequences were aligned with ClustalW2 using default settings (http://www.ebi.ac.uk/Tools/msa/clustalw2/). ZmURP, RGH3α *Zea mays*; Os, *Oryza sativa*; At, *Arabidopsis thaliana*; Hs, *Homo sapiens*.

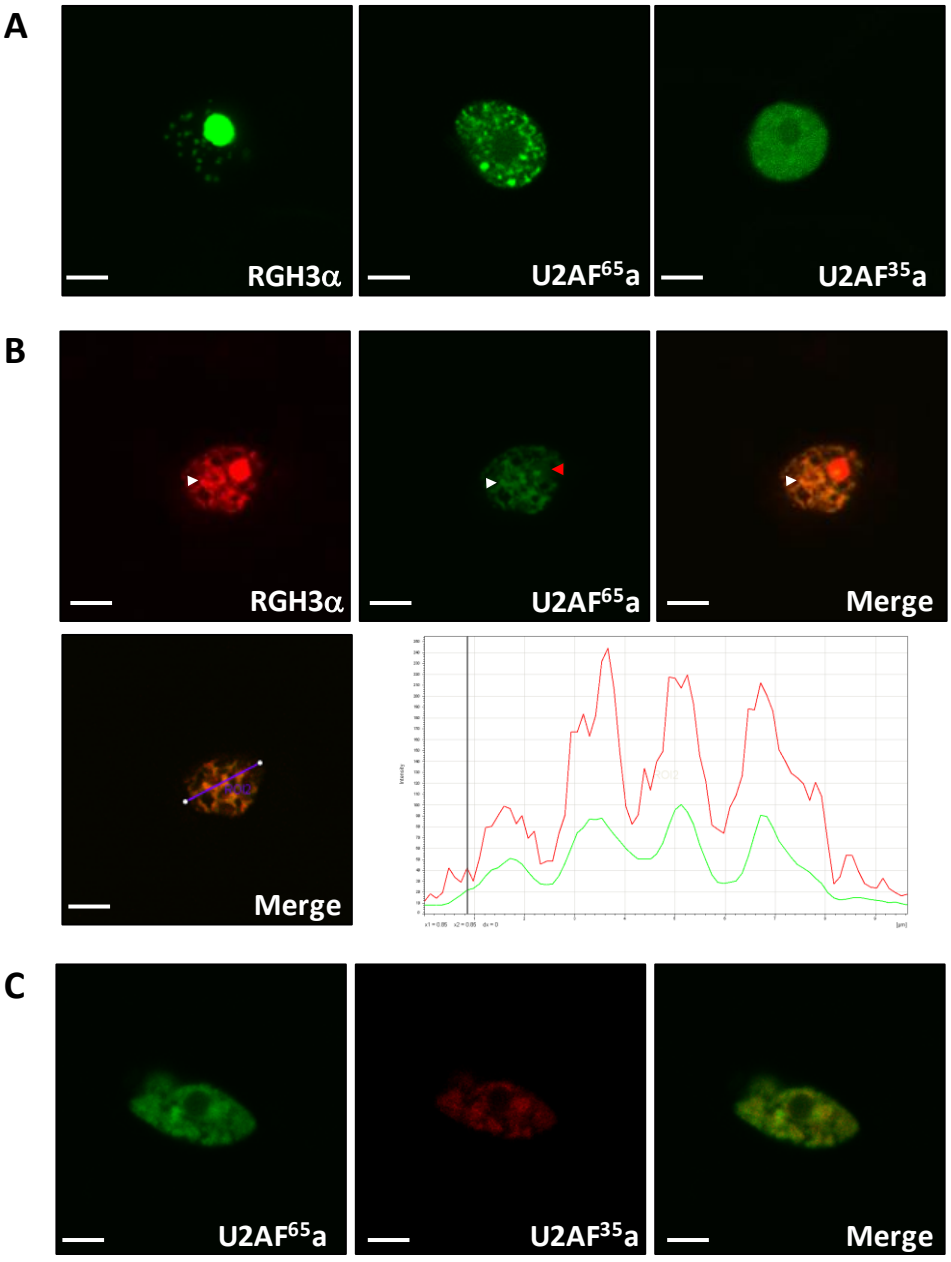

Figure 3-2. RGH3α co-localization with U2AF^{65}a. A) Individual localization of RGH3α, U2AF^{65}a, and U2AF^{35}a fused to GFP. All proteins localize to nuclear speckles but only RGH3α is observed in the nucleolus. B) Co-expression of RGH3α-RFP with GFP-U2AF^{65}a demonstrates that both proteins co-localize in nuclear speckles (white arrowheads) and disperse in the cytoplasm at potential sites of transcription. Also, U2AF^{65}a is re-localized to the nucleolus (red arrowhead) due to a potential interaction with RGH3α. Lower panel demonstrates overlapping of fluorescent signals from a z-stack layer of the above image. C) Co-localization of U2AF^{65}a with U2AF^{35}a throughout the cytoplasm indicates that both protein are very likely to interact having a conserved function as that described in metazoans (scale bars = 5µm).

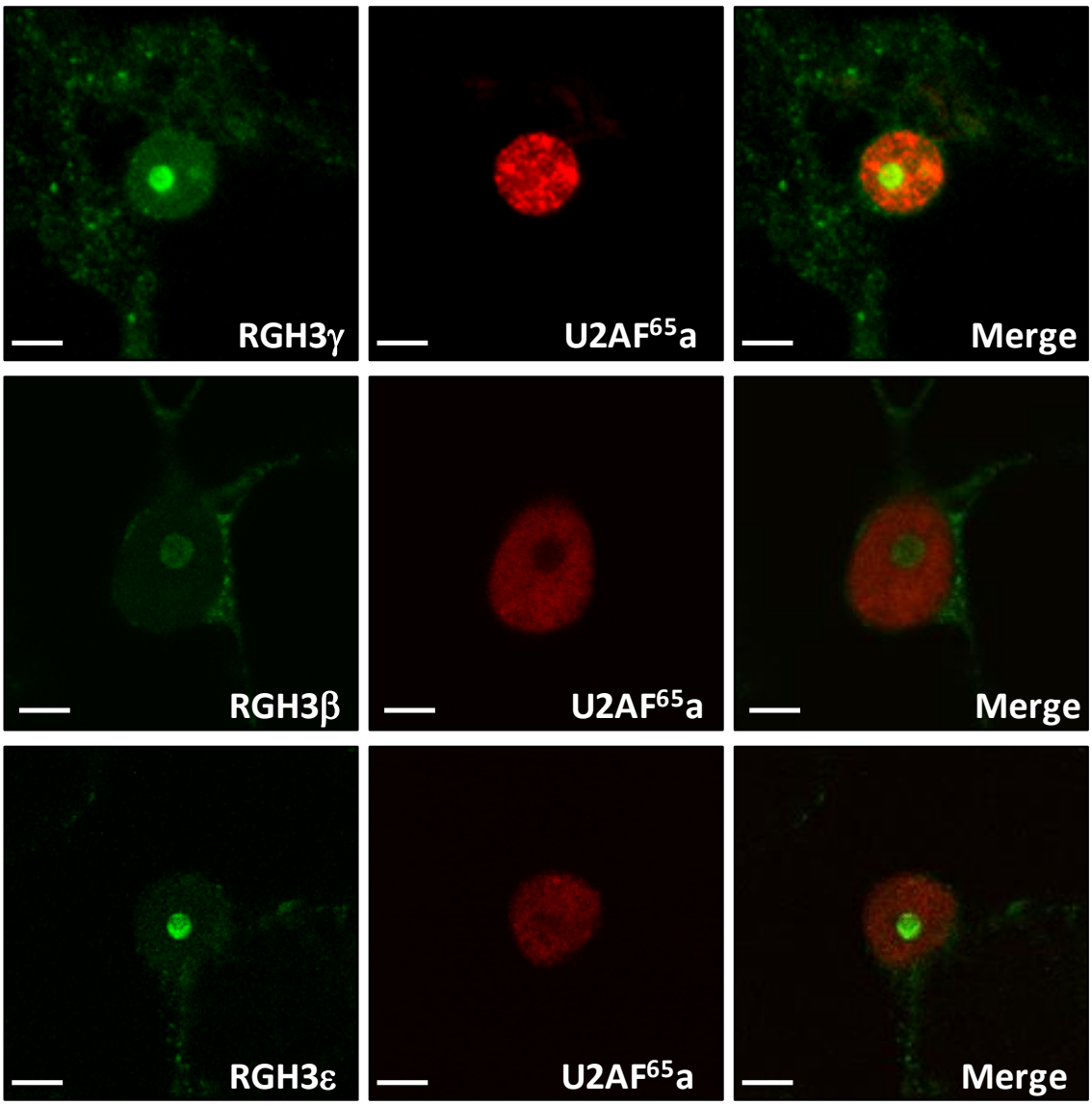

Figure 3-3. RGH3 truncated protein variants fail to co-localize with U2AF^{65}a. To test the effect of domain deletions in RGH3 function, three representative RGH3 truncated protein variants (RGH3γ, RGH3β and RGH3ε) were fused to GFP and transiently co-express with U2AF^{65}a-RFP in *N. benthamiana.* Consistent with previous observations, none of the RGH3 protein variants are recruited to spliceosomal speckles and are concentrated in the nucleolus (left panels). These proteins fail to co-locallize with U2AF^{65}a (right panels) which remains concentrated in speckles and disperse in the nucleoplasm (center panels) (scale bars = 5μm).

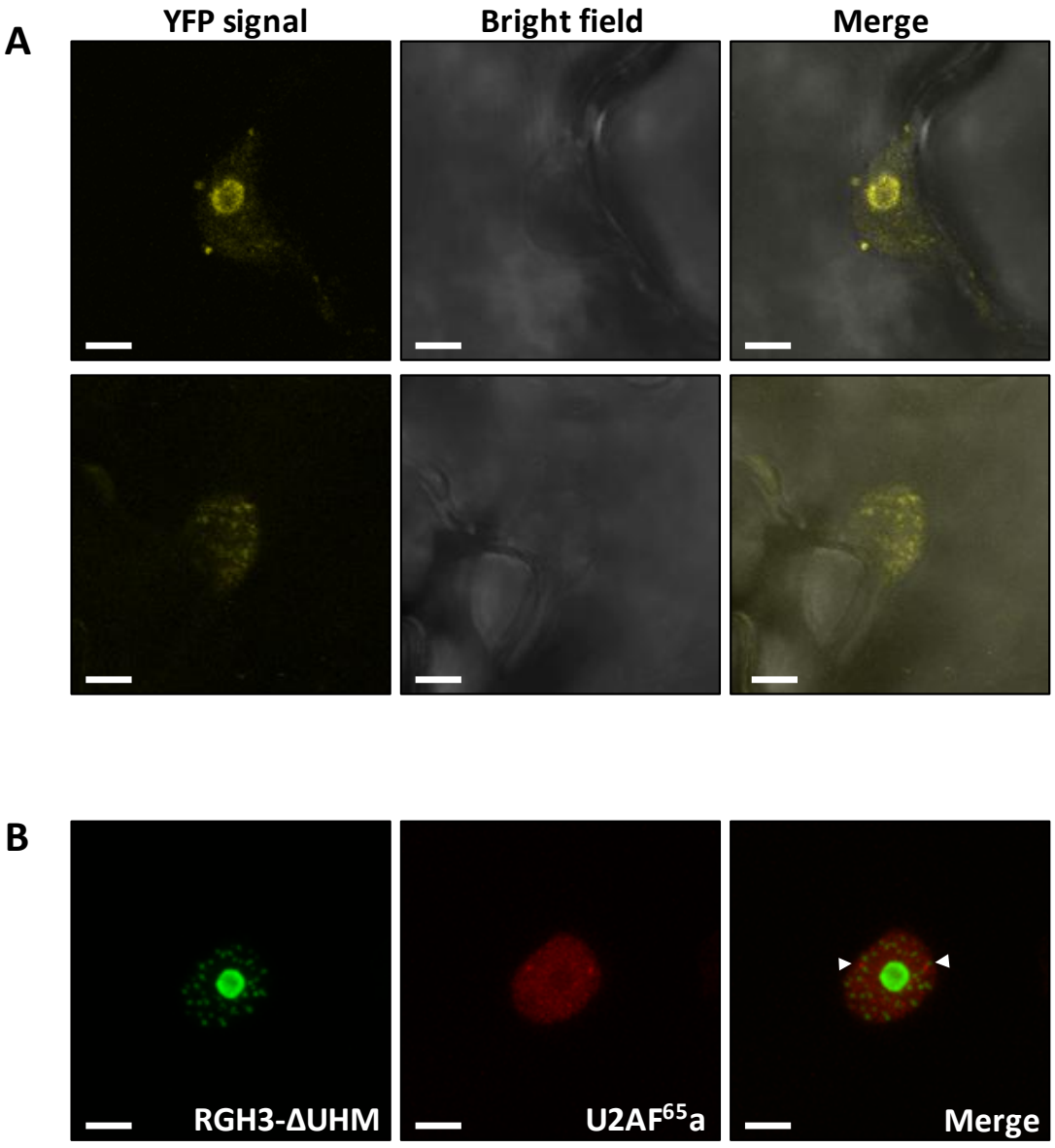

Figure 3-4. RGH3α interacts with U2AF^{65}a *in planta.* A) BiFC assays demonstrate the ZMURP interacts with U2AF^{65}a in the nucleus. Most samples show fluorescent signal from areas surrounding the nucleolus (top row), while others emit signal from structures that resemble nuclear speckles (bottom row). Two independent biological replicates were performed confirming the results. B) RGH3-ΔUHM deletion construct fused to GFP fails to colocalize with U2AF^{65}a-RFP in transient co-expression assays in *N. benthamiana.* RGH3-ΔUHM is seen in the nucleolus and nuclear speckles while U2AF^{65}a-RFP is disperse in the cytoplasm and concentrated in speckles. Interestingly, both proteins seem to localize in different speckles in the cytoplasm (white arrowhead) (scale bars = 5µm).

CHAPTER 4
ANALYSIS OF THE *rgh3-umu1* HYPOMORPHIC ALLELE

Introduction

Transposable elements (TE) are movable DNA fragments that can insert into new chromosomal locations. These elements often duplicate themselves during transposition to amplify within the genome. Originally discovered in maize by Barbara McClintok, transposable elements are present in almost all studied organisms representing up to 80% of the genome in some species (Feschotte et al., 2002). Eukaryotic TEs are typically organized in two groups based on their mode of transposition: class I elements include all retrotransposons which mobilize via an RNA intermediate; class II elements are transposed as DNA (Hua-Van et al., 2005). Both classes are subdivided into autonomous elements capable of producing the necessary machinery for transposition and non-autonomous elements that require the proteins encoded by an autonomous element in order to transpose.

Regardless of the type of element or mode of transposition, TEs are considered to be a major force in genome evolution and rearrangement. Often referred as ‘selfish’ or parasitic DNA, transposable elements influence genome size through their ability to self-replicate and amplify (Feschotte et al., 2002). Moreover, transposition or homologous recombination between transposon-rich areas influences genome organization. TEs also contribute to chromosome form and function by maintaining and regulating heterochromatic areas, where they tend to concentrate, such as centromeres and telomeres (Slotkin and Martienssen, 2007). The mutagenic nature of TEs and their potential effects on gene expression can lead to novel functional variation within a genome. Even though the great majority of transposable elements present in a genome

are generally inactive, active elements transpose and can insert themselves within or near genic sequences. For example, in maize Class I retroelements greatly outnumber Class II DNA-elements. Though, the former are usually clustered in methylated and heterochromatic gene poor areas. On the other hand, DNA transposons are mainly found in unmethylated, genetically active euchromatic regions (Bennetzen, 2000). The activity and function of transposable elements is believed to be the product of an evolutionary adaptation in order for the elements to survive and propagate within a host genome. In principle, a high abundance of transposable elements in gene-rich areas could have harmful, deleterious effects on a host genome (Hua-Van et al., 2005). Thus, it is in the best interest of both genomes and TEs to alleviate or remove these deleterious effects. Consequentially, many TEs have co-adapted by adopting strategies such as transposition bias into non-coding regions, excising from exonic sequences through splicing, or activation in specific tissues (Kidwell and Lisch, 1997). These patterns of transpositions impact on host organisms include the diversification of protein function through introduction of new introns or exons. In maize, a clear example of such an adaptation was recently demonstrated by studies using *Helitron* mobile elements. *Helitrons* show a peculiar capability of capturing gene fragments within the element which are then moved as the element moves. Barbaglia et al. (2012) confirmed that the gene fragments captured by the element are transcribed and demonstrated that through alternative splicing, transcripts containing fragments from different genes are created. Surprisingly, the authors also found transcripts containing joined exons from genes within the *Helitron* and exons in close proximity to the mobile element insertion site. Thus, the *Helitrons* mobile elements have the ability to create new transcripts that could

be selected naturally to improve host genome's fitness. Other examples in maize include *Ds* alleles of the alcohol dehydrogenase (*Adh)* and the ADP-glucose glucosyl-tranferase (*Wx*). Both alleles contain a *Ds*-element inserted in an exon and encode wild-type-sized proteins that show an intermediate phenotype when the activator *Ac* element is not present and *Ds* cannot transpose (Purugganan and Wessler, 1992). This phenomenon is possible due to the splicing of the *Ds*-element from the pre-mRNA. Other evolutionary adaptation include activation of transcription within a TE that affect transcription of targeted or nearby genes and influence gene expression in a tissue- or developmental-specific manner (Girard and Freeling, 1999; Slotkin and Martienssen, 2007).

Transposons are useful tools to generate mutants for to the study of molecular biology and functional genomics. In maize, transposable elements have been used as endogenous mutagens in forward and reverse genetic programs to tag and clone genes (Settles, 2009). Transposon-tagging uses the sequence of DNA-transposons to identify mutagenized genes. The known transposon sequence can be used to enrich for DNA adjacent to the transposon insertion (Brutnell, 2002). Multiple mutagenized populations using a variety of transposable element families such as *Ac/Ds*, *En/Spm*, and *Mu* exist in maize genes (Brutnell, 2002; Settles, 2009). The UniformMu population, developed by McCarty et al. (2005), harnesses a high mutagenic frequency produced by the high-copy-number native *Mutator* element present in maize. The *Mutator* element is a class II DNA-element showing strong insertional bias towards gene-rich, non-repetitive areas of the genome particularly near the 5'end region of a gene (McCarty et al., 2005). The *rough endosperm3* (*rgh3*) seed mutant was original identified by Fajardo (2008) from

the UniformMu population while screening for seeds showing nonautonomous functions in endosperm and embryo development. A *Mu*-insertion within the first exon of a gene coding for a U2AF^{35} related protein (URP) was found to be tightly linked to the seed phenotype (Fouquet et al., 2011). In this chapter, I show that the mapped *Mu*-insertion produces a hypomorphic allele producing a functionally deficient protein that is likely compromised in its ability to participate in spliceosomal complexes.

Results

Mu-transposon Insertion Produces a r*gh3-umu1* Hypomorphic Allele

The *rough endosperm* (*rgh*) class of seed mutants have a typical rough surface to the endosperm that is also termed etched or pitted (Scanlon et al., 1994; Neuffer et al., 1997). A subset of mutant seeds from the *UniformMu* population showing a *rgh* phenotype were screened by B-A translocation in order to identify mutants affecting endosperm–embryo developmental interactions (Fouquet et al., 2011). The *rgh3* mutant became of particular interest because its uncovering crosses suggested that a mutated endosperm negatively impacts wild-type embryo development (Fouquet et al., 2011). A *Mutator* transposable element insertion in the long arm of chromosome 5 was found to be tightly linked with the *rgh3* phenotype; the *Mu*-induced allele was then referred as *rgh3-umu1* (Fajardo, 2008; Fouquet et al., 2011).

The linked *Mu*-element was inserted within the first exon of *Rgh3* coding sequence (see Figure 2-1 for coding sequence scheme). The *rgh3-umu1* allele was initially believed to be a null allele. However, RT-PCR analysis on normal and mutant tissues using *Rgh3* specific primers flanking the *Mu*-insertion region amplified a single size-shifted band on all mutant tissues when compared to normal (data not shown). Sequencing analysis of the RT-PCR products revealed the presence of a 141 bp

segment of the *Mu*-element terminal inverted repeat (TIR) produced by alternative splicing of the *Rgh3* transcript. I performed additional RT-PCR amplifications utilizing mutant seedling tissue, and cloned and sequence nine full-length cDNAs. All nine cDNA clones contained the 141bp *Mu*-element indel fragment formed by junction of a cryptic donor site found within the *Mu*-element with exon 3 of the *Rgh3* transcript. Surprisingly, and even though the *Mu* insertion deleted 12 amino acids of the wild type protein, the fragment introduced 47 new amino acids that did not alter the coding frame (Figure 4-1). Interestingly, eight out of the nine clones were predicted to code for a full length protein similar to RGH3α (Figure 4-1), while the remaining product could code for a protein similar to RGH3γ. These data argues that *rgh3-umu1* is not a null allele but rather a hypomorphic allele.

The unexpected low abundance of transcripts coding for RGH3 truncated proteins in mutant tissues argues that the *Mu*-insertion may somehow interfere with the alternative splicing mechanism that give raise to alternatively spliced isoforms in normal *Rgh3* tissues. Interestingly, a similar example where a *Mu* insertion alters the splicing pattern of a transcript was previously observed in two mutant alleles of the maize *Alcohol dehydrogenase1* (*Adh1*) gene (Ortiz and Stromer, 1990). Moreover, another UniformMu seed mutant identified in our laboratory shows the exact same *Mu*-element splicing pattern observed in *rgh3-umu1* (Jeffery Gustin, unpublished data).

RGH3$^{umu}\alpha$ is Found *in vivo* and Shows Partial Co-localization With U2AF^{65}

To test if the *rgh3-umu1* transcript codes for a protein *in vivo,* I extracted protein from normal and mutant 24DAP seeds, and modified and unmodified mutant seedling (see Chapter 5 for details on mutant seedlings). Western blots were then probed with

the anti-RGH3 N-terminal peptide antibody which detected bands of similar sizes in the normal as well as the mutant seed tissues (Figure 4-2A). From the data presented in chapter 2, the heaviest band in normal tissues (~125kDa) belongs to RGH3α. This indicates that the band of similar size detected in mutant seed tissues probably is RGH3$^{umu}\alpha$. Surprisingly, and contrary to what was observed in Figure 2-2, the C-terminal antibody did not detect the bands is see tissues. Similarly, these proteins were not detected in mutant seedling tissues either. To confirm translation and size of RGH3$^{umu}\alpha$, I also cloned the *rgh3-umu1 sequence* and express it *in vitro* using wheat germ extracts. Proteins were detected with the same N- and C-terminal anti-RGH3 antibodies. As demonstrated by the blotted membranes shown in Figure 4-2B, *rgh3-umu1* produces a protein that is similar in size to RGH3α and is detected with both antibodies. The results indicate that RGH3$^{umu}\alpha$ is being produced and that the amino acids introduced by the *Mu*-element do not considerably alter the protein's electromobility. As mentioned in chapter 2, technical difficulties in the protein extraction step and western blot analyzes prevented further experiments to confirm these results.

To determine the functional consequences of the RGH3$^{umu}\alpha$ protein sequence, I fused RGH3$^{umu}\alpha$ to GFP to examine the sub-cellular localization of the protein by transient expression analysis in *N. benthamiana*. In Figure 4-3A, RGH3$^{umu}\alpha$ localizes to the nucleolus and is also disperse in the nucleoplasm. Data from three independent biological replicates indicate that the recruitment of RGH3$^{umu}\alpha$ to nuclear speckles is impaired when compared to normal RGH3α protein.

In chapter 3, the normal RGH3α protein is shown to interact with U2AF^{65} *in vivo* (Figure 3-4). To determine if the RGH3$^{umu}\alpha$ protein also interacts with U2AF^{65}, I

performed co-localization studies between GFP-RGH3$^{umu}\alpha$ and U2AF^{65}a-RFP fusion proteins in *N. benthamiana*. RGH3$^{umu}\alpha$ is able to partially co-localize with U2AF^{65}a but not as readily as the normal RGH3α (Figure 4-3B). Moreover, RGH3$^{umu}\alpha$ does not always co-localize with U2AF^{65} (Figure 4-3C) and frequently shows a non-overlapping signal similar to that of RGH3 truncated protein variants (Figure 3-3). These data argue that the hypomorphic RGH3$^{umu}\alpha$ protein is likely to be functional but not to the same extent as a normal RGH3α protein. The *Mu*-insertion may also affect protein turnover or expression as suggested by the relative to normal RGH3α. Combined, these data suggest the *Mu* insertion causes altered subnuclear localization and reduced RGH3 protein function.

Rgh3 is Required For a Subset of RNA Splicing Events

In human cells, URP participates in both U2 and U12 splicing and is essential for cell culture survival (Tronchere et al., 1997; Shen et al., 2010). In maize, I showed that RGH3α interacts with U2AF^{65}a, suggesting an orthologous role in splicing between humans and maize. To determine whether RGH3$^{umu}\alpha$ alters splicing, we selected 21 maize genes that are alternatively spliced in maize, Arabidopsis, and rice (Wang and Brendel, 2006b; Table 4-1). I performed semi-quantitative RT-PCR assays with RNA from normal and mutant tissues to compare splicing patterns (Figure 4-4). Only three genes had differences in isoform usage in *rgh3-umu1* and wild-type tissues. GRMZM2G165901 encodes a Gly-rich RNA binding protein with two major splice variants (Figure 4-4A). Sequencing of these isoforms found a noncanonical intron retention with GU-CG dinucleotides at the 59 and 39 splice sites instead of GU-AG. The *rgh3* mutation reduces the level of intron retention in seed and seedling tissue,

suggesting noncanonical splicing is more efficient in the mutant. The smaller variant contained different splice junctions in *rgh3* and the wild-type with *rgh3* shifting the splice acceptor site by three bases.

GRMZM2G051276 encodes a putative inositol monophosphatase with three splice variants (Figure 4-4B). Sequencing of these variants found complex alternative splicing within annotated intron 10. First, a four-base exon is found that results from splicing of two noncanonical introns with UU-AA and AU-UU dinucleotides. Second, an alternative variant contains only a single intron between the annotated exons 10 and 11. Assuming the same donor site is used in the alternative variant, the dinucleotides for the predicted intron are UU-UU, indicating this intron is also noncanonical. Relative to the wild-type, the alternative, retained intron variant accumulates to higher levels in *rgh3* seed tissue but to lower levels in endosperm culture.

GRMZM2G081642 encodes a protein of unknown function, which is conserved within plants and algal species. Sequencing of four variants identified an alternative donor site for intron 2 in which the shorter variant is a GC-AG intron (Figure 4-4C). The longer variant splices with a canonical GU-AG intron, and the *rgh3* tissues have higher levels of this canonical donor site. Intron 3 has three alternative acceptor sites that all have canonical dinucleotides. Combined, these data suggest that RGH3α modulates splice site selection for a subset of non-canonical introns. Furthermore, the data further strengthen the argument that RGH3$^{umu}\alpha$ function is variable but not inhibited.

Discussion

Transposable elements are exceptional mutagenic agents capable of affecting individual genes as well as whole genomes (Bennetzen, 2000). The ability of TEs to

move about the genome inactivating or altering gene expression has become an essential tool for molecular and functional genomic studies in maize (Settles, 2009). In fact, over the last few years efforts to harness the properties of TEs as mutagenic tools have resulted in the creation of multiple large-scale mutagenesis collections that mostly utilize *Ac/Ds* or *Mutator* elements (May et al., 2003; Fernandes et al., 2004; Kolkman et al., 2005; McCarty et al., 2005; Ahern et al. 2009).

The *Mutator* (*Mu*) transposable element is an ancient element and is regarded as the most mutagenic plant transposon studied to date (Brutnell, 2002; Lisch, 2002). *Mu* transpositions have the tendency to occur to single copy or low-copy-number sequences, showing high preference for 5' end and intron regions of gene loci. Additionaly, *Mu*-elements have the ability to move to any chromosome in the genome making it the ideal genetic tool for mutagenesis projects (Xia-Min and Lisch, 2006). The general assumption is that *Mu*-elements, or other TEs, create null alleles of genes upon insertion often revealing a gene's functional importance through altered plant phenotypes. Nevertheless, multiple examples in maize have demonstrated that *Mu*-elements are capable of influencing gene regulation post-transcriptionaly or post-translationaly exposing interesting molecular features of TEs as well as the targeted gene (reviewed in Girard and Freeling, 1999; Weil and Wessler, 1992). The *rgh3-umu1* hypomorphic allele is yet another example of how *Mu*-elements can alter gene expression or protein function. Together with the *Adh1-S3034* allele (Ortiz and Stromer, 1990), *rgh3-umu1* is the second published example showing a conserved splicing pattern. Moreover, our lab recently found an additional example arguing that this event is a more general phenomenon rather than isolated events. Taken together, these

events involving *Mu*-elements argue that this active TE has been able to survive by decreasing its deleterious impact on the host genome and potentially increasing protein diversity. Moreover, as explained in Fouquet et al. (2011), the hypomorphic *rgh3-umu1* allele reveals the developmental importance of *Rgh3.* This insight might not have been possible if *rgh3* had been null allele due to a probable gametophytic lethal phenotype. In conclusion, the data suggest that *Rgh3* may be required in important seed and seedling developmental functions in maize. Most importantly, they demonstrate that the effect of this type of *Mu*-insertion could enhance the study of alleles with fundamental implications in development and that typically are knock-out lethal.

Previously, I demonstrated that RGH3α interacts with the U2AF^{65} protein through the UHM domain (Figure 3-3), a behavior consistent to that observed in humans (Tronchere et al., 1997; Shen et al., 2010). Moreover, even though its function has not been extensively analyzed in humans, data obtained from analyses with RGH3 natural isoforms demonstrate that the RS-domain of RGH3 enhances nuclear localization and is required for recruitment of the protein to spliceosomal speckles and complexes. Based on studies in vertebrates, the RS-domain of U2AF^{35} is required for protein-protein interactions with splicing factors of the SR-family in order to accurately identify the 3' end SS (Pacheco et al., 2004; Mollet et al., 2006). The *rgh3-umu1* hypomorphic allele codes for a full length RGH3$^{umu}\alpha$, with an additional 47 aa in the acidic domain, partially retaining its ability to participate in spliceosomal complexes. Nevertheless, *rgh3-umu1* is seed and seedling lethal, indicating that the RGH3 acidic domain is also necessary for protein function. Even though I can only speculate about its function, it is possible that the acidic domain enhances protein-protein interaction with U2AF^{65} or with

an additional splicing factor to strengthen the recognition of splicing signals. As indicated by studies on the human splicing factor SFSR1 (Cho et al., 2011), regulatory domains of a protein can mask or expose additional functional domains in order to regulate participation of the protein in complexes. Thus, an additional function of the acidic domain could be to regulate the ability of RGH3 to form protein-protein complexes by masking the UHM domain or by regulating the function of the RS-domain.

Finally, despite the severity of the *rgh3-umu1* mutant phenotypes, alternative splicing events are not severely disrupted in *rgh3-umu1* tissues. This indicates that RGH3 is involved in only a few essential regulatory splicing events (Figure 4-4). In plants, similar examples of developmental abnormalities produced by knockdown and misexpression alleles of U2- and U12-specific splicing factors have been observed previously (Kalyna et al., 2003; Wang and Brendel, 2006; Ali et al., 2007; Kim et al., 2010). Likewise, overexpression of proteins involved in splicing have also shown a range of pleiotropic effects that can negatively impact plant development (Lopato et al., 1999; Kalyna et al., 2003). Consequently, proper tissue- or developmental-regulation of essential splicing factors is crucial to ensure adequate protein expression levels and proper plant development. Poor regulation, expression, or stability of the $RGH3^{umu}\alpha$ protein likely compromises its function and causes abnormal seed development. In conclusion, the data suggest that the hypomorphic allele *rgh3-umu1* may not be a rare allele produced by the alternative splicing of the *Mu*-element, exposing a likely adaptation of this element for its continual survival within the maize genome. Furthermore, the lethal nature of the *rgh3-umu1* phenotype indicates that the acidic

domain of RGH3 is necessary for proper protein function in the regulation of a reduce set of alternative splicing events.

Materials and Methods

In vitro Transcription/Translation and Western Blot Analysis of RGH3 Proteins

Rgh3α and *Rgh3*$^{umu1}\alpha$ transcripts were cloned and tested as previously described in the materials and methods section of chapter 2.

Subcellular Localization Studies

Rgh3$^{umu1}\alpha$ transcript was cloned and tested as previously described in the materials and methods section of chapter 2.

RT-PCR Analyses of Alternatively Spliced Maize Genes

Gene candidates were selected from a large list of genes found in the Alternative Splicing in Plants database (ASiP, Wang and Brendel, 2006b) All genes showed alternative splicing evidences in *Arabidopsis* and Rice. Selected genes were blasted against the Maize reference genome (www.maizesequence.org). Repetitive genes, genes with no orthologous in Maize, or genes that are member of a large family were discarded. ESTs from remaining genes were then analyzed for evidences of alternative splicing. In the end, 22 genes were selected for RT-PCR analysis.

Total RNA was extracted from multiple maize normal and *rgh3/rgh3* tissues as described by Reid et al. (2006) using 10 mL of extraction buffer/gr of fresh weight. RT-PCR was performed on first-strand cDNA and synthesized by Superscript III (Invitrogen) using gene-specific primer sets described in table 4-1. Amplified fragments were visualized by electrophoresis on 2% agarose gels (0.5x TBE) and stained in ethidium bromide solution (~0.1 µg/mL).

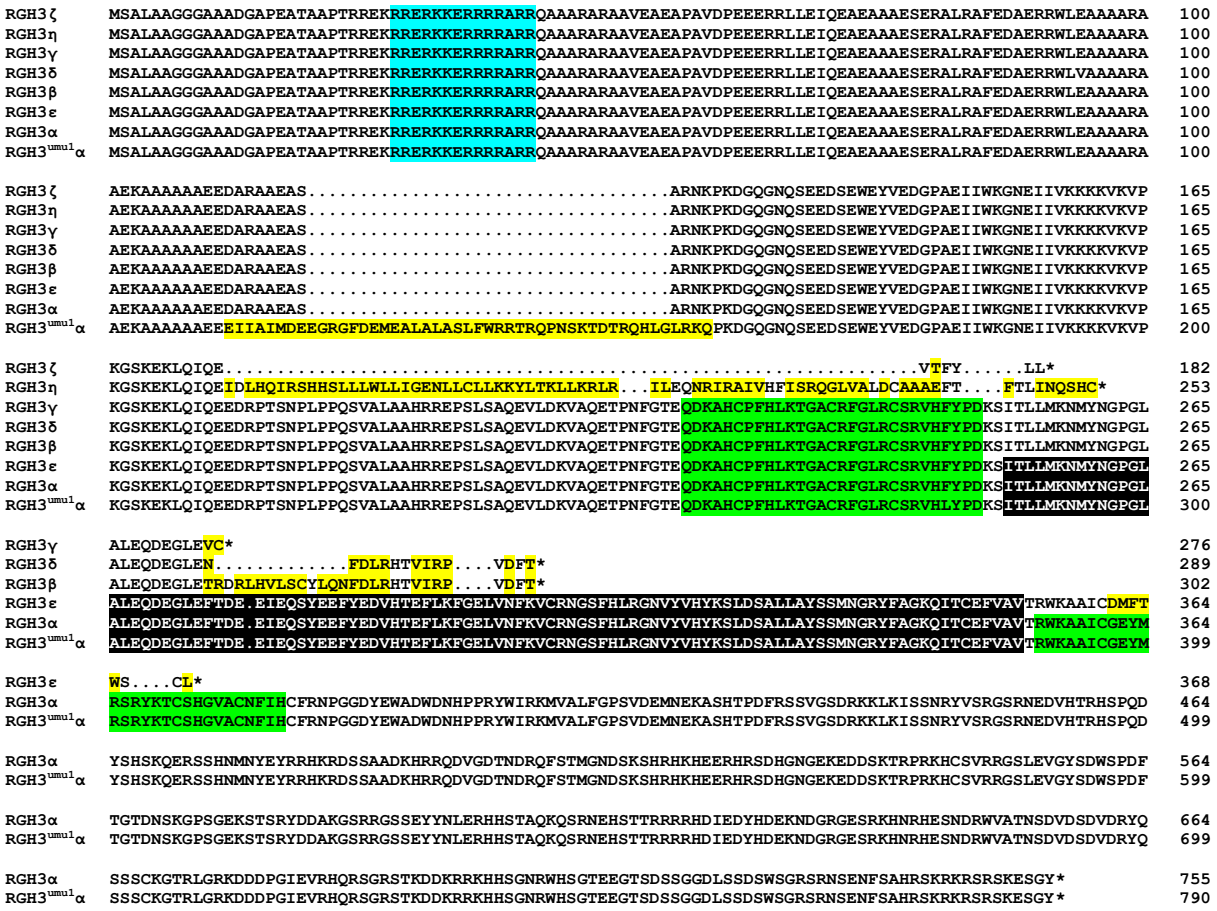

Figure 4-1. Multiple sequence alignment of RGH3 protein isoforms and mutant allele. Greek letters indicate wild-type isoforms, RGH3^{umu1}α is the predicted full-length protein product of the *rgh3-umu1* allele. The nuclear localization signal is highlighted in blue. The zinc finger domains are highlighted in green, and the UHM domain is in white letters with black shading. Residues that are not identical to the RGH3α protein are highlighted in yellow. Conserved domains were identified with Prosite scans (http://www.expasy.ch/tools/scanprosite/). The sequences were aligned with ClustalW2 using default settings (http://www.ebi.ac.uk/Tools/msa/clustalw2/).

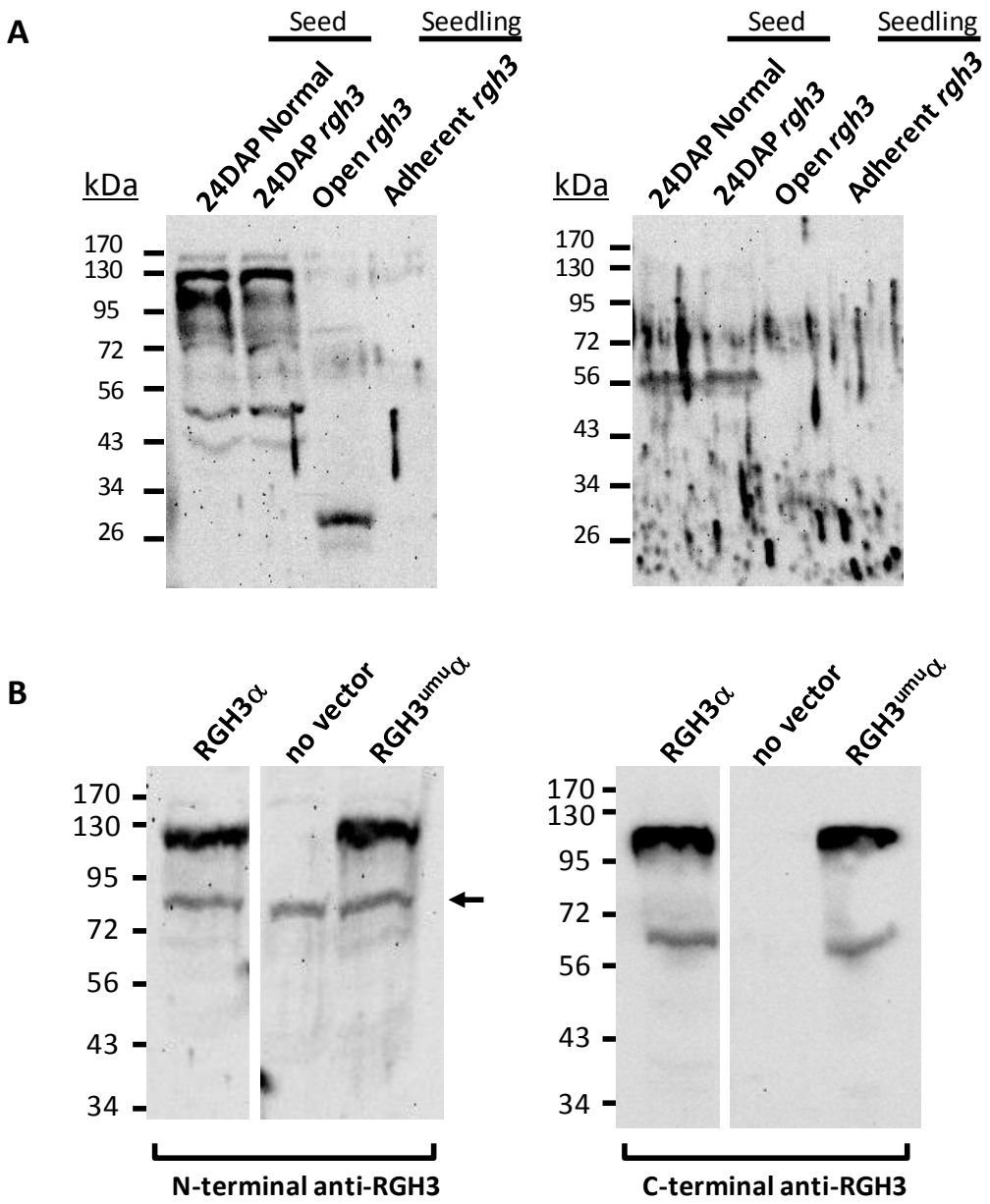

Figure 4-2. *rgh3-umu1* allele produces a full length protein of comparable size to RGH3α protein. Western blot analyses of tissue protein extract (A) and *in vitro* reactions (B) detected with N-terminal (left panel) or C-terminal (right panel) anti-RGH3 antibody. A) The N-terminal Ab detects multiple bands in both normal and rgh3 seed tissues. On the contrary, the C-terminal Ab only detects a single band at ~56 kDa. A single small band was detected in *rgh3* Open seedling by the N-terminal Ab. B) in vitro expressed proteins are of similar size at ~125kDa and are detected with both Ab indicating production of a full length protein RGH3^{umu}α that is found in mutant tissue. Each panel shows two images from a same western blot membrane processed to facilitate comparison between samples.

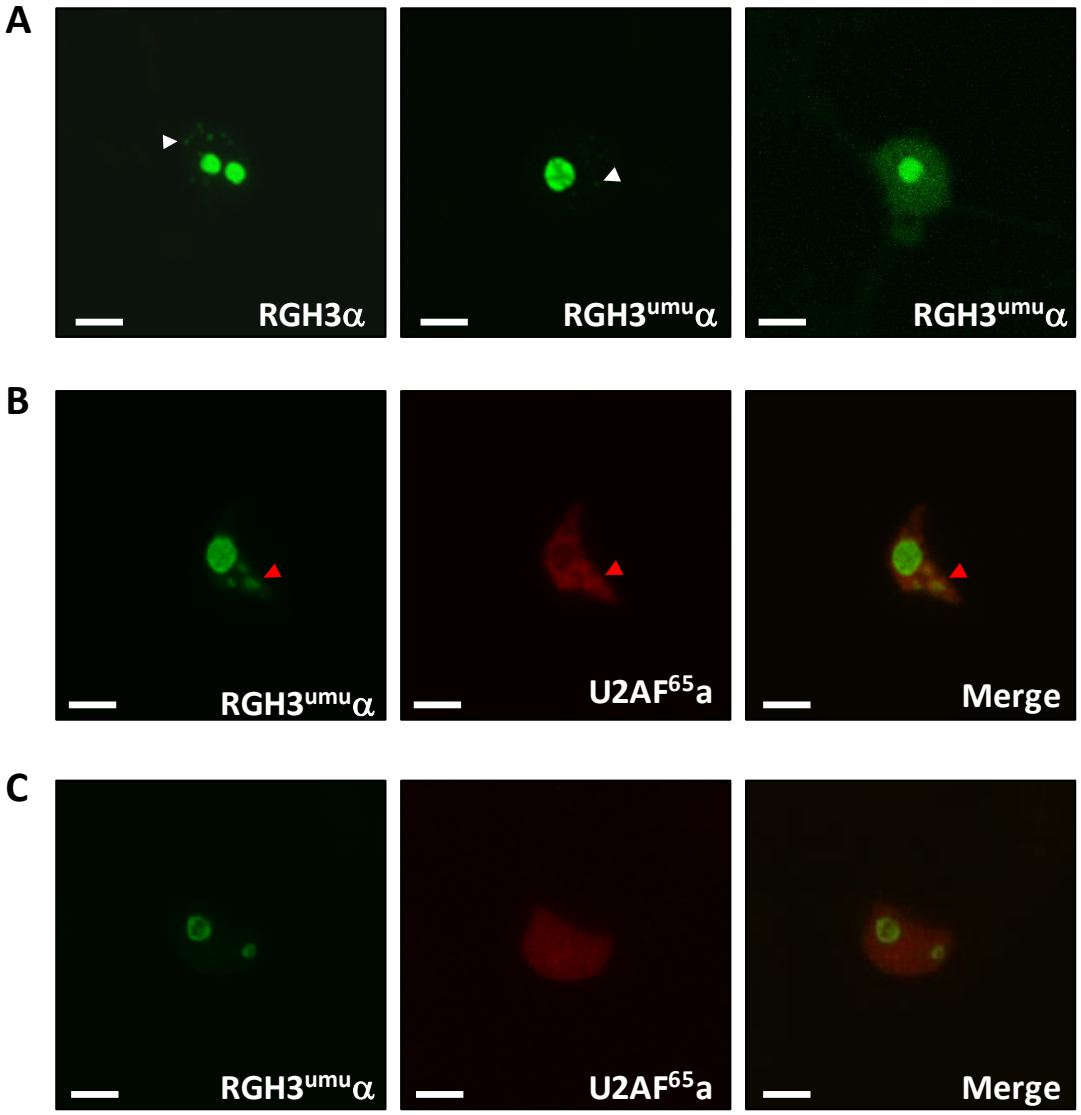

Figure 4-3. RGH3$^{umu}\alpha$ is recruited to speckles and partially co-localizes with U2AF^{65}a. RGH3$^{umu}\alpha$ was fused with GFP and analyzed through transient expression in *N. benthamiana* by itself (A) or with U2AF^{65}a-RFP (B-C). A) Multiple experimental replicates demonstrate that RGH3$^{umu}\alpha$ is consistently found in the nucleolus but transiently localizes to nuclear speckles and not as readily as RGH3α (white arrowhead). B-C) The hypomorphic protein partially co-localizes with U2AF^{65}a in the nucleoplasm (red arrowhead) but not in the nucleolus (B). The co-localization between the two proteins was rarely present and often times no co-localization was observed (C) (scale bar = 5µm).

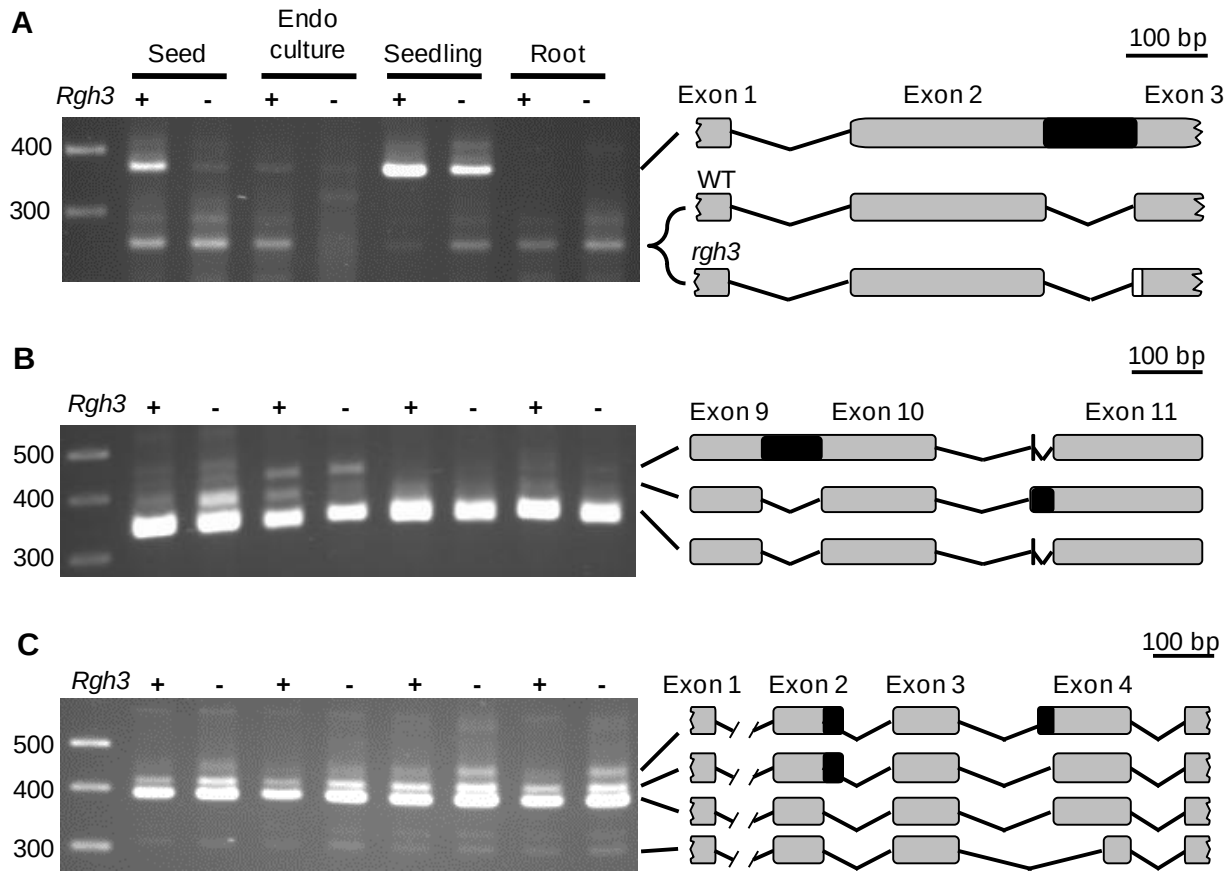

Figure 4-4. RNA splicing defects in *rgh3* detected with semi-quantitative RT-PCR. Schematics show intron-exon structures as determined by cloning and sequencing of the RT-PCR products. Exons are numbered according to the B73 annotation of release 5b.60. Black boxes indicate retained intron sequences. A) GRMZM2G165901 has a novel splice acceptor site in *rgh3* with three additional bases included in the mutant intron (white box). B) GRMZM2G051276 has a four base exon between annotated exons 10 and 11. The retained segment of annotated intron 10 also shifts the splice acceptor sequence by two bases relative to the 4 base exon. C) GRMZM2G081642 has an alternative donor site at the 3' end of exon 2 and an alternative acceptor site at the 5' end of exon 4.

Table 4-1. Maize genes surveyed for splicing defects in *rgh3-umu1* mutants

Rice Homolog	Maize Gene	No. isoforms*	Left primer	Right primer
Qualitative difference between rgh3 and WT isoform levels:				
Os02g07350	GRMZM2G051276	3	TTACGGTCTTCGATCGCTCT	GGAGCTTGGCTCTCTTGAAA
Os02g10440	GRMZM2G081642	3	CCATGATCGAGCAGTTCGT	TTGTCCCGAAGACATGACAC
Os03g46770	GRMZM2G165901	2	AGAATGCCTTCGCCTCCTAC	GAAGCGAACGGTAACACGAT
No qualitative difference detected:				
Os01g03100	GRMZM2G003930	3	CCAGACTTCGAAGCTTCCTC	CAGCTGTTGTGGAGAAGTATAACC
Os01g09120	GRMZM2G176506	>4	CTTAAATGGGCCAGGCACTA	GGGATGGTCTCAACAGGCTA
Os02g10920	GRMZM2G117069	3	GCTGCCCGTACTACCAACC	GGCGCTGCAGACTTTTTAAC
Os02g44230	GRMZM2G112830	3	AGGCGTAAAATCGAGGAGGT	TCTCATCACAGGGGACATGA
Os02g58730	GRMZM2G375002	2	TGTTTTTCTGCTCGCCTTTT	AATTGCTGCGTCAAACACTG
Os03g16140	GRMZM2G073567	1	GCAGATCGCTGCATCAAATA	ATCCCTGCCGTAAGGAGAGT
Os03g20900	GRMZM2G009060	3	AAGAGCAGAAAGGGCATTCA	GCCAAACTCAAAGGACTCCA
Os03g60370	GRMZM2G151967	1	GGAGGCCGAAACAACTCTTA	GGTCTTGACGCCTACAGCTC
Os04g35380	GRMZM2G047705	3	CTACCTGAGCGATGCAATCA	AGTGCAACCCTGTCAAATCC
Os05g02120	GRMZM2G141873	3	GCATCATTCTCTGGGGATGT	AGTGGCTAGTCCACCAGCAG
Os07g30840	GRMZM2G171745	>4	AGCTGATCGCGCAGTTCT	CGCTGAGCTGCTTTAGCTTT
Os07g42660	GRMZM2G048846	3	GGTCTTCGTGAAAACCATCG	AGCCTCTTGCACCATGTCTT
Os07g47630	GRMZM2G436092	3	CCCTAGCCTCGAGCTCTATC	GCAGCCTCCACTCTTAGACG
Os11g34210	GRMZM2G091433	3	GCAAGCAGATGAGGATGTTG	TTCATCTCCACAGGAATCAGC
Os11g35870	GRMZM2G118385	2	GCGGTGTACGGAGACGAC	CAAGGCCCTTACTTTCCACA
Os12g29990	GRMZM2G420055	3	GGCTCTTGACGATGCAAAAT	TCTTTATGCTGTCGCATTGG
Os12g37970	GRMZM2G130149	>4	GAGAGAGGAGACTCGCAAGG	GAGGACGACGATGGAGACAT
Os12g43600	GRMZM2G080603	3	CGCAACATCACCGTCAAC	ACACAGATGGGCAACAACAA

*Observed by RT-PCR in 14DAP seeds, endosperm culture, 10-12 days after planting seedling shoot and seedling root

CHAPTER 5
MAPPING OF THE *rgh3* SEEDLING GENETIC MODIFIER AND THE *dek*9700* LOCUS

Introduction

The maize *Rough endosperm3* (*Rgh3*) gene shows significant homology to the human URP splicing factor The human gene was shown to be required in splicing activities *in vitro*, interact with U2AF^{65}, and participate in both U2- and U12 type spliceosomal complexes (Tronchere et al., 1997; Will et al., 2004; Shen et al., 2010). Originally identified by Fajardo (2008), the maize *rgh3* mutant affects seed and seedling development. In seeds, the *rgh3* allele shows a range in the severity of seed phenotype which includes cell differentiation defects for basal endosperm transfer cell layer (BETL) and the embryo surrounding region (ESR) (Fouquet et al., 2011). About 50% of *rgh3* seeds are able to germinate in soil, which correspond to the heaviest, least severe class of seed phenotype. However, *rgh3* seedlings are lethal at about 15 days after germination. Surprisingly, mutant endosperms are able to proliferate more readily than normal endosperms in tissue culture indicating that *Rgh3* is required to repress cell proliferation (Fouquet et al., 2011). I showed that *rgh3-umu1* is a hypomorphic allele and that a small number of alternative RNA splicing events are affected in the mutant. Moreover, I found that *Rgh3* transcript is alternatively spliced and that RGH3α protein is able to interact with U2AF^{65}, consistent with the behavior of human URP. Taken together, the data strongly indicate that RGH3 is the ortholog of human URP. The *rgh3-umu1* allele provides a unique protein variant that can identify the targets and interacting proteins that are most sensitive to altered URP/RGH3 function. These gene targets or interacting proteins are expected to help explain the *rgh3-umu1* phenotype and give insight into the function of the RGH3 acidic domain.

Maize (*Zea mays*) is a highly diverse plant species. It comprises several hundred inbred lines as well as landraces showing a tremendous level of variation in morphological traits as well as polymorphism at the DNA level. Maize is thought to come from a single domestication event (Matsuoka et al., 2002), yet the level of genotypic variation translates into a rich source of natural allelic variation that can modify mutant phenotypes. Inbred or landrace genetic modifiers can sometimes be the product of multiple factors, or the effect of single genes, which interact or are targets of a mutant locus (Lopes et al., 1995; Vollbrecht et al., 2000).

Map-based or positional cloning is a robust approach to uncover the genetic cause of a phenotype. Positional cloning of mutants from defined genetic backgrounds can be simplified relative to mapping from undefined or mixed backgrounds (Jander et al., 2002). The first step of positional cloning is identifying a map position to bin level resolution. This is most efficiently done using a core set of distributed markers (Bortiri et al., 2006). The two most widely used molecular marker systems in plant breeding and genetics are SNPs and SSRs (Appleby et al., 2009). In contrast to SNPs, SSRs are more amenable to small scale experiments such as a mapping a single mutant using bulk segregant analysis (BSA) (Gallavotti et al., 2008; Thompson et al., 2009). Individual SSR markers can be more easily added or removed from a mapping experiment and a smaller number of samples are generally analyzed. The cost of SSR mapping can be further reduced by identifying marker sets that are most useful for the specific mapping experiment. BSA mapping requires co-dominant markers. The Maize Mapping Project characterized hundreds of SSRs and deposited images of inbred screening gels on the MaizeGDB database (Sharopova et al. 2002). These screening

gels included F_1 or mixed DNA samples to test for co-dominant alleles for only 1–2 inbred pairs of the 11 inbreds screened. Similarly, Fu et al. (2006) developed and characterized insertion–deletion polymorphism (IDP) markers within 24 inbreds but did not screen mixed or F1 DNA samples. In addition, the IDP markers were not screened for the predominant inbred used in public mutagenesis projects, W22. Our anecdotal experience suggested that size differences in SSR alleles were not sufficient to predict co-dominant markers.

The *rgh3-umu1* allele was originally isolated in the maize W22 background. Though, when crossed into other inbred backgrounds such as B73 or Mo17 a modified seedling phenotype was observed. I decided to identify the seedling phenotype genetic modifier locus through a map-based cloning approach. In order to make the mapping more efficient, and due to the above mentioned issues, I analyze 505 SSR markers for co-dominant polymorphisms between the B73, W22, and Mo17 maize inbred lines to identify a distributed marker set. The distributed marker set allows efficient mapping of mutants from B73 and W22 public mutagenesis experiments. The effectiveness of the marker set was further demonstrated by the cloning of a *dek* mutant (*dek**9700*) from the UniformMu mutagenesis population (McCarty et al., 2005) by BSA.

Results

A Dominant *rgh3* Seedling Modifier is Present in the B73 Maize Inbred

To facilitate the mapping of the *Mu*-element linked to the *rgh3* mutant phenotype, several F_2 mapping populations were generated in maize inbred backgrounds including B73 and Mo17 (Fajardo, 2008). In the B73 F_2, the normal to mutant seed ratio deviated from an expected 3:1 to about 15:1. In the W22 background, a fraction of mutant seeds germinate and exhibit aberrant development with adherent, narrow leaves resulting in

hooked seedlings that died 15 to 18 days after planting (Figure 5-1). Interestingly, mutant seedlings in F_2 populations from crosses to B73 also showed stunted growth but were characterized by an open, round leaf morphology (Figure 5-1). Due to the unexpected low frequency of mutant seeds, I decided to plant normal looking seeds to test for the presence of *rgh3/rgh3* seedlings in this population. Both types of mutant seedlings, open leaves or the adherent, hooked phenotype were observed. The distortion of the normal to mutant seed ratio as well as the observed difference in seedling phenotype suggests the presence of genetic modifiers of *rgh3* in the B73 background. To test this hypothesis, at least at the seedling level, I back-crossed segregating F_1 progeny from a B73 X *rgh3*/+ cross with a *rgh3*/+ W22 tester parent line to create a BC_1 population (Figure 5-2A). Seeds from six segregating ears were separated between normal and mutant groups and 148 *rgh3/rgh3* seeds were planted. About 55% of the seeds germinated and their seedling phenotypes were scored (Table 5-1). The phenotypic difference between modified looking seedlings was named open, and unmodified was referred as adherent. The modified phenotype showed a gradient in severity (Figure 5-2B). Consequently, seedlings were grouped into two main categories, open or adherent, which were then sub-divided between a high and low-confidence phenotype (Table 5-1). Overall, the final ratio between both seedling phenotypes was close to 1:1, which suggests a single dominant seedling modifier from the B73 parent. In order to map this modifier locus, I needed molecular markers for mapping in B73 and W22 inbred parents.

Development of a Distributed SSR Marker Set

Each of the 505 SSR markers was tested for co-dominant polymorphisms between B73/Mo17, B73/W22, and Mo17/W22. A co-dominant marker was defined as useful for

F_2 mapping when it amplified easily resolved size polymorphisms between at least two inbreds and showed a novel banding pattern in a 1:1 mix of a pair of inbred DNA samples (Figure 5-3A, umc1538). Based on these criteria, 238 (47.1%) of the markers amplified co-dominant markers with each pair of inbreds having 154–170 useful markers (Figure 5-3B; Table 5-2). Less than 10% of the 505 markers had three distinct alleles for the inbreds, such as umc1538 (Figure 5-3A). 31.9% of the markers had two alleles, while 8.5% of the markers showed co-dominance in just one of the three inbred pairs. The remaining 267 markers that were not useful for BSA gave the following amplification patterns: 76 amplified a single allele in all inbreds (Figure 5-3A, umc1288); 129 amplified alleles that were difficult to resolve or score (Figure 5-3A, umc1590 and phi402893); and 62 failed to amplify.

I selected a distributed, co-dominant marker set that could be used for all of the inbred pairs. For this analysis, I divided the IBM2 2008 Neighbors map coordinates by a factor of 4 to account for the genetic expansion of the IBM population (Lee et al., 2002). This is the average conversion factor for marker coordinates between the Genetic 2008 and IBM2 2008 Neighbors maps at MaizeGDB. Using these predicted genetic distances, the average distance between all polymorphic markers for each inbred pair was <14 cM (Table 5-2). A minimally redundant set of distributed markers would be spaced at 50 cM intervals to ensure that all phenotypic variations can be detected by at least one marker. Based on the size of the IBM2 2008 Neighbors map, 50 markers would be needed for a minimal distributed set. However, only 43 markers that contained three distinct alleles in B73, Mo17, and W22 were observed and these markers are not distributed uniformly. To account for these issues, I selected a distributed set of 85

markers that includes some redundancy (Figure 5-4). 64–71 of the distributed markers are polymorphic for each inbred pair, and these polymorphic markers are spaced at a mean genetic interval of 27–29 cM (Table 5-2).

The distributed markers contain some gaps in which the closest marker is >25 cM (Figure 5-4). The majority of the gaps in coverage are at the ends of chromosomes with the largest gap located on the long arm of chromosome 8. This gap is likely to be caused by an error in the IBM2 2008 Neighbors map. The terminal locus of this arm is annotated as *Empty pericarp4* (*Emp4*) and adds approximately 46 cM of genetic distance to the map. *Emp4* has been mapped to chromosome 1 via translocations, molecular markers, and molecular cloning (Gutierrez-Marcos et al., 2007). After excluding the *Emp4* locus from the map, 7.3–9.6% of the genome is predicted to be located >25 cM from a marker depending upon the inbred pair (Table 5-2).

Mapping of *rgh3* Seedling Genetic Modifier

To map the genetic modifier of *rgh3* seedling phenotype, I expanded the BC_1 mapping population to increase the number of meiotic products available for testing by BSA. Mutant seeds, and also seeds from the normal-looking group were planted and genomic DNA was extracted from germinated mutant seedlings. Individual DNAs from adherent and open *rgh3/rgh3* seedlings were pooled by groups based on their phenotype. Both DNA pools were tested with the distributed marker set under identical PCR conditions along with a control 1:1 mix between W22:B73 DNA. Since the population was back-crossed to the W22 parent but the modifier is dominant in the B73 background, modified open individuals should be heterozygous at the modifier locus. Thus, an increased number of markers should become heterozygous in the modified open DNA pool as they map closer and become linked to the modifier location. The

adherent DNA pool should show the inverse distortion towards W22 for the same markers indicating absence of the domain modifier allele. Out of all tested markers throughout the genome, markers mapping to the short arm of chromosome 9 showed the most distortion towards B73. Therefore, I began fine mapping analyses with individual DNA samples and tested them with SSR markers showing distortion in that area. A marker located at ~13.2Mb and at ~5.3Mb (umc1170 and umc1867, respectively) indicated B73 allele enrichment towards the proximal end of the chromosome (Figure 5-5A). To analyze the complete area, I designed and tested additional SSR markers at 0.5, 8.2, 9.7, and 10.9Mb. Enrichment for the B73 allele was observed in the segment between markers located at 8.2 and 10.9Mb indicating mapping of the modifier locus (Figure 5-5B). All three markers in the region showed very similar recombination frequencies preventing fine mapping.

Mapping of *dek*9700* from UniformMu Population

To test the effectiveness of the distributed set to map a mutant allele from public mutagenesis population, I completed BSA for the *dek*9700* mutant from the transposon tagging UniformMu population developed at the University of Florida (McCarty et al., 2005). Ideally, this mapping experiment should also to test the reliability in which a marker can detect different allele in a mixed DNA F_2 background and if this information is enough to distinguish between a linked versus an unlinked marker. That is, it will demonstrate if the alleles from markers in the distributed set can accurately be distinguished when the marker is linked to a locus of interest. Then, to complete the mapping experiment, I tested Mo17/UniformMu and B73/UniformMu F_2 mapping populations and found segregation distortion on chromosome 6 in both populations.

PCRs of *dek* mutant individuals from the Mo17/UniformMu population were used to refine the map position (Figure 5-6A). Some of the *dek* mutant kernels in this experiment were exceptionally small, and I extracted DNA from whole dried kernels. Surprisingly, these had relatively little maternal DNA contamination, and recombinants could be scored with reasonable confidence (Figure 5-6A, lanes 34–45). These experiments mapped the *dek* mutation to a 10.6 cM interval between umc1063 and umc1653 (Figure 5-6B). The predicted genetic distance between these markers is 20.5 cM suggesting that the conversion factor between IBM2 2008 Neighbors map and the Genetic 2008 consensus map coordinates may be an underestimate. We scored the mapping population for bnlg345 to obtain additional recombinants. When bnlg345 is included, the total interval observed between bnlg345 and umc1653 is similar to the expected distance with 32.3 cM observed versus 39 cM expected. These data suggest the IBM 2008 Neighbors map provides accurate map distance estimates when the W22 inbred is used as a mapping parent. This mapping experiment was able to place the *dek*-9700* mutant isolate to an approximately 10 cM interval in bin 6.07 indicating that the markers provide a robust resource for mapping UniformMu mutants.

Discussion

The maize RGH3 protein is the homolog of human URP, a protein involved in assembly and function of the U2 and U12 spliceosome (Tronchere et al., 1997; Shen et al., 2011). *Rgh3* is required for proper endosperm-embryo development in the maize seed by influencing endosperm cell differentiation (Fouquet et al, 2011). Identifying additional protein interactors and/or target genes could greatly enhance our ability to understand the biochemical processes that are most sensitive to reduced *Rgh3* function. Genetic modifiers can provide valuable information towards understanding the

biochemical bases of the trait (Lopes et al., 1995; Vollbrecht et al., 2000). Maize is a highly diverse organism providing a vast array of natural genetic variation that can be used to identify genetic modifiers. The hypomorphic *rgh3-umu1* allele was originally discovered in the W22 background where it shows a variable range in seed phenotype (Fajardo, 2008). Interestingly, when crossed into the maize B73 inbred background, the seed phenotype becomes partially suppressed and the seedling phenotype has an open, rounded leaf morphology (Figure 5-1). Through mapping experiments using a distributed SSR marker set I was able to localize the modifier loci to a 3Mb region on the short arm of chromosome 9 (Figure 5-5B). Further fine mapping was not possible due to conflicting recombinant individuals in the open and adherent phenotypic classes, which is possibly due to the range of seedling phenotypes. Overlap between the expressivity of the B73 and W22 phenotypes may lead to scoring errors and mis-assignment of individuals to the incorrect genotypic class. Mis-assignment of open and adherent genotypes would decrease the resolution within the linkage region.

The specific cause of the variable, modified seedling phenotypes is not known. It is possible that the hypomorphic *rgh3-umu1* allele contributes to the modified seedling phenotype range. In the W22 background, *rgh3-umu1* shows a variable range of phenotypes and I have shown that the $RGH3^{umu1}\alpha$ protein also has variablesub-nuclear localization as well as variable levels of co-localization with $U2AF^{65}$.. Alternatively, the variability could be produced by the modifier locus rather than the $RGH3^{umu1}\alpha$ protein. However, the seedling phenotype variability was also observed in advanced back-crossed populations, such as BC_4 where the background is almost entirely W22 (data

not shown) supporting the idea that variability in RGH3$^{umu1}\alpha$ function is the more likely cause of the observed range in seedling phenotype.

One possible approach to obtain a higher resolution map position for the rgh3 modifier would be to use the extreme or high confidence open and adherent seedling phenotypes. Currently, the available number of meiotic products in the high confidence groups is too few to improve resolution. It would be necessary to expand the mapping population to increase the number of available high confidence meiotic products. However, the time and effort required to generate the larger population would make the approach inefficient to clone the locus. Other approaches are possible to find targets and interacting proteins, such as protein complex pull-down assays or RNA immunoprecipitation assays. These biochemical approaches are more likely to provide mechanistic information in a more cost-effective manner.

The development of common mutagenesis resources in the B73 and W22 inbreds creates the opportunity for map-based cloning in well defined genetic backgrounds. By using defined inbreds, a smaller number of markers can be selected to ensure that mutants are mapped in a BSA (Liu et al. 2009). A SNP marker system has been developed for BSA mapping of maize mutants. The authors recommend 1,016 markers be used for mapping mutants in uncharacterized inbred combinations such as those involving W22 (Liu et al. 2009). The large number of markers required, high initial set-up costs, and requirement to complete hundreds of BSA mappings before SNP genotyping becomes cost-effective makes this technology less accessible for small research groups that are interested in mapping a few mutants at a time. Consequently, I focused on

identifying distributed markers with technology that is more suitable for lower throughput genotyping.

Despite the large number of maize molecular markers already developed and characterized, it is not simple to identify distributed, co-dominant markers for pairs of inbreds. The presented data suggest approximately 30% of mapped SSRs will be useful for BSA of any given pair of divergent inbreds. This is twice the frequency at which maize SNP markers are expected to produce quantitative co-dominant markers for a pair of divergent inbreds (Liu et al. 2009) and is consistent with the higher information content of SSR loci (Hamblin et al. 2007). It is important to note that the analysis was restricted to 4% TBE agarose gel electrophoresis. An additional 25% of the markers analyzed in this study amplified products that could potentially be scored using higher resolution agarose and capillary electrophoresis techniques.

The distributed markers I selected provide a simple marker technology to enable mapping of mutants from public mutagenesis resources. By surveying for W22 polymorphisms, mutants from multiple public transposon mutagenesis populations and one EMS mutagenesis population can now be mapped more readily (Cowperthwaite et al., 2002; Till et al., 2004; Kolkman et al., 2005; McCarty et al., 2005; Ahern et al., 2009). Although I was not able to precisely localize the *rgh3* modifier locus, the usefulness of the marker set was demonstrated by mapping the *dek*-9700* mutant isolate to an approximately 10 cM interval in bin 6.07. Two seed mutants, *su2* and *dek*-1104*, have recombination data that place these loci >20 cM proximal to *dek*-9700* (Scanlon et al. 1994). Nine other seed mutant isolates including 2 named loci, *dek19* and *emb3*, have been mapped to the long arm of chromosome 6 using B-A

translocations (Chang and Neuffer 1994; Scanlon et al. 1994; Neuffer and England 1995; Heckel et al. 1999). Complementation tests with *dek19* and *emb3* will be needed prior to assigning *dek*-9700* a locus name.

Finally, I estimate that 7.3–9.6% of the genome, depending on the inbred combination, is likely to be outside the range of the distributed marker set to detect linkage in a BSA. Thus, it is expected that 90–93% of mutants should be able to be mapped with a single F_2 mapping population using the marker set reported here. However, segregation distortion for many SSR markers is readily apparent when one allele constitutes 75% of the DNA sample (Carson et al. 2004; Jones et al. 2007), which corresponds to a Haldane genetic distance of 35 cM. Assuming this upper bound is a realistic expectation for BSA, the coverage of the distributed markers will be greater and the success rate in using this marker system should approach 96–98%. The frequency of success could be increased further by generating two F2 populations with crosses between all three inbreds.

Materials and Methods

Plant Material

Back-crossed mapping populations were designed as described by scheme on figure 5-2. Seeds from BC_1 progeny from *rgh3/+ W22 tester* x F_1 crosses (figure 5-2) were scored visually for normal or *rgh3/rgh3* phenotype and were then planted in soil under controlled conditions. Two week old mutant seedlings were scored for open or adherent phenotypes and DNA was extracted from seedling leaf as described by Settles et al. (2004). For BSA analysis, samples were pooled to create open and adherent bulks. Individual DNA was tested with linked SSR markers under identical conditions described below and recombinants were scored. For SSR markers testing, genomic

DNA from the B73, W22, and Mo17 inbreds was extracted from 2 week old seedlings as described (Settles et al., 2004).

The proof-of-concept mapping experiment was completed by crossing a UniformMu *defective kernel* (*dek*) isolate, *dek*-9700*, to the B73 and Mo17 inbreds. F_2 mapping populations were generated by self-pollinating F_1 plants. Mature mutant and normal seeds were selected visually from segregating ears of the B73/UniformMu and Mo17/UniformMu populations. For DNA extraction, the seeds were imbibed in water overnight, and the pericarp was removed. For normal seeds, the embryo was dissected for the extraction. For the *dek* mutant seeds, the entire endosperm and embryo was used in reduced grain-fill mutants. In the Mo17/UniformMu population, 11/42 *dek* kernels had very severe grain fill defects, and the pericarp was included in the DNA extraction for these kernels. DNA extraction was completed as described (Settles et al., 2004), except that 1 mL of extraction buffer was used and the homogenized sample was centrifuged prior to phenol:chloroform:isoamyl alcohol extraction. This step sedimented the starch gel formed by the DNA extraction buffer. The supernatant was transferred to a new microcentrifuge tube prior to completing the remaining steps of the extraction. For the BSA PCR, the individual samples were pooled to create normal and mutant bulks. The samples were amplified with the distributed marker set. Individual mutant DNA samples were tested by PCR with the linked SSR markers, and map distances were calculated using Haldane's map function (Haldane, 1919).

SSR Testing and Scoring

Purified DNA was diluted to a concentration of ~20 ng/µL just prior to PCR. Each SSR marker was amplified from the three inbreds and from 1:1 mixes of B73/W22,

Mo17/W22, and B73/Mo17 DNA. SSR markers were tested from a commercial primer set of 480 primer pairs with 477 non-redundant markers (M8818-1SET, Sigma-Aldrich Co., St. Louis, MO). The remaining 28 primer pairs were selected to fill gaps in the distributed sets from the IBM2 2008 Neighbors Frame 2 genetic map at the MaizeGDB database (http://www.maizegdb.org). Primer sequences for these markers are available at MaizeGDB. Each marker was tested under common PCR conditions (0.25 µM primers, 150 µM each dNTP, ~40 ng DNA, and GoTaq® PCR mix, Promega Co., Madison, WI). Thermocycling conditions were 94°C for 40 s, 57°C for 45 s, and 72°C for 40 s with 34 cycles. Amplified fragments were visualized by electrophoresis on 14 cm, 4% agarose gels (0.5% TBE) at 90 V for ~2h and stained in ethidium bromide solution (~0.1 µg/mL). Co-dominant polymorphisms were scored visually from the gel images. If a marker failed to amplify or gave multiple products within an inbred DNA, the PCR was repeated 2–3 times to confirm the amplification pattern.

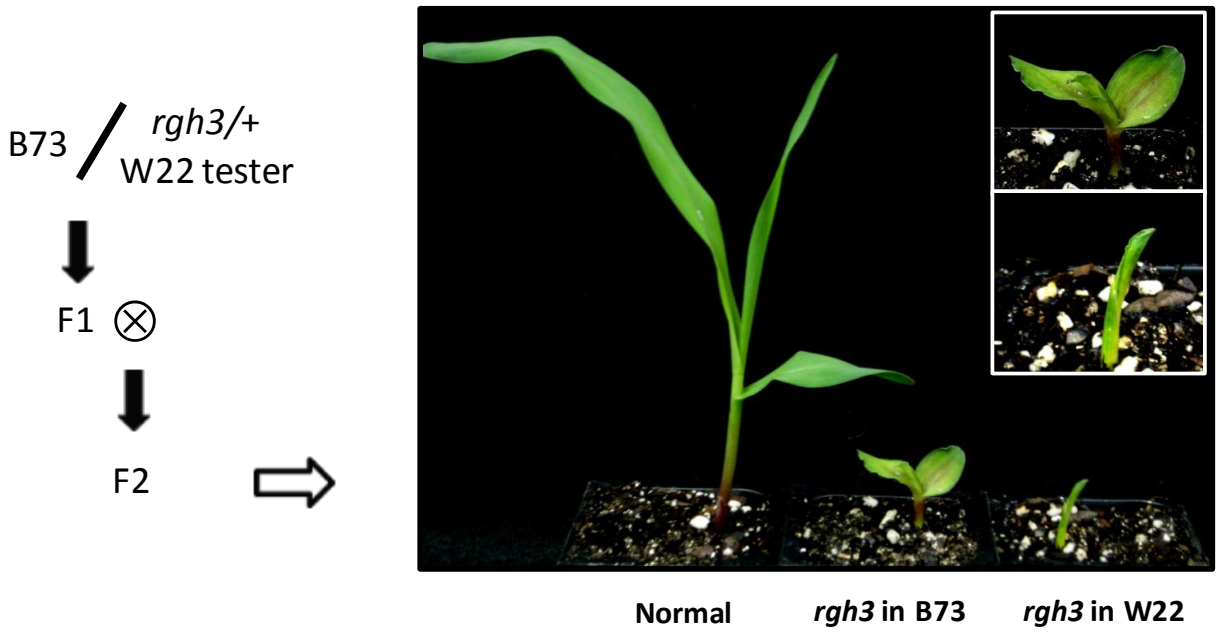

Figure 5-1. Abnormal *rgh3* seedling phenotype on multiple maize genetic backgrounds. *rgh3-umu1* was isolated in the W22 maize inbred background and was crossed to B73 inbred to create mapping populations (left panel scheme). In W22 background, the *rgh3* seedling phenotype is characterized by an stunted growth, and adherent, hooked leaf morphology. When crossed into B73 inbred, a modified seedling showing stunted growth, open leaf morphology was observed (images on the right).

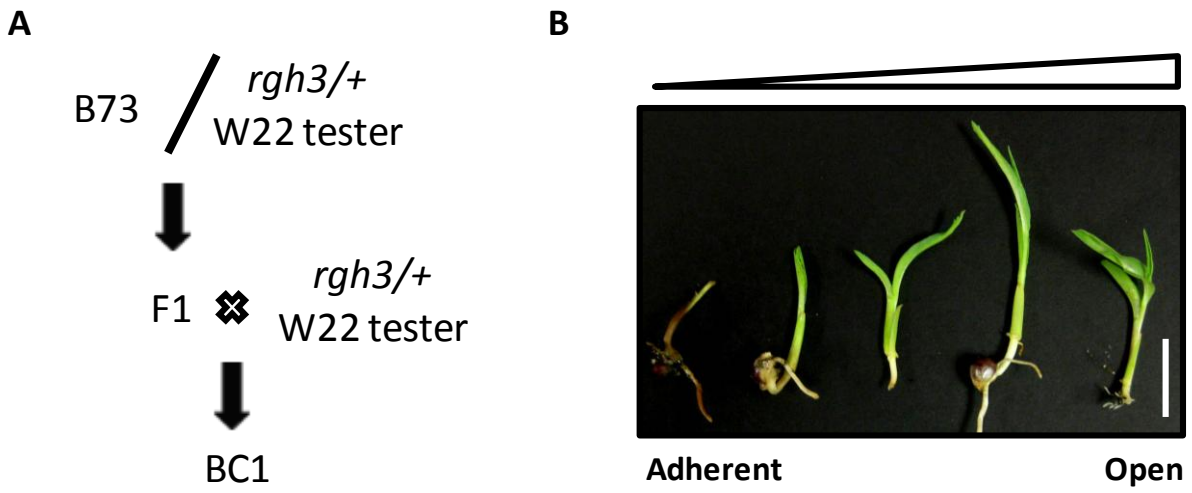

Figure 5-2. Gradient of abnormal *rgh3* seedling phenotype. A) Left panel scheme shows crossing design to create introgressions of lines segregating for the modifier loci into a W22 *rgh3*/+ tester line. B) Gradient of abnormal *rgh3/rgh3* seedling phenotypes observed in BC1 population. Adherent denotes unmodified phenotype; open denotes modified phenotype (scale bar = 0.5in).

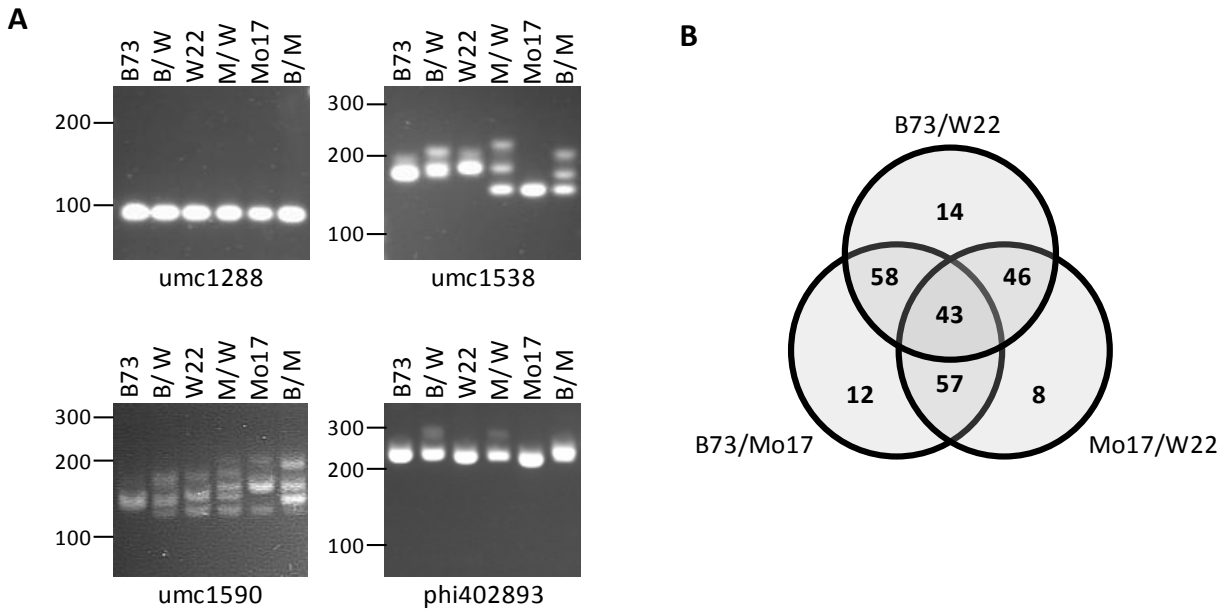

Figure 5-3. Screen for polymorphic SSR markers in B73, Mo17, and W22. A) Examples of PCR scores: umc1288 amplifies a single allele; umc1538 has three co-dominant alleles; umc1590 and phi402893 have polymorphic alleles that are not suitable for BSA. DNA size markers (bp) are indicated for each image. B/W, M/W, and B/M indicate 1:1 mixes of B73:W22, Mo17:W22, and B73:Mo17 inbred DNA. B) Venn diagram showing the number of co-dominant markers found for each pair of inbreds.

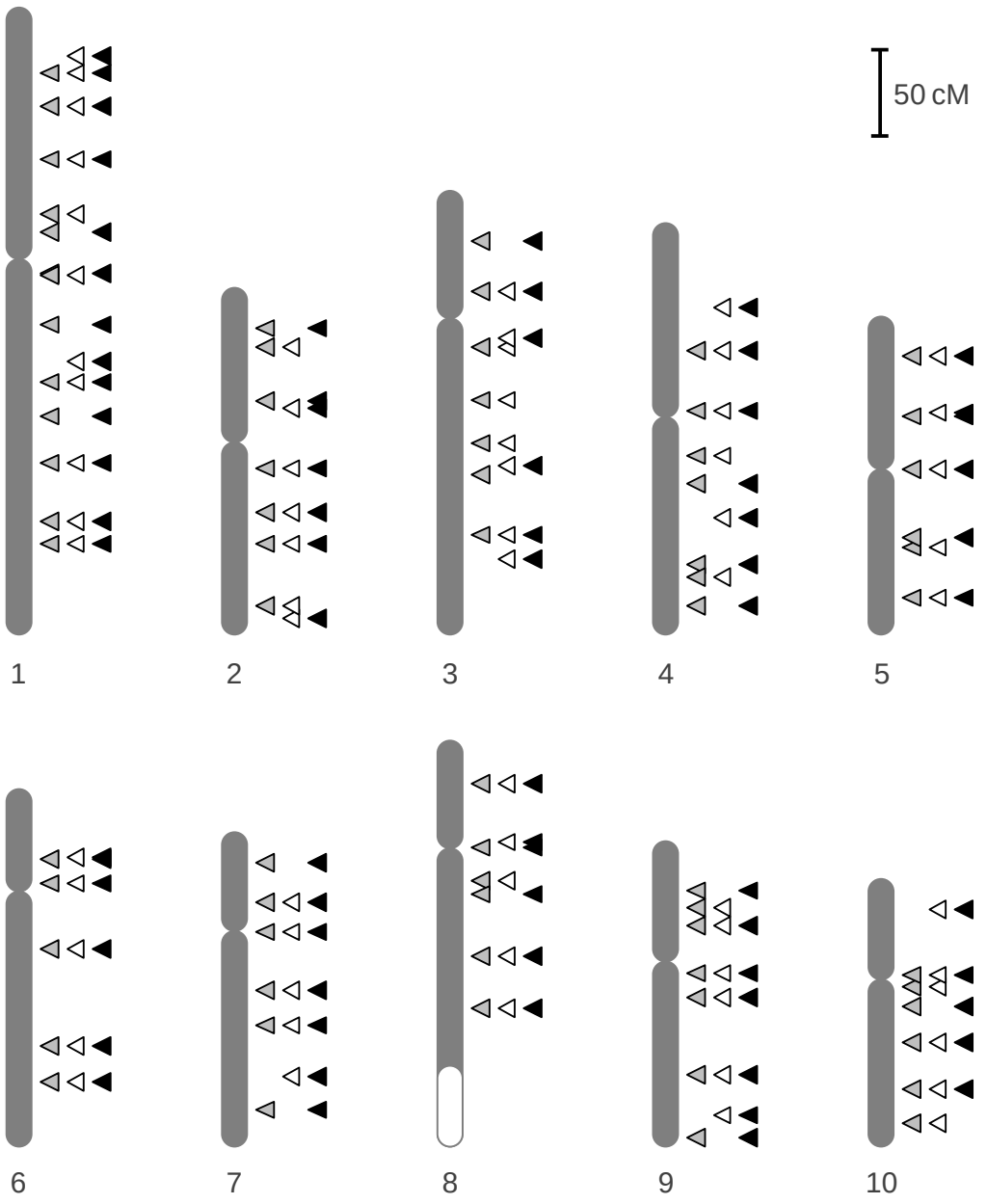

Figure 5-4. Genetic map of the distributed marker sets. Markers are positioned based on their genetic coordinates on the IBM2 2008 Neighbors genetic map. Gray, open, and black arrowheads are polymorphic markers for the B73/Mo17, B73/W22, and W22/Mo17 inbred pairs, respectively. The open oval on chromosome 8 indicates the genetic distance added to the map due to misplacement of *Emp4.* Supplementary table 2 gives map locations and expected co-dominant polymorphisms for the distributed marker set.

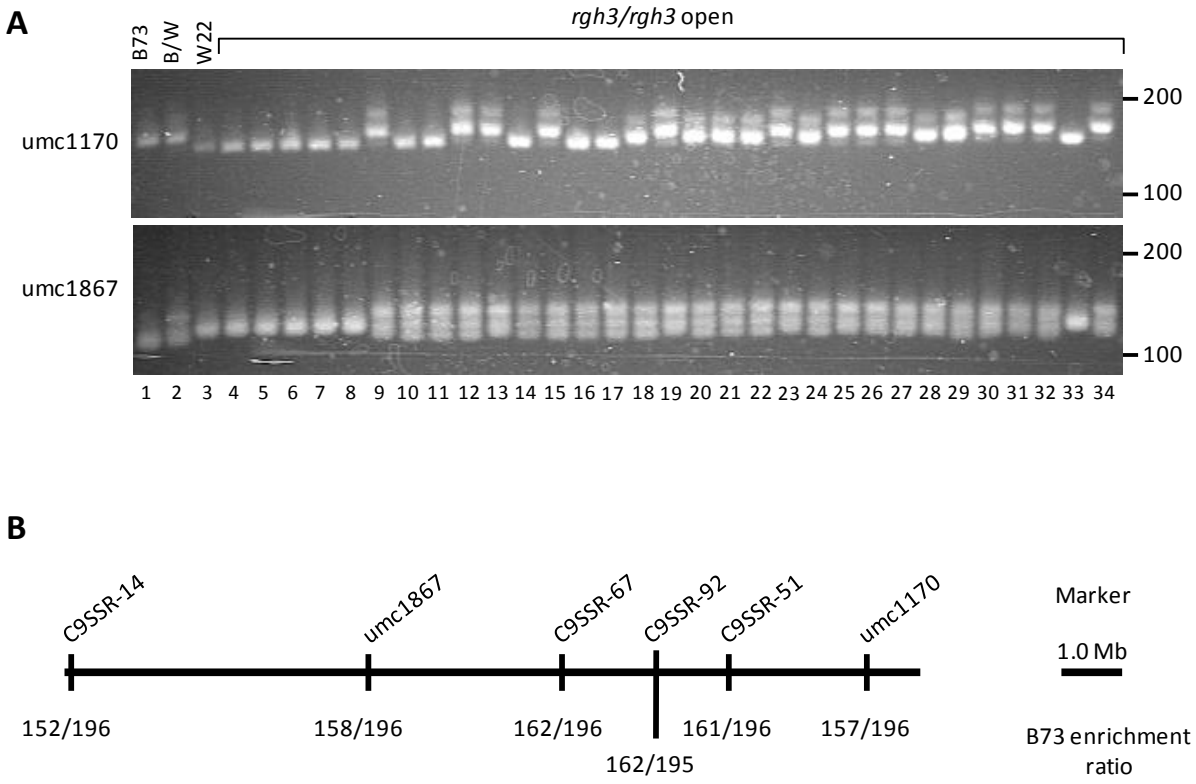

Figure 5-5. Mapping of the *rgh3-umu1* seeding genetic modifier. A) Gel images demonstrate increasing recombination frequency of heterozygous open individuals from umc1170 located at 13.2Mb towards umc1867 found at 5.3Mb. Lanes 3 to 8 show individuals showing no recombination indicating likely phenotyping mis-assignments. B) Physical map of the short arm of chromosome 9 showing B73 heretozygous enrichment by multiple SSR markers tested in the area. Markers C9SSR-67, C9SSR-92, and C9SSR-51 showed the highest distortion towards B73 mapping the modifier locus.

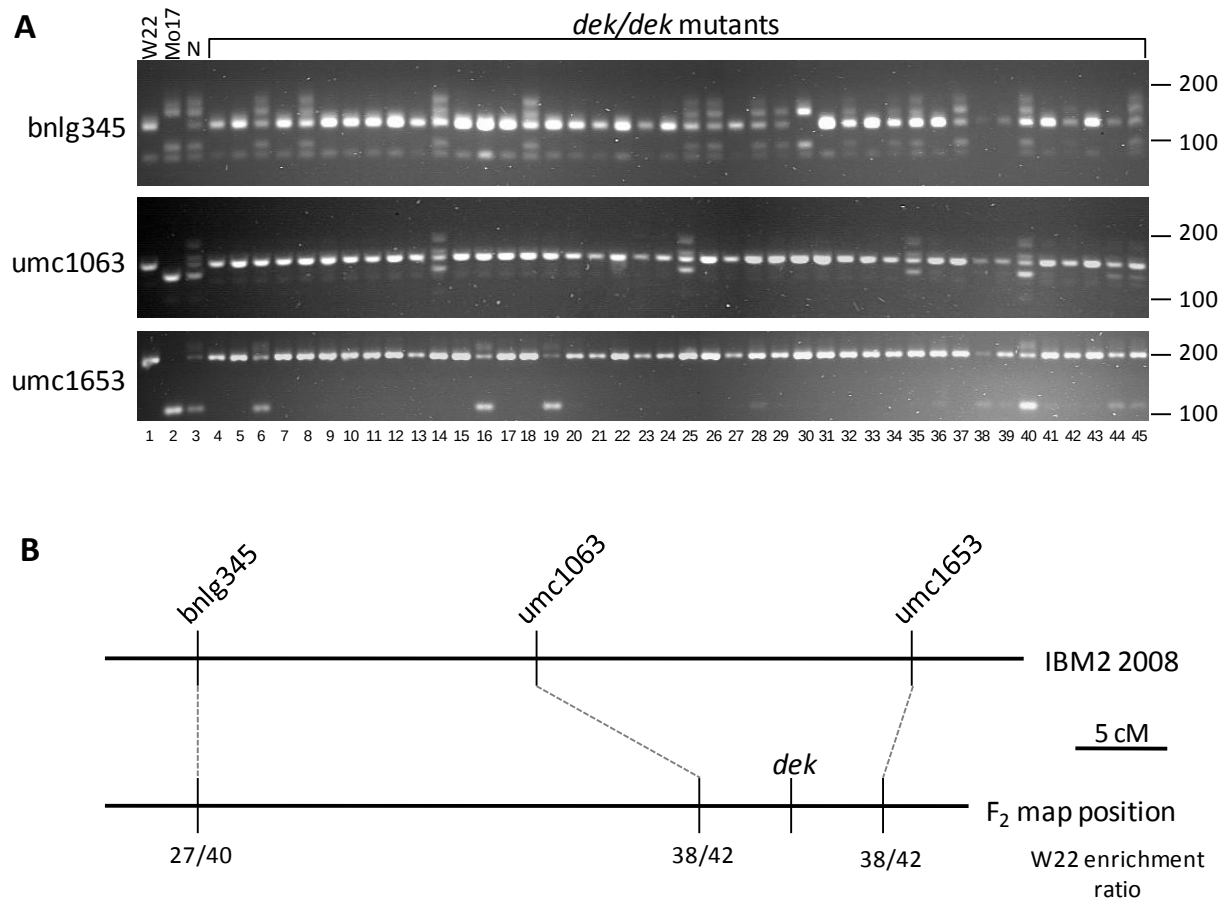

Figure 5-6. Fine map of the UniformMu *dek* locus. A) Recombination frequencies were scored using 42 individual *dek/dek* individuals from the Mo17/UniformMu F_2 mapping population. Lanes 34-45 are PCR products from severe *dek* seeds in which maternal pericarp tissue was included in the DNA extraction. Weak amplification of the Mo17 allele was attributed to contaminating parental DNA in these samples. B) Comparison of the predicted genetic distances based on IBM2 2008 Neighbors map coordinates and the observed recombination frequencies relative to the *dek* locus.

Table 5-1. Modified vs. unmodified scoring of BC1 *rgh3/rgh3* seedlings

Culture	*Planted*	*Germinated*	*Open high*	*Open low*	*Adherent high*	*Adherent low*	*No Score*
08C-0078	11	7	1	0	2	1	3
08C-0079	45	18	1	3	8	1	4
08C-0080	16	10	6	1	1	0	2
08C-0081	19	11	2	1	3	3	2
08C-0082	33	25	7	4	6	3	5
08C-0083	24	13	5	1	3	3	1
Total	**148**	**84**	**22**	**10**	**23**	**11**	**17**

Table 5-2. Number of co-dominant markers identified for each inbred pair from 505 SSR markers

	Total polymorphic markers				***Distributed marker set***			
Inbred comparison	No. of markers	Average map interval (cM)	S.D. (cM)	% Map Covered	No. of markers	Average map interval (cM)	S.D. (cM)	% Map Covered
B73/Mo17	170	12.0	12.3	91.7	70	26.9	13.2	91.1
B73/W22	161	13.0	11.8	94.6	64	29.1	12.3	90.4
Mo17/W22	154	13.5	11.6	94.5	71	26.6	12.9	92.7
Union of 3 inbred pairs	238	8.9	9.1	95.6	85	22.7	12.0	94.7

CHAPTER 6
CONCLUSIONS

Regulation of RGH3

An ever increasing amount of evidence demonstrates the relevance of alternative splicing as a regulatory mechanism in higher eukaryotes. Even though more is understood about the mechanisms and roles of alternative splicing in vertebrates and metazoans, research in plants is beginning to demonstrate the biological consequences of alternative splicing. The study of the *rgh3* mutant has provided additional understanding regarding the involvement of alternative splicing in the development of endosperm and embryo tissues in maize seeds (Fouquet et al., 2011). The goal of this dissertation is to develop knowledge about the regulation and function of RGH3 in order to understand the mechanisms of alternative splicing in development.

The data presented here suggests the RGH3 splicing factor is regulated by alternative splicing of its pre-mRNA. Regulation through alternative splicing seems to be a common mechanism of control of splicing related genes as demonstrated by multiple studies with SR-proteins and other splicing factors (Iida and Go, 2006; Chung et al., 2007; Lareau et al., 2007a; Palusa et al., 2007; Barta et al., 2008). Thus, alternative splicing of *Rgh3* and the observed localization of RGH3 protein isoforms provide additional evidences for this type of regulatory mechanism among splicing factors. Moreover, results from quantitative RT-PCR analyzes conducted by Fouquet et al. (2011), and western blot analyzes using anti-RGH3 antibodies (see Chapter 2, 3, & 4) suggest that regulation of *Rgh3* may be influenced in a tissue- or developmental-manner. Despite these observations, it is not yet clear if the main impacts of *Rgh3* splicing take place at the mRNA or protein level. Therefore, additional experiments will

be required to test these possibilities. Additionally, even though the presented data does not evidently indicate its participation, regulation of *Rgh3* by the RUST mechanism is still plausible and therefore should also be tested. Furthermore, future work at the protein level should also provide a clearer picture regarding RGH3 regulation and its role in development.

RGH3 Splicing and Post-Splicing Activities

The spliceosome is a very dynamic molecular machine formed by multiple smaller protein-RNA complexes (Wahl et al., 2009). In humans, it has been shown that URP protein participates in both U2- and U12-type spliceosomes (Tronchere et al., 1997; .Shen et al., 2010). Localization of the functional RGH3α protein to nuclear speckles, and its co-localization and interaction with U2AF^{65} strongly indicates RGH3 participates in the U2-type spliceosome. The aberrant splicing of several genes in the *rgh3-umu1* mutant showed a trend for disrupting introns with non-canonical dinucleotides suggesting RGH3 also participates in the U12 spliceosome. Future experiments to pull-down RGH3 protein complexes should provide additional information regarding where and when RGH3 participates in the spliceosome.

Identifying additional protein-protein interactions with RGH3 should also provide information to explain the localization of RGH3 to the nucleolus. The nucleolus is involved in a wide range of functions from rRNA and RNPs processing, to RNA silencing and post-transcriptional mRNA regulation (Brown and Shaw, 2008). In addition, many splicing factors have been implicated in post-splicing mechanisms including mRNA nuclear export, NMD, and mRNA translation (Long and Caceres, 2009). The localization of RGH3 to the nucleolus suggests an additional role in post-

transcriptional activities beyond splicing potentially including roles in the exon-exon junction complex or NMD.

rgh3-umu1 Allele Reveals Developmental Roles for RGH3

The *rgh3-umu1* allele provides a unique opportunity to study a hypomorphic protein of the URP family. Knockdowns of URP in human cell cultures are lethal, while the *rgh3-umu1* mutant does not affect endosperm cell viability in culture (Shen et al., 2010; Fouquet et al., 2011). Prior to identifying the *rgh3-umu1* allele, it was not feasible to use genetics to study URP orthologs due to the lethal nature of knockout alleles. Through the study of the *rgh3-umu1* allele, I was able to investigate RGH3 impacts in alternative splicing demonstrating its involvement in a reduce number of splicing events. Unfortunately, at this stage it is difficult to estimate the number of genes targeted by RGH3, and the low number of genes affected in *rgh3-umu1* may not serve as a reference for a whole genome scale. Nevertheless, the data argues that RGH3 is involved in regulation of genes that are likely to have significant impact on the development of the maize seed. The hypomorphic allele also facilitated studies of RGH3 domain functions and will likely allow studies of protein-protein interactions in a defective background. From these data and the data presented by Fouquet et al. (2011) regarding the phenotypic impacts of *rgh3-umu1* on seed development, it can be argued that RGH3 is involved in the regulation of proteins that are important for proper cell differentiation. Defects in these proteins in turn affect development of the endosperm and influence development of the embryo. Rgh3 has a separate embryo-specific function that translates into the observed seedling phenotypes (Fouquet et al., 2011).

Finally, through the use of a distributed set of SSR markers, I was able to place a *rgh3-umu1* seedling allele modifier to a 3 Mb interval on the short arm of chromosome 9

(Figure 5-5). The mapped interval contains too many genes to discriminate among all of them in search of a candidate gene. In addition, the range of *rgh3-umu1* seedling phenotypes has hindered the possibilities to fine-map a potential gene. Though, by combining this map position with results from other experiments that could uncover RGH3 gene targets or interacting proteins that map within this interval it will be possible to find a candidate gene. These additional experiment may include RNA-sequence analyzes, RNA-immunoprecipitation analyzes, or protein-complexes pull-down experiments among others. The nature of the modifier gene is still a matter of speculation. However, since the changes in phenotype are observed at the seedling and not at the seed level it could be argued that the modifier gene acts in embryo tissues specifically. Given the variability of the seedling phenotype, and the hypomorphic nature of $RGH3^{umu}\alpha$ protein, it is also probable that the modifier gene directly interacts with RGH3 either in a protein complex or as a target transcript. Taken together, it can be argued that the seedling modifier is either a splicing factor that interacts with RGH3 or a gene target, such as a transcription factor, that impacts seedling development.

LIST OF REFERENCES

Abdalla, K.O., Thomson, J.A., Rafudeen, M.S. (2009). Protocols for nuclei isolation and nuclear protein extraction from theresurrection plant *Xerophyta viscosa* for proteomic studies. Analytic Bioch **384,** 365–367

Ahern, K.R., Deewatthanawong, P., Schares, J., Muszynski, M., Weeks, R., Vollbrecht, E., Duvick, J., Brendel, V.P., and Brutnell, T.P. (2009). Regional mutagenesis using Dissociation in maize. Methods **49,** 248-254.

Ali, G.S., and Reddy, A.S. (2008a). Regulation of alternative splicing of pre-mRNAs by stresses. Curr Top Microbiol Immunol **326,** 257-275.

Ali, G.S., and Reddy, A.S. (2008b). Spatiotemporal organization of pre-mRNA splicing proteins in plants. Curr Top Microbiol Immunol **326,** 103-118.

Ali, G.S., Palusa, S.G., Golovkin, M., Prasad, J., Manley, J.L., and Reddy, A.S. (2007). Regulation of plant developmental processes by a novel splicing factor. PLoS One **2,** e471.

Appleby, N., Edwards, D., and Batley, J. (2009). New technologies for ultra-high throughput genotyping in plants. Methods Mol Biol **513,** 19-39.

Banerjee, H., Rahn, A., Gawande, B., Guth, S., Valcarcel, J., and Singh, R. (2004). The conserved RNA recognition motif 3 of U2 snRNA auxiliary factor (U2AF 65) is essential in vivo but dispensable for activity in vitro. RNA **10,** 240-253.

Barbaglia, A.M., Klusman, K.M., Higgins, J., Shaw, J.R., Hannah, L.C., and Lal, S.K. (2012). Gene capture by Helitron transposons reshuffles the transcriptome of maize. Genetics **190,** 965-975.

Barta, A., Kalyna, M., and Lorković, Z.J. (2008). Plant SR proteins and their functions. Curr Top Microbiol Immunol **326,** 83-102.

Bennetzen, J.L. (2000). Transposable element contributions to plant gene and genome evolution. Plant Mol Biol **42,** 251-269.

Black, D.L. (2003). Mechanisms of alternative pre-messenger RNA splicing. Annu Rev Biochem **72,** 291-336.

Bortiri, E., Jackson, D., and Hake, S. (2006). Advances in maize genomics: the emergence of positional cloning. Curr Opin Plant Biol **9,** 164-171.

Brown, J.W., and Shaw, P.J. (2008). The role of the plant nucleolus in pre-mRNA processing. Curr Top Microbiol Immunol **326,** 291-311.

Brutnell, T.P. (2002). Transposon tagging in maize. Funct Integr Genomics **2,** 4-12.

Carson, M.L., Stuber, C.W., and Senior, M.L. (2004). Identification and Mapping of Quantitative Trait Loci Conditioning Resistance to Southern Leaf Blight of Maize Caused by Cochliobolus heterostrophus Race O. Phytopathology **94,** 862-867.

Chang, M.T., Neuffer M.G. (1994). Endosperm-embryo interaction in maize. Maydica **39,** 9–18

Chang, Y.F., Imam, J.S., and Wilkinson, M.F. (2007). The nonsense-mediated decay RNA surveillance pathway. Annu Rev Biochem **76,** 51-74.

Chen, M., and Manley, J.L. (2009). Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches. Nat Rev Mol Cell Biol **10,** 741-754.

Cho, S., Hoang, A., Sinha, R., Zhong, X.Y., Fu, X.D., Krainer, A.R., and Ghosh, G. (2011). Interaction between the RNA binding domains of Ser-Arg splicing factor 1 and U1-70K snRNP protein determines early spliceosome assembly. Proc Natl Acad Sci U S A **108,** 8233-8238.

Citovsky, V., Lee, L.Y., Vyas, S., Glick, E., Chen, M.H., Vainstein, A., Gafni, Y., Gelvin, S.B., and Tzfira, T. (2006). Subcellular localization of interacting proteins by bimolecular fluorescence complementation in planta. J Mol Biol **362,** 1120-1131.

Cowperthwaite, M., Park, W., Xu, Z., Yan, X., Maurais, S.C., and Dooner, H.K. (2002). Use of the transposon Ac as a gene-searching engine in the maize genome. Plant Cell **14,** 713-726.

Diao, X.M., and Lisch, D. (2006a). Mutator transposon in maize and MULEs in the plant genome. Yi Chuan Xue Bao **33,** 477-487.

Docquier, S., Tillemans, V., Deltour, R., and Motte, P. (2004). Nuclear bodies and compartmentalization of pre-mRNA splicing factors in higher plants. Chromosoma **112,** 255-266.

Domon, D., Lorkovic, Z.J.,Valcarcel, J., and Filipowicz, W. (1998). Multiple Forms of the U2 Small Nuclear Ribonucleoprotein Auxiliary Factor U2AF Subunits Expressed in Higher Plants. J Bio Chem. **273,** 34603-34610

Dowhan, D.H., Hong, E.P., Auboeuf, D., Dennis, A.P., Wilson, M.M., Berget, S.M., and O'Malley, B.W. (2005). Steroid hormone receptor coactivation and alternative RNA splicing by U2AF65-related proteins CAPERalpha and CAPERbeta. Mol Cell **17,** 429-439.

Fang, Y., Hearn, S., and Spector, D.L. (2004). Tissue-specific expression and dynamic organization of SR splicing factors in Arabidopsis. Mol Biol Cell **15,** 2664-2673.

Fajardo, D. (2008). Putative Role for RNA Splicing in Maize Endosperm-Embryo

Developmental Interactions is Revealed by the *rough endosperm 3* (*rgh3*) Seed Mutant. PhD Dissertation. University of Florida, Gainesville, Florida.

Fernandes, J., Dong, Q., Schneider, B., Morrow, D.J., Nan, G.L., Brendel, V., and Walbot, V. (2004). Genome-wide mutagenesis of Zea mays L. using RescueMu transposons. Genome Biol **5,** R82.

Feschotte, C., Jiang, N., and Wessler, S.R. (2002). Plant transposable elements: where genetics meets genomics. Nat Rev Genet **3,** 329-341.

Filichkin, S.A., Priest, H.D., Givan, S.A., Shen, R., Bryant, D.W., Fox, S.E., Wong, W.K., and Mockler, T.C. (2010). Genome-wide mapping of alternative splicing in Arabidopsis thaliana. Genome Res **20,** 45-58.

Fouquet, R., Martin, F., Fajardo, D.S., Gault, C.M., Gómez, E., Tseung, C.W., Policht, T., Hueros, G., and Settles, A.M. (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.

Frilander, M.J., and Steitz, J.A. (1999). Initial recognition of U12-dependent introns requires both U11/5' splice-site and U12/branchpoint interactions. Genes Dev **13,** 851-863.

Fu, Y., Wen, T.J., Ronin, Y.I., Chen, H.D., Guo, L., Mester, D.I., Yang, Y., Lee, M., Korol, A.B., Ashlock, D.A., and Schnable, P.S. (2006). Genetic dissection of intermated recombinant inbred lines using a new genetic map of maize. Genetics **174,** 1671-1683.

Gallavotti, A., Barazesh, S., Malcomber, S., Hall, D., Jackson, D., Schmidt, R.J., and McSteen, P. (2008). sparse inflorescence1 encodes a monocot-specific YUCCA-like gene required for vegetative and reproductive development in maize. Proc Natl Acad Sci U S A **105,** 15196-15201.

Girard, L., and Freeling, M. (1999). Regulatory changes as a consequence of transposon insertion. Dev Genet **25,** 291-296.

Golovkin, M., and Reddy, A.S. (1998). The plant U1 small nuclear ribonucleoprotein particle 70K protein interacts with two novel serine/arginine-rich proteins. Plant Cell **10,** 1637-1648.

Gutiérrez-Marcos, J.F., Dal Prà, M., Giulini, A., Costa, L.M., Gavazzi, G., Cordelier, S., Sellam, O., Tatout, C., Paul, W., Perez, P., Dickinson, H.G., and Consonni, G. (2007). empty pericarp4 encodes a mitochondrion-targeted pentatricopeptide repeat protein necessary for seed development and plant growth in maize. Plant Cell **19,** 196-210.

Haldane, J.B.S. (1919). The combination of linkage values and the calculation of distance between the loci of linked factors. J Genet **8,** 299–309

Hamblin, M.T., Warburton, M.L., and Buckler, E.S. (2007). Empirical comparison of Simple Sequence Repeats and single nucleotide polymorphisms in assessment of maize diversity and relatedness. PLoS One **2,** e1367.

Hastings, M.L., Allemand, E., Duelli, D.M., Myers, M.P., and Krainer, A.R. (2007). Control of pre-mRNA splicing by the general splicing factors PUF60 and U2AF(65). PLoS One **2,** e538.

Heckel, T., Werner, K., Sheridan, W.F., Dumas, C., and Rogowsky, P.M. (1999). Novel phenotypes and developmental arrest in early embryo specific mutants of maize. Planta **210,** 1-8.

Horton, R.M., Hunt, H.D., Ho, S.N., Pullen, J.K., and Pease, L.R. (1989). Engineering hybrid genes without the use of restriction enzymes: gene splicing by overlap extension. Gene **77,** 61-68.

Hua-Van, A., Le Rouzic, A., Maisonhaute, C., and Capy, P. (2005). Abundance, distribution and dynamics of retrotransposable elements and transposons: similarities and differences. Cytogenet Genome Res **110,** 426-440.

Iida, K., and Go, M. (2006). Survey of conserved alternative splicing events of mRNAs encoding SR proteins in land plants. Mol Biol Evol **23,** 1085-1094.

Isshiki, M., Tsumoto, A., and Shimamoto, K. (2006). The serine/arginine-rich protein family in rice plays important roles in constitutive and alternative splicing of pre-mRNA. Plant Cell **18,** 146-158.

Jander, G., Norris, S.R., Rounsley, S.D., Bush, D.F., Levin, I.M., and Last, R.L. (2002). Arabidopsis map-based cloning in the post-genome era. Plant Physiol **129,** 440-450.

Jones, E.S., Sullivan, H., Bhattramakki, D., and Smith, J.S. (2007). A comparison of simple sequence repeat and single nucleotide polymorphism marker technologies for the genotypic analysis of maize (Zea mays L.). Theor Appl Genet **115,** 361-371.

Kalyna, M., Lopato, S., and Barta, A. (2003). Ectopic expression of atRSZ33 reveals its function in splicing and causes pleiotropic changes in development. Mol Biol Cell **14,** 3565-3577.

Kalyna, M., Simpson, C.G., Syed, N.H., Lewandowska, D., Marquez, Y., Kusenda, B., Marshall, J., Fuller, J., Cardle, L., McNicol, J., Dinh, H.Q., Barta, A., and Brown, J.W. (2012). Alternative splicing and nonsense-mediated decay modulate expression of important regulatory genes in Arabidopsis. Nucleic Acids Res **40,** 2454-2469.

Karimi, M., Inzé, D., and Depicker, A. (2002). GATEWAY vectors for Agrobacterium-mediated plant transformation. Trends Plant Sci **7,** 193-195.

Karimi, M., Depicker, A., and Hilson, P. (2007). Recombinational cloning with plant gateway vectors. Plant Physiol **145,** 1144-1154.

Keren, H., Lev-Maor, G., and Ast, G. (2010). Alternative splicing and evolution: diversification, exon definition and function. Nat Rev Genet **11,** 345-355.

Kidwell, M.G, Lisch, D. (1997). Transposable elements as sources of variation in animals and plants. Proc. Natl. Acad. Sci. U.S.A. **94**, 7704-7711

Kielkopf, C.L., Lücke, S., and Green, M.R. (2004). U2AF homology motifs: protein recognition in the RRM world. Genes Dev **18,** 1513-1526.

Kielkopf, C.L., Rodionova, N.A., Green, M.R., and Burley, S.K. (2001). A novel peptide recognition mode revealed by the X-ray structure of a core U2AF35/U2AF65 heterodimer. Cell **106,** 595-605.

Kim, S.H., Koroleva, O.A., Lewandowska, D., Pendle, A.F., Clark, G.P., Simpson, C.G., Shaw, P.J., and Brown, J.W. (2009). Aberrant mRNA transcripts and the nonsense-mediated decay proteins UPF2 and UPF3 are enriched in the Arabidopsis nucleolus. Plant Cell **21,** 2045-2057.

Kitagawa, K., Wang, X., Hatada, I., Yamaoka, T., Nojima, H., Inazawa, J., Abe, T., Mitsuya, K., Oshimura, M., and Murata, A. (1995). Isolation and mapping of human homologues of an imprinted mouse gene U2af1-rs1. Genomics **30,** 257-263.

Kolkman, J.M., Conrad, L.J., Farmer, P.R., Hardeman, K., Ahern, K.R., Lewis, P.E., Sawers, R.J., Lebejko, S., Chomet, P., and Brutnell, T.P. (2005). Distribution of Activator (Ac) throughout the maize genome for use in regional mutagenesis. Genetics **169,** 981-995.

Koroleva, O.A., Calder, G., Pendle, A.F., Kim, S.H., Lewandowska, D., Simpson, C.G., Jones, I.M., Brown, J.W., and Shaw, P.J. (2009). Dynamic behavior of Arabidopsis eIF4A-III, putative core protein of exon junction complex: fast relocation to nucleolus and splicing speckles under hypoxia. Plant Cell **21,** 1592-1606.

Lamond, A.I., and Spector, D.L. (2003). Nuclear speckles: a model for nuclear organelles. Nat Rev Mol Cell Biol **4,** 605-612.

Lareau, L.F., Brooks, A.N., Soergel, D.A., Meng, Q., and Brenner, S.E. (2007a). The coupling of alternative splicing and nonsense-mediated mRNA decay. Adv Exp Med Biol **623,** 190-211.

Lareau, L.F., Inada, M., Green, R.E., Wengrod, J.C., and Brenner, S.E. (2007b). Unproductive splicing of SR genes associated with highly conserved and ultraconserved DNA elements. Nature **446,** 926-929.

Lee, M., Sharopova, N., Beavis, W.D., Grant, D., Katt, M., Blair, D., and Hallauer, A. (2002). Expanding the genetic map of maize with the intermated B73 x Mo17 (IBM) population. Plant Mol Biol **48,** 453-461.

Lewis, B.P., Green, R.E., and Brenner, S.E. (2003). Evidence for the widespread coupling of alternative splicing and nonsense-mediated mRNA decay in humans. Proc Natl Acad Sci U S A **100,** 189-192.

Lisch, D. (2002). Mutator transposons. Trends Plant Sci **7,** 498-504.

Liu, S., Chen, H.D., Makarevitch, I., Shirmer, R., Emrich, S.J., Dietrich, C.R., Barbazuk, W.B., Springer, N.M., and Schnable, P.S. (2010). High-throughput genetic mapping of mutants via quantitative single nucleotide polymorphism typing. Genetics **184,** 19-26.

Long, J.C., and Caceres, J.F. (2009). The SR protein family of splicing factors: master regulators of gene expression. Biochem J **417,** 15-27.

Lopato, S., Kalyna, M., Dorner, S., Kobayashi, R., Krainer, A.R., and Barta, A. (1999). atSRp30, one of two SF2/ASF-like proteins from Arabidopsis thaliana, regulates splicing of specific plant genes. Genes Dev **13,** 987-1001.

Lopes, M.A., Takasaki, K., Bostwick, D.E., Helentjaris, T., and Larkins, B.A. (1995). Identification of two opaque2 modifier loci in quality protein maize. Mol Gen Genet **247,** 603-613.

Lorkovic, Z.J., Lehner, R., Forstner, C., and Barta, A. (2005). Evolutionary conservation of minor U12-type spliceosome between plants and humans. RNA **11,** 1095-1107.

Lorković, Z.J., Hilscher, J., and Barta, A. (2004). Use of fluorescent protein tags to study nuclear organization of the spliceosomal machinery in transiently transformed living plant cells. Mol Biol Cell **15,** 3233-3243.

Lorković, Z.J., Hilscher, J., and Barta, A. (2008). Co-localisation studies of Arabidopsis SR splicing factors reveal different types of speckles in plant cell nuclei. Exp Cell Res **314,** 3175-3186.

Lu, T., Lu, G., Fan, D., Zhu, C., Li, W., Zhao, Q., Feng, Q., Zhao, Y., Guo, Y., Huang, X., and Han, B. (2010). Function annotation of the rice transcriptome at single-nucleotide resolution by RNA-seq. Genome Res **20,** 1238-1249.

Marquez, Y., Brown, J.W., Simpson, C., Barta, A., and Kalyna, M. (2012). Transcriptome survey reveals increased complexity of the alternative splicing landscape in Arabidopsis. Genome Res **22,** 1184-1195.

Matlin, A.J., Clark, F., and Smith, C.W. (2005). Understanding alternative splicing: towards a cellular code. Nat Rev Mol Cell Biol **6,** 386-398.

Matsuoka, Y., Vigouroux, Y., Goodman, M.M., Sanchez G, J., Buckler, E., and Doebley, J. (2002). A single domestication for maize shown by multilocus microsatellite genotyping. Proc Natl Acad Sci U S A **99,** 6080-6084.

May, B.P., Liu, H., Vollbrecht, E., Senior, L., Rabinowicz, P.D., Roh, D., Pan, X., Stein, L., Freeling, M., Alexander, D., and Martienssen, R. (2003). Maize-targeted mutagenesis: A knockout resource for maize. Proc Natl Acad Sci U S A **100,** 11541-11546.

McCarty, D.R., Settles, A.M., Suzuki, M., Tan, B.C., Latshaw, S., Porch, T., Robin, K., Baier, J., Avigne, W., Lai, J., Messing, J., Koch, K.E., and Hannah, L.C. (2005). Steady-state transposon mutagenesis in inbred maize. Plant J **44,** 52-61.

McGlincy, N.J., and Smith, C.W. (2008). Alternative splicing resulting in nonsense-mediated mRNA decay: what is the meaning of nonsense? Trends Biochem Sci **33,** 385-393.

Mollet, I., Barbosa-Morais, N.L., Andrade, J., and Carmo-Fonseca, M. (2006). Diversity of human U2AF splicing factors. FEBS J **273,** 4807-4816.

Moore, M.J., and Sharp, P.A. (1993). Evidence for two active sites in the spliceosome provided by stereochemistry of pre-mRNA splicing. Nature **365,** 364-368.

Neuffer, M.G., England, D. (1995) Induced mutations with confirmed locations. Maize Genetics Coop News Lett 69:43–46

Neuffer, M.G., Coe, E.H., and Wessler, S.R. (1997). Mutants of Maize. (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press)

Nilsen, T.W., and Graveley, B.R. (2010). Expansion of the eukaryotic proteome by alternative splicing. Nature **463,** 457-463.

Ortiz, D.F., and Strommer, J.N. (1990). The Mu1 maize transposable element induces tissue-specific aberrant splicing and polyadenylation in two Adh1 mutants. Mol Cell Biol **10,** 2090-2095.

Palusa, S.G., and Reddy, A.S. (2010). Extensive coupling of alternative splicing of pre-mRNAs of serine/arginine (SR) genes with nonsense-mediated decay. New Phytol **185,** 83-89.

Palusa, S.G., Ali, G.S., and Reddy, A.S. (2007). Alternative splicing of pre-mRNAs of Arabidopsis serine/arginine-rich proteins: regulation by hormones and stresses. Plant J **49,** 1091-1107.

Patel, A.A., and Steitz, J.A. (2003). Splicing double: insights from the second spliceosome. Nat Rev Mol Cell Biol **4,** 960-970.

Pendle, A.F., Clark, G.P., Boon, R., Lewandowska, D., Lam, Y.W., Andersen, J., Mann, M., Lamond, A.I., Brown, J.W., and Shaw, P.J. (2005). Proteomic analysis of the Arabidopsis nucleolus suggests novel nucleolar functions. Mol Biol Cell **16,** 260-269.

Pribat, A., Noiriel, A., Morse, A.M., Davis, J.M., Fouquet, R., Loizeau, K., Ravanel, S., Frank, W., Haas, R., Reski, R., Bedair, M., Sumner, L.W., and Hanson, A.D. (2010). Nonflowering plants possess a unique folate-dependent phenylalanine hydroxylase that is localized in chloroplasts. Plant Cell **22,** 3410-3422.

Proost, S., Van Bel, M., Sterck, L., Billiau, K., Van Parys, T., Van de Peer, Y., and Vandepoele, K. (2009). PLAZA: a comparative genomics resource to study gene and genome evolution in plants. Plant Cell **21,** 3718-3731.

Purugganan, M., and Wessler, S. (1992). The splicing of transposable elements and its role in intron evolution. Genetica **86,** 295-303.

Quesada, V., Macknight, R., Dean, C., and Simpson, G.G. (2003). Autoregulation of FCA pre-mRNA processing controls Arabidopsis flowering time. EMBO J **22,** 3142-3152.

Reddy, A.S.N. (2001). Nuclear pre-mRNA splicing in plants. Crit. Rev. Plant Sci. **20,** 523–71

Reddy, A.S. (2007). Alternative splicing of pre-messenger RNAs in plants in the genomic era. Annu Rev Plant Biol **58,** 267-294.

Reddy, A.S., Day, I.S., Göhring, J., and Barta, A. (2012). Localization and dynamics of nuclear speckles in plants. Plant Physiol **158,** 67-77.

Reid, K.E., Olsson, N., Schlosser, J., Peng, F., and Lund, S.T. (2006). An optimized grapevine RNA isolation procedure and statistical determination of reference genes for real-time RT-PCR during berry development. BMC Plant Biol **6,** 27.

Scanlon, M.J., Stinard, P.S., James, M.G., Myers, A.M., and Robertson, D.S. (1994). Genetic analysis of 63 mutations affecting maize kernel development isolated from Mutator stocks. Genetics **136,** 281-294.

Selenko, P., Gregorovic, G., Sprangers, R., Stier, G., Rhani, Z., Krämer, A., and Sattler, M. (2003). Structural basis for the molecular recognition between human splicing factors U2AF65 and SF1/mBBP. Mol Cell **11,** 965-976.

Settles, A.M., Latshaw, S., and McCarty, D.R. (2004). Molecular analysis of high-copy insertion sites in maize. Nucleic Acids Res **32,** e54.

Settles, A.M., Holding, D.R., Tan, B.C., Latshaw, S.P., Liu, J., Suzuki, M., Li, L., O'Brien, B.A., Fajardo, D.S., Wroclawska, E., Tseung, C.W., Lai, J., Hunter, C.T., Avigne, W.T., Baier, J., Messing, J., Hannah, L.C., Koch, K.E., Becraft, P.W., Larkins, B.A., and McCarty, D.R. (2007). Sequence-indexed mutations in maize using the UniformMu transposon-tagging population. BMC Genomics **8,** 116.

Settles, A.M. (2009). Transposon tagging and reverse genetics. In: Kriz A, Larkens B (eds) Molecular genetic approaches to maize improvement. Springer, Berlin, pp 143–160

Sharopova, N., McMullen, M.D., Schultz, L., Schroeder, S., Sanchez-Villeda, H., Gardiner, J., Bergstrom, D., Houchins, K., Melia-Hancock, S., Musket, T., Duru, N., Polacco, M., Edwards, K., Ruff, T., Register, J.C., Brouwer, C., Thompson, R., Velasco, R., Chin, E., Lee, M., Woodman-Clikeman, W., Long, M.J., Liscum, E., Cone, K., Davis, G., and Coe, E.H. (2002). Development and mapping of SSR markers for maize. Plant Mol Biol **48,** 463-481.

Shav-Tal, Y., Blechman, J., Darzacq, X., Montagna, C., Dye, B.T., Patton, J.G., Singer, R.H., and Zipori, D. (2005). Dynamic sorting of nuclear components into distinct nucleolar caps during transcriptional inhibition. Mol Biol Cell **16,** 2395-2413.

Shen, H., Zheng, X., Luecke, S., and Green, M.R. (2010). The U2AF35-related protein Urp contacts the 3' splice site to promote U12-type intron splicing and the second step of U2-type intron splicing. Genes Dev **24,** 2389-2394.

Slotkin, R.K., and Martienssen, R. (2007). Transposable elements and the epigenetic regulation of the genome. Nat Rev Genet **8,** 272-285.

Spector, D.L., and Lamond, A.I. (2011). Nuclear speckles. Cold Spring Harb Perspect Biol **3**.

Stauffer, E., Westermann, A., Wagner, G., and Wachter, A. (2010). Polypyrimidine tract-binding protein homologues from Arabidopsis underlie regulatory circuits based on alternative splicing and downstream control. Plant J **64,** 243-255.

Tanabe, N., Yoshimura, K., Kimura, A., Yabuta, Y., and Shigeoka, S. (2007). Differential expression of alternatively spliced mRNAs of Arabidopsis SR protein homologs, atSR30 and atSR45a, in response to environmental stress. Plant Cell Physiol **48,** 1036-1049.

Tavanez, J.P., Madl, T., Kooshapur, H., Sattler, M., and Valcárcel, J. (2012). hnRNP A1 proofreads 3' splice site recognition by U2AF. Mol Cell **45,** 314-329.

Thompson, B.E., Bartling, L., Whipple, C., Hall, D.H., Sakai, H., Schmidt, R., and Hake, S. (2009). bearded-ear encodes a MADS box transcription factor critical for maize floral development. Plant Cell **21,** 2578-2590.

Till, B.J., Reynolds, S.H., Weil, C., Springer, N., Burtner, C., Young, K., Bowers, E., Codomo, C.A., Enns, L.C., Odden, A.R., Greene, E.A., Comai, L., and Henikoff, S. (2004). Discovery of induced point mutations in maize genes by TILLING. BMC Plant Biol **4,** 12.

Tillemans, V., Dispa, L., Remacle, C., Collinge, M., and Motte, P. (2005). Functional distribution and dynamics of Arabidopsis SR splicing factors in living plant cells. Plant J **41,** 567-582.

Tillemans, V., Leponce, I., Rausin, G., Dispa, L., and Motte, P. (2006). Insights into nuclear organization in plants as revealed by the dynamic distribution of Arabidopsis SR splicing factors. Plant Cell **18,** 3218-3234.

Tronchère, H., Wang, J., and Fu, X.D. (1997). A protein related to splicing factor U2AF35 that interacts with U2AF65 and SR proteins in splicing of pre-mRNA. Nature **388,** 397-400.

Tzfira, T., Tian, G.W., Lacroix, B., Vyas, S., Li, J., Leitner-Dagan, Y., Krichevsky, A., Taylor, T., Vainstein, A., and Citovsky, V. (2005). pSAT vectors: a modular series of plasmids for autofluorescent protein tagging and expression of multiple genes in plants. Plant Mol Biol **57,** 503-516.

Vollbrecht, E., Reiser, L., and Hake, S. (2000). Shoot meristem size is dependent on inbred background and presence of the maize homeobox gene, knotted1. Development **127,** 3161-3172.

Wahl, M.C., Will, C.L., and Lührmann, R. (2009). The spliceosome: design principles of a dynamic RNP machine. Cell **136,** 701-718.

Wang, B.B., and Brendel, V. (2004). The ASRG database: identification and survey of Arabidopsis thaliana genes involved in pre-mRNA splicing. Genome Biol **5,** R102.

Wang, B.B., and Brendel, V. (2006a). Molecular characterization and phylogeny of U2AF35 homologs in plants. Plant Physiol **140,** 624-636.

Wang, B.B., and Brendel, V. (2006b). Genomewide comparative analysis of alternative splicing in plants. Proc Natl Acad Sci U S A **103,** 7175-7180.

Wang, E.T., Sandberg, R., Luo, S., Khrebtukova, I., Zhang, L., Mayr, C., Kingsmore, S.F., Schroth, G.P., and Burge, C.B. (2008). Alternative isoform regulation in human tissue transcriptomes. Nature **456,** 470-476.

Well, C.F., Wessler, S.R (1990). The effects of plant transposable element inserfion on transcription initiation and RNA processing. Annu Rev Plant Physiol Plant Mol Biol **41,** 527-552

Wessler, S.R., Bureau, T.E., and White, S.E. (1995). LTR-retrotransposons and MITEs: important players in the evolution of plant genomes. Curr Opin Genet Dev **5,** 814-821.

Will, C.L., Schneider, C., Hossbach, M., Urlaub, H., Rauhut, R., Elbashir, S., Tuschl, T., and Lührmann, R. (2004). The human 18S U11/U12 snRNP contains a set of novel proteins not found in the U2-dependent spliceosome. RNA **10,** 929-941.

Wise, A.A., Liu, Z., and Binns, A.N. (2006). Three methods for the introduction of foreign DNA into Agrobacterium. Methods Mol Biol **343,** 43-53.

Wu, S., Romfo, C.M., Nilsen, T.W., and Green, M.R. (1999). Functional recognition of the 3' splice site AG by the splicing factor U2AF35. Nature **402,** 832-835.

Xing, Y., and Lee, C. (2006). Alternative splicing and RNA selection pressure--evolutionary consequences for eukaryotic genomes. Nat Rev Genet **7,** 499-509.

Zamore, P.D., Patton, J.G., and Green, M.R. (1992). Cloning and domain structure of the mammalian splicing factor U2AF. Nature **355,** 609-614.

Zhang, X.N., and Mount, S.M. (2009). Two alternatively spliced isoforms of the Arabidopsis SR45 protein have distinct roles during normal plant development. Plant Physiol **150,** 1450-1458.

BIOGRAPHICAL SKETCH

Federico Martin was born in Santa Fe, Argentina. He graduated from El Portal high school in December 1999. He then moved to Tempe, Arizona to become part of the men's swimming varsity team at Arizona State University where he began his undergraduate studies in January 2001. While an undergraduate student, Federico worked as a research assistant in the Shool of Life Sciences under the supervision of Dr. Willem Vermaas studying hydrogen photoproduction in the cyanobacteria *Shynechocystis sp pcc* 6803. He then switched laboratories to work as a research assistant in the Center for Infectious Diseases and Vaccinology under the supervision of Dr. Guy Cardineau. The research topic was plant-based vaccine production methods. Federico graduated in May 2005 with a Bachelors of Science degree in molecular biosciences and biotechnology. After his graduation, he continued working as a research technician under Dr. Guy Cardineau's supervision. In August 2006, he was accepted in the Plant Molecular and Cellular Biology (PMCB) program at the University of Florida. In January 2007, he initiated his Ph.D project under the supervision of Dr. A. Mark Settles studying the function of the ROUGH ENDOSPERM3 (RGH3) protein in the model system *Zea mays.*